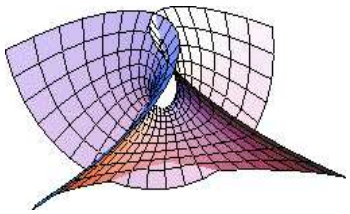
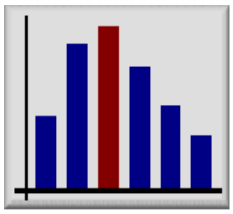


Lehrstuhl für Statistik
Institut für Mathematik
Universität Würzburg



Simulation and Estimation in Multivariate Generalized Pareto Models

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades
der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von

René Michel

aus

Würzburg

April 2006

"Irren ist menschlich,
aber um richtig Fehler zu machen,
braucht man einen Computer."

Harald Lesch, α-Centauri

Danksagung

Die vorliegende Arbeit entstand im Laufe meiner Tätigkeit am Lehrstuhl für Statistik der Universität Würzburg. An erster Stelle sei hier Herrn Prof. Dr. Michael Falk für die hervorragende Betreuung herzlichst gedankt. Er unterstützte mich in jeder Hinsicht und hatte für meine Probleme jeglicher Art stets ein offenes Ohr. Die lebhaften und ergiebigen Diskussionen werden mir immer in sehr guter Erinnerung bleiben.

Den weiteren Mitarbeitern des Lehrstuhls für Statistik gebührt mein Dank für eine äußerst angenehme Arbeitsatmosphäre, die die letzten drei Jahre unvergesslich macht. Hierbei seien Karin Krumpholz, die "gute Seele" des Lehrstuhls, sowie natürlich Dipl.-Math. Christian Weiß, PD Dr. Frank Marohn, Prof. Dr. Rainer Göb und Dr. Martin Kleinsteuber genannt.

Nicht vergessen möchte ich auch Dr. Richard Greiner, der mir vom ersten Tag meines Studiums an bis heute immer als Ansprechpartner zur Verfügung stand.

Ganz besonders bedanken möchte ich mich bei meiner Freundin Anja, die mich immer unterstützt und ermuntert hat und die einfach immer für mich da war.

Meiner Familie, die mich meinen Weg hat gehen lassen und mir dabei immer zur Seite stand, und meinen Freunden, die immer für die nötige Ablenkung sorgten, möchte ich meine Dankbarkeit zum Ausdruck bringen.

Beim Bayerischen Landesamt für Umwelt, insbesondere bei Herrn Rudolf Neusiedl, möchte ich mich für den Datensatz in Kapitel 7 bedanken.

Zum Schluß sei allen denen ein Dankeschön ausgesprochen, die hier nicht namentlich Erwähnung fanden, aber zur Entstehung dieser Arbeit beigetragen haben.

Würzburg, im April 2006 René Michel

Contents

Chapter 1

Introduction

The field of extreme value theory has been a very active research area in recent years. This is due to the fact that it is of relevance in many practical problems such as the analysis of

- sea levels (Tawn (1988, [72])),
- tidal waves (Coles and Tawn (1991, [10]),
- wind speeds (deHaan and deRonde (1998, [33])),
- rainfall (Schlather and Tawn (2003, [62])),
- metallurgical data (Beirlant et al. (2005, [3])),
- financial data (Hsing et al. (2004, [40])),
- currency exchange rates (Hauksson et al. (2001, [35])),
- insurance data (Reiss and Thomas (2001, [57])),

only to name a few. From the naturally broad variety of literature concerning extreme value theory we refer to Reiss and Thomas (2001, [57]), Coles (2001, [8]) and Beirlant et al. (2005, [3]) for general introductions to the field. The probabilistic background is discussed in detail, for example, in Falk et al. (2004, [21]), Resnick (1987, [59]) and Galambos (1978, [29]).

The goal of the present text is to investigate and simulate multivariate generalized Pareto distributions (GPDs). Different to the univariate case, where GPDs have turned out to be crucial models for the peaks-over-threshold (POT) approach by the Balkema–deHaan–Pickands Theorem (see Section 1.4 of Reiss and Thomas (2001, [57])), the role of multivariate GPDs in the framework of extreme value theory is, still, under scrutiny. Recent results, however, in Rootzén and Tajvidi (2005, [60]), Tajvidi (1996, [71]), Beirlant et al. (2005, [3]) and Falk et al. (2004, [21]) show that multivariate GPDs can again be used for (multivariate) POT approaches.

A fact illustrating that multivariate GPDs are a recent research topic is that there are slightly varying definitions of multivariate GPDs under investigation. For this manuscript we stick to the definition of Section 5.1 of Falk et al. (2004, [21]), which was originally introduced by Kaufmann and Reiss (1995, [45]) for the bivariate case. There are two other definitions closely related to this one. The first one is to be found in Paper B of Tajvidi (1996, [71]), and the other is presented in Section 8.3 of Beirlant et al. (2005, [3]) and is investigated in more detail in Rootzén

and Tajvidi (2005, [60]). These definitions are both identical to ours in the region of interest, see Section 10.1 of Reiss and Thomas (2001, [57]) and Section 8.3 of Beirlant et al. (2005, [3]). In contrast to the univariate case it is not intuitively clear, how exceedances over high thresholds are to be defined. A different approach with a different understanding of extreme observations, which results in a single-parameter family also called multivariate generalized Pareto distributions, is described in Balkema and Embrechts (2004, [2]). Extreme value distributions (EVDs) are also often used as models for threshold exceedances due to a point process limiting result, see for example Coles and Tawn (1991, [10]), Joe et al. (1992, [43]) or Coles and Tawn (1994, [11]).

A short list with points which have to be worked on for a better understanding of multivariate GPDs is given in the introduction of Tajvidi (1996, [71]). Among other things, the need for simulations and methods for parametric estimation is mentioned. The computation of the asymptotic distributions of these estimators and the investigation of alternative definitions to the one proposed there, are suggested as well. These points briefly describe the program of this text.

In Chapter 2 we begin with the definitions and notations necessary for the understanding of the present text. We will also give the most important examples of multivariate GPDs which will be used throughout this manuscript.

Chapter 3 will be dedicated to the simulation of multivariate GPDs. In many complex practical situations, problems cannot be solved by analytical methods. Then one has to rely on Monte-Carlo methods which use simulations. The importance of simulation techniques which provide multivariate GPD random vectors is, therefore, obvious if one wants to do Monte-Carlo simulations in GPD models. An application could be the computation of the Value-at-Risk of a portfolio by Monte-Carlo methods as suggested in Section 7 of Brommundt (2003, [5]). The elliptical distributions used there could reasonably be changed to GPDs, since one is interested in the tails of the corresponding distributions. Real portfolios consist of several hundred or even a few thousand stocks, thus, it is clear that simulation in high dimensions is crucial.

The simulation methods presented here can also be used for a first check of new statistical testing or estimation procedures in multivariate GPD models, before applying them to real data. This is important, since in simulated data sets one controls the entire setup, whereas the background of real data is commonly unknown. This will be a frequent application of the simulation algorithms throughout this manuscript.

There has been relatively little work concerned with the simulation of multivariate extreme value distributions and, except Tajvidi (1996, [71]) where only the need for simulations is stated, and Proposition 6.3.6 of Falk et al. (2004, [21]), where the bivariate Marshall-Olkin GPD is simulated, there is no publication known to the author referring to the simulation of multivariate GPDs. An up-to-date summary of the work done on the simulation of multivariate EVDs can be found in Stephenson (2003, [70]). An additional reference not noted there, are Sections 9.1 and 9.2 of Reiss and Thomas (2001, [57]) where simulation techniques for the bivariate Marshall-Olkin, Hüsler-Reiss and the logistic extreme value distributions are mentioned. Most of the simulation methods for extreme value distributions known today deal with the bivariate or the trivariate case like in Section 3 of Ghoudi et al. (1998, [32]). Stephenson (2003, [70]), where symmetric and asymmetric extreme value distributions of logistic type are simulated, is the only source known to the author for a simulation algorithm in general dimension.

The main simulation algorithm presented here will use the Shi transformation and the Shi coordinates, which will turn out to be the crucial tools of the simulation algorithm. This transformation was first introduced by Shi (1995, [64]) and was used in Stephenson (2003, [70]) for

the simulation of multivariate EVDs of logistic type. As the simulation algorithm in Stephenson (2003, [70]), our algorithm simulating GPDs of logistic type will be formulated for arbitrary dimension and will also be computationally applicable in high dimensions.

In the sequel we will present another algorithm simulating GPDs comparable to the one simulating EVDs in Ghoudi et al. (1998, [32]). Another related algorithm simulating EVDs is the one from Section 2 of Nadarajah (1999, [51]), except that we will simulate from a direct representation, and not a limiting one as it is done there. Our algorithm will also be formulated for arbitrary dimension, although, due to computational reasons, it can only be used in low dimensions. However, this algorithm has the ability to simulate distributions outside of the logistic case.

In Chapters 4, 5 and 6 we will turn to the estimation in GPD models, both in a nonparametric and a parametric way. Publications known to the author concerning estimation in GPD models are again Paper B of Tajvidi (1996, [71]), Section 5.2 of Falk et al. (2004, [21]) and Section 10.2 of Reiss and Thomas (2001, [57]). In the first reference, estimation is done by maximum likelihood methods for a few restricted parametric models in the bivariate case. We will do more detailed investigations of ML methods in GPD models, also for higher dimensions, in Chapter 6. In the second reference above a nonparametric estimator of the Pickands dependence function is presented for the bivariate case, and in the third the canonical dependence function is estimated. Our nonparametric estimation methods in Chapters 4 and 5 will be aimed at the angular (or spectral) density, which is in the bivariate case the second derivative of the Pickands dependence function, see Section 9.3 of Beirlant et al. (2005, [3]) or Remark 2.2 in Falk and Reiss (2003, [25]). The reason for our concentration on the angular density is given below.

There is a broad variety of literature dealing with the estimation in EVD models. Most of it is concerned with finding nonparametric estimators of the Pickands dependence function. Starting with Pickands (1981, [53]) a lot of estimators for the bivariate extreme value case have been found by Deheuvels and Martynov (1996, [14]), Deheuvels (1991, [13]), Deheuvels and Tiago de Oliveira (1989, [15]), Smith et al. (1990, [68]), Hall and Tajvidi (2000, [34]), Capéraà et al. (1997, [6]), Capéraà and Fougères (2000, [7]), Genest et al. (1995, [31]), Abdous et al. (1999, [1]), Jiménez et al. (2001, [41]) and Heffernan (2004, [36]), to give a short list. If these estimators are smooth enough one can, in the bivariate case, derive estimators for the angular density of them by just twice differentiating, see above.

In Einmahl et al. (1997, [19]) and Einmahl et al. (2001, [18]) the angular distribution of observations, which come from the domain of attraction of a bivariate extreme value distribution, is estimated. Drees and Huang (1998, [16]) are concerned with the estimation of the exponent measure and optimal rates of convergence there, likewise in the domain of attraction of an extreme value distribution. Drees and Huang (1995, [17]) also give upper bounds for the rates of convergence of estimators of the angular distribution.

Most of the references cited above deal only with the bivariate setup which is a general tendency in the extreme value literature, see Section 9.4 of Beirlant et al. (2005, [3]). This is only natural, since it is often the case in extreme value theory that things tend to get complicated when going from the bivariate to the trivariate case. That is due to the fact that the Pickands dependence function, which governs these models, is a univariate function in the bivariate case. Therefore, the step from dimension 2 to 3 is, for the dependence function, the step from dimension 1 to 2. This leads to more complicated formulas, and not every assertion valid in the bivariate case carries over to the trivariate case. Some new examples of this phenomenon in GPD models will be shown in Theorem 2.3.12, Corollary 2.3.19 and Theorem 4.2.1. However, higher dimensional

cases are naturally of great practical importance, as stated for example in Tawn (1988, [72]).

Parametric estimators for the extreme value case (also in higher dimensions) are presented in Tawn (1988, [72]), Coles and Tawn (1991, [10]), Joe et al. (1992, [43]), Coles and Tawn (1994, [11]), Coles et al. (1999, [9]) and Section 3.6 of Kotz and Nadarajah (2000, [47]), along with nonparametric estimators of the angular density. We want to carry them over to the GPD case, since the angular density is a useful tool in the visual investigation of the tail dependence structure. The concept of tail dependence is a very useful one to describe dependencies in extreme value theory, see for example Section 6.1 of Falk et al. (2004, [21]) or Section 8.4 of Coles (2001, [8]). We will use kernel methods for the estimation of the angular density which also have been successfully applied in Tawn (1988, [72]), Joe et al. (1992, [43]), Coles et al. (1999, [9]) or Abdous et al. (1999, [1]).

Chapter 4 starts with the nonparametric estimation in bivariate case. The estimation of the angular density is possible there with random variables which follow, after suitable scaling, exactly the angular distribution.

The problem in going from dimension 2 to 3 mentioned before, leads to a slightly different approach in the estimation of the angular density for the general multivariate case in Chapter 5. Here it can only be estimated by random variables which follow the angular distribution asymptotically.

In Chapter 6 we will introduce parametric methods for the estimation in GPD models. We will use maximum likelihood (ML) approaches as in Coles (2001, [8]), Coles and Tawn (1991, [10]), Joe et al. (1992, [43]), Coles and Tawn (1994, [11]) or Coles et al. (1999, [9]). We will also give an approach via relative frequencies inspired by Falk (1998, [20]). Tawn (1988, [72]) shows that the ML estimators in the extreme value case tend to behave badly when close to the case of independence. We will be confronted with that in the GPD case, too. The estimation via relative frequencies will have its strength there. However, it will be inferior to the ML methods in cases of high dependence.

For all estimators the asymptotic normality will be shown under suitable regularity conditions. We develop the asymptotic normality under the assumption that the underlying random vectors follow exactly a GPD. This is a common approach in the univariate case, see Section 5.6 of Beirlant et al. (2005, [3]). In practice one will naturally encounter observations, which are not exactly but only asymptotically distributed by a GPD as a result of the POT approach. In the univariate case this concept is described by the well known δ-neighborhoods, see for example Section 2.2 of Falk et al. (2004, [21]). A suggestion for a multivariate analogue of these δ-neighborhoods can be found in Kaufmann and Reiss (1995, [45]) or Sections 5.3 and 5.4 of Falk et al. (2004, [21]). Another alternative will be presented in Section 5.3 of this manuscript.

Finally in Chapter 7 we will apply the presented methods to a real hydrological data set containing water discharges of the rivers Altmühl and Danube in southern Bavaria, Germany. We will investigate whether extreme water levels tend to appear together alongside these rivers. Investigations like this could lead to a better understanding of the dynamics of extreme floods, which have recently occurred quite often in this area, causing damages amounting to several million euros.

With the final remarks in Chapter 8 we will conclude the present text.

Chapter 2

Multivariate Generalized Pareto Distributions

This chapter intends to clarify definitions and notations which will be used throughout this manuscript. The most important one is of course the definition of the multivariate generalized Pareto distribution. The definition used here originates from Section 5.1 of Falk et al. (2004, [21]) and was first introduced by Kaufmann and Reiss (1995, [45]) for the bivariate case. An alternative definition can be found in Paper B of Tajvidi (1996, [71]) and in Section 8.3 of Beirlant et al. (2005, [3]). The second one is discussed in more detail in Rootzén and Tajvidi (2005, [60]). Both are closely related to each other and are identical to our definition in the region of interest, where all components of a random vector are large, see Section 10.1 of Reiss and Thomas (2001, [57]) and Section 8.3 of Beirlant et al. (2005, [3]). They differ from each other and from our definition in areas, where only some components are large. While our definition does not make assumptions on the structure of the generalized Pareto distribution for areas, where not all components are large, both other definitions make specific assumptions on these areas.

Section 2.1 will begin by introducing the Pickands coordinates, which are of great importance to the field of extreme value theory and which we will frequently use. For the role of Pickands coordinates in the entire extreme value framework we refer to Section 5 of Falk et al. (2004, [21]).

In Section 2.2 we give the definitions of multivariate generalized Pareto and extreme value distributions. Alongside we will introduce other important objects belonging to GPDs and EVDs which are necessary for their investigation.

The most important examples of GPDs, namely the logistic case and various generalizations of it, will be presented in Section 2.3 together with some theoretical results on these distributions. These include a generalization of a known counterexample, showing that GP functions are not necessarily distribution functions, to arbitrary dimension $d \geq 3$ and the non-uniqueness of the angular measure for GPDs.

2.1 Pickands Coordinates

Definition 2.1.1
For $d \in \mathbb{N}$ the set

$$R_d := \left\{ x = (x_1, \ldots, x_d) \in (0,\infty)^d \,\middle|\, \sum_{i=1}^{d} x_i < 1 \right\}$$

defines the (open) *unit simplex* in $\mathbb{R}^d$. By $\overline{R_d}$ we denote the closed unit simplex with the representation

$$\overline{R_d} = \left\{ x \in [0,\infty)^d \,\middle|\, \sum_{i=1}^{d} x_i \leq 1 \right\}.$$

$\Diamond$

The unit simplex R_d will play an important role in the following. We will need especially the d–dimensional volume of R_d at several occasions.

Lemma 2.1.2
We have

$$\mathrm{vol}(R_{c,d}) = \int_{R_{c,d}} 1\, dx = \frac{c^d}{d!}, \qquad d \in \mathbb{N},$$

and especially

$$\mathrm{vol}(R_d) = \mathrm{vol}\left(\overline{R_d}\right) = \frac{1}{d!},$$

where $c > 0$ and

$$R_{c,d} := \left\{ (x_1, \ldots, x_d) \in (0,\infty)^d \,\middle|\, \sum_{i=1}^{d} x_i < c \right\}.$$

Proof:
We will show the assertion by induction. We have for the initial step $d = 1$

$$R_{c,1} = (0,c) \Rightarrow \int_{R_{c,1}} 1\, dx = \int_{(0,c)} 1\, dx = c = \frac{c^1}{1!}.$$

Let the induction hypothesis hold for $d-1$. We will show next the induction step $d-1 \to d$.

We have by Fubini's Theorem (see for example Fristedt and Gray (1997, [28]), Section 9.2), where $\mathbf{1}_A$ denotes the indicator function of a set A, i.e., $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ else,

$$\begin{aligned}
\mathrm{vol}(R_{c,d}) &= \int_{\mathbb{R}^d} \mathbf{1}_{R_{c,d}}(x_1, \ldots, x_d)\, dx_1 \cdots dx_d \\
&= \int_0^c \int_{\mathbb{R}^{d-1}} \mathbf{1}_{R_{c-x_d,d-1}}(x_1, \ldots, x_{d-1})\, dx_1 \cdots dx_{d-1}\, dx_d \\
&= \int_0^c \mathrm{vol}(R_{c-x_d,d-1})\, dx_d = \int_0^c \frac{(c-x_d)^{d-1}}{(d-1)!}\, dx_d = \frac{1}{(d-1)!}\left[-\frac{(c-x_d)^d}{d} \right]_0^c = \frac{c^d}{d!}.
\end{aligned}$$

We conclude now $\mathrm{vol}(R_d) = \mathrm{vol}(R_{1,d}) = \frac{1}{d!} = \mathrm{vol}\left(\overline{R_d}\right)$ since the boundary set ∂R_d is a null set with regard to the d–dimensional Lebesgue measure.

$\square$

The next two related Pickands transformations will be crucial for large parts of this manuscript.

Definition 2.1.3
For $d \in \mathbb{N}$, $d \geq 2$, define the transformation $T_P : (-\infty, 0]^d \backslash \{0\} \to \overline{R_{d-1}} \times (-\infty, 0)$ by

$$T_P(x) := \left(\frac{x_1}{x_1 + \ldots + x_d}, \ldots, \frac{x_{d-1}}{x_1 + \ldots + x_d}, x_1 + \ldots + x_d \right) =: (z_1, \ldots, z_{d-1}, c). \tag{2.1}$$

T_P is called *transformation to (standard) Pickands coordinates* $z := (z_1, \ldots, z_{d-1}), c$. z is called the *angular component* and c is called the *radial component*.

$\lozenge$

For information on important properties of T_P, see Section 5.4 in Falk et al. (2004, [21]) and Falk and Reiss (2005, [27]). The mapping T_P is one-to-one with inverse function

$$T_P^{-1}(z, c) = c \left(z_1, \ldots, z_{d-1}, 1 - \sum_{i=1}^{d-1} z_i \right). \tag{2.2}$$

It has the same geometrical interpretation as standard polar coordinates with the difference that polar coordinates use the euclidian norm $|| \cdot ||_2$ for the angular and radial component, whereas Pickands coordinates use the sum norm $|| \cdot ||_1$.

Definition 2.1.4
For $d \in \mathbb{N}$, $d \geq 2$, define the transformation $T_F : (-\infty, 0)^d \to R_{d-1} \times (-\infty, 0)$ by

$$T_F(x) := \left(\frac{\frac{1}{x_1}}{\frac{1}{x_1} + \ldots + \frac{1}{x_d}}, \ldots, \frac{\frac{1}{x_{d-1}}}{\frac{1}{x_1} + \ldots + \frac{1}{x_d}}, \frac{1}{x_1} + \ldots + \frac{1}{x_d} \right) =: (z_1, \ldots, z_{d-1}, c). \tag{2.3}$$

T_F is called *transformation to Pickands coordinates* $z := (z_1, \ldots, z_{d-1})$, c *with regard to Fréchet margins*. z is again called the *angular component* and c the *radial component*.

$\lozenge$

The transformation (2.3) is, as its name already indicates, closely related to the Pickands transformation T_P from Definition 2.1.3. In addition we have applied here the transformation $y \mapsto \frac{1}{y}$ to the components of x. This is the transformation which conveys exponentially distributed random variables to Fréchet distributed random variables. Therefore, the choice of the name of T_F. The mapping T_F is one-to-one with inverse function

$$T_F^{-1}(z, c) = \frac{1}{c} \left(\frac{1}{z_1}, \ldots, \frac{1}{z_{d-1}}, \frac{1}{1 - \sum_{i=1}^{d-1} z_i} \right). \tag{2.4}$$

Since we are going to apply the density transformation theorem it will be important to know the determinant of the Jacobian matrix of the inverse of T_F, which we give in the next lemma.

Lemma 2.1.5
The determinant of the Jacobian matrix of the transformation T_F^{-1} is

$$\det \left(J_{T_F^{-1}}(z, c) \right) = -\frac{1}{c^{d+1}} \cdot \frac{1}{z_1^2 \cdot z_2^2 \cdot \ldots \cdot z_{d-1}^2 \left(1 - \sum_{i=1}^{d-1} z_i \right)^2}.$$

Proof:
As one easily checks we have for the Jacobian matrix

$$J_{T_F^{-1}}(z,c) = \begin{pmatrix} \frac{\partial x_1}{\partial c} & \frac{\partial x_1}{\partial z_1} & \frac{\partial x_1}{\partial z_2} & \cdots & \frac{\partial x_1}{\partial z_{d-2}} & \frac{\partial x_1}{\partial z_{d-1}} \\ \frac{\partial x_2}{\partial c} & \frac{\partial x_2}{\partial z_1} & \frac{\partial x_2}{\partial z_2} & \cdots & \frac{\partial x_2}{\partial z_{d-2}} & \frac{\partial x_2}{\partial z_{d-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial x_{d-2}}{\partial c} & \frac{\partial x_{d-2}}{\partial z_1} & \frac{\partial x_{d-2}}{\partial z_2} & \cdots & \frac{\partial x_{d-2}}{\partial z_{d-2}} & \frac{\partial x_{d-2}}{\partial z_{d-1}} \\ \frac{\partial x_{d-1}}{\partial c} & \frac{\partial x_{d-1}}{\partial z_1} & \frac{\partial x_{d-1}}{\partial z_2} & \cdots & \frac{\partial x_{d-1}}{\partial z_{d-2}} & \frac{\partial x_{d-1}}{\partial z_{d-1}} \\ \frac{\partial x_d}{\partial c} & \frac{\partial x_d}{\partial z_1} & \frac{\partial x_d}{\partial z_2} & \cdots & \frac{\partial x_d}{\partial z_{d-2}} & \frac{\partial x_d}{\partial z_{d-1}} \end{pmatrix}$$

of T_F^{-1} the representation

$$J_{T_F^{-1}}(z,c) = \begin{pmatrix} \frac{1}{z_1}\left(-\frac{1}{c^2}\right) & \frac{1}{c}\left(-\frac{1}{z_1^2}\right) & 0 & \cdots & 0 & 0 \\ \frac{1}{z_2}\left(-\frac{1}{c^2}\right) & 0 & \frac{1}{c}\left(-\frac{1}{z_2^2}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{z_{d-2}}\left(-\frac{1}{c^2}\right) & 0 & 0 & \cdots & \frac{1}{c}\left(-\frac{1}{z_{d-2}^2}\right) & 0 \\ \frac{1}{z_{d-1}}\left(-\frac{1}{c^2}\right) & 0 & 0 & \cdots & 0 & \frac{1}{c}\left(-\frac{1}{z_{d-1}^2}\right) \\ \frac{1}{1-\sum_{i=1}^{d-1} z_i}\left(-\frac{1}{c^2}\right) & \frac{1}{c}\frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \frac{1}{c}\frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \cdots & \frac{1}{c}\frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \frac{1}{c}\frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} \end{pmatrix}.$$

We compute the determinant with the known calculation rules for the determinant (see, for example, Chapter VI, §2 in Lang (1966, [48])). In the first step we factor out $\frac{1}{c}$ from the last $d-1$ columns, as well as $-\frac{1}{c^2}$ once from the first column and, thus, obtain

$$\begin{aligned} \det\left(J_{T_F^{-1}}(z,c)\right) &= -\frac{1}{c^{d+1}} \cdot \det \begin{pmatrix} \frac{1}{z_1} & -\frac{1}{z_1^2} & 0 & \cdots & 0 & 0 \\ \frac{1}{z_2} & 0 & -\frac{1}{z_2^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{z_{d-2}} & 0 & 0 & \cdots & -\frac{1}{z_{d-2}^2} & 0 \\ \frac{1}{z_{d-1}} & 0 & 0 & \cdots & 0 & -\frac{1}{z_{d-1}^2} \\ \frac{1}{1-\sum_{i=1}^{d-1} z_i} & \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \cdots & \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} \end{pmatrix} \\ &= -\frac{1}{c^{d+1} z_1} \cdot \det \begin{pmatrix} 0 & -\frac{1}{z_2^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\frac{1}{z_{d-2}^2} & 0 \\ 0 & 0 & \cdots & 0 & -\frac{1}{z_{d-1}^2} \\ \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \cdots & \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} & \frac{1}{\left(1-\sum_{i=1}^{d-1} z_i\right)^2} \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{c^{d+1}z_1^2}\cdot\det\begin{pmatrix} \frac{1}{z_2} & -\frac{1}{z_2^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{z_{d-2}} & 0 & \cdots & -\frac{1}{z_{d-2}^2} & 0 \\ \frac{1}{z_{d-1}} & 0 & \cdots & 0 & -\frac{1}{z_{d-1}^2} \\ \frac{1}{1-\sum_{i=1}^{d-1}z_i} & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} & \cdots & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} \end{pmatrix} \\
= & -\frac{1}{c^{d+1}\cdot z_1\cdot z_2^2\cdot\ldots\cdot z_{d-1}^2\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)^2}\cdot\underbrace{(-1)^{d-2}}_{\text{column changes}}\cdot\underbrace{(-1)^{d-2}}_{\text{negative signs}} \\
& -\frac{1}{c^{d+1}z_1^2z_2}\cdot\det\begin{pmatrix} 0 & -\frac{1}{z_3^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{1}{z_{d-1}^2} \\ \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} & \cdots & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} \end{pmatrix} \\
& -\frac{1}{c^{d+1}z_1^2z_2^2}\cdot\det\begin{pmatrix} \frac{1}{z_3} & -\frac{1}{z_3^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{z_{d-1}} & 0 & \cdots & -\frac{1}{z_{d-1}^2} \\ \frac{1}{1-\sum_{i=1}^{d-1}z_i} & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} & \cdots & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} \end{pmatrix} \\
= & -\frac{1}{c^{d+1}\cdot z_1\cdot z_2^2\cdot\ldots\cdot z_{d-1}^2\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)^2}-\frac{1}{c^{d+1}\cdot z_1^2\cdot z_2\cdot z_3^2\cdot\ldots\cdot z_{d-1}^2\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)^2} \\
& -\frac{1}{c^{d+1}z_1^2z_2^2}\det\begin{pmatrix} \frac{1}{z_3} & -\frac{1}{z_3^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{z_{d-1}} & 0 & \cdots & -\frac{1}{z_{d-1}^2} \\ \frac{1}{1-\sum_{i=1}^{d-1}z_i} & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} & \cdots & \frac{1}{\left(1-\sum_{i=1}^{d-1}z_i\right)^2} \end{pmatrix} \\
= & \ldots \\
= & -\frac{1}{c^{d+1}}\left(\frac{1}{z_1\cdot z_2^2\cdot z_3^2\cdot\ldots\cdot z_{d-2}^2\cdot z_{d-1}^2\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)^2}\right. \\
& \quad +\frac{1}{z_1^2\cdot z_2\cdot z_3^2\cdot\ldots\cdot z_{d-2}^2\cdot z_{d-1}^2\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)^2} \\
& \quad +\ldots \\
& \quad +\frac{1}{z_1^2\cdot z_2^2\cdot z_3^2\cdot\ldots\cdot z_{d-2}^2\cdot z_{d-1}\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)^2} \\
& \quad \left.+\frac{1}{z_1^2\cdot z_2^2\cdot z_3^2\cdot\ldots\cdot z_{d-2}^2\cdot z_{d-1}^2\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)}\right) \\
= & -\frac{1}{c^{d+1}}\cdot\frac{z_1+\ldots+z_{d-1}+1-\sum_{i=1}^{d-1}z_i}{z_1^2\cdot z_2^2\cdot z_3^2\cdot\ldots\cdot z_{d-2}^2\cdot z_{d-1}^2\cdot\left(1-\sum_{i=1}^{d-1}z_i\right)^2}
\end{aligned}
$$

$$= -\frac{1}{c^{d+1}} \cdot \frac{1}{z_1^2 \cdot z_2^2 \cdot z_3^2 \cdot \ldots \cdot z_{d-2}^2 \cdot z_{d-1}^2 \cdot \left(1 - \sum_{i=1}^{d-1} z_i\right)^2},$$

as asserted.

□

2.2 Multivariate Generalized Pareto and Extreme Value Distributions

In this section we will introduce multivariate extreme value and generalized Pareto distributions. We assume that the reader is familiar with the family of univariate extreme value and generalized Pareto distributions. We refer to Sections 1.3 and 1.4 of Reiss and Thomas (2001, [57]) for a summary on these distributions.

Definition 2.2.1
An *extreme value distribution (EVD) with negative exponential margins* is defined by the distribution function

$$G(x_1, \ldots, x_d) = \exp\left(\left(\sum_{i=1}^{d} x_i\right) D\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, \ldots, \frac{x_{d-1}}{\sum_{i=1}^{d} x_i}\right)\right)$$

for $x_1, \ldots, x_d < 0$. Thereby the function $D : \overline{R_{d-1}} \to [0,1]$ is called *Pickands dependence function* and has to satisfy (2.6) and (2.7) below.

Let $G_1, \ldots, G_d$ be univariate extreme value distributions. Then the transformation of an EVD with negative exponential margins to the margins $G_1, \ldots, G_d$, i.e.,

$$G_*(x_1, \ldots, x_d) := G\left(\log\left(G_1(x_1)\right), \ldots, \log\left(G_d(x_d)\right)\right)$$

where $x_i \in \text{supp}(G_i) := \overline{\{x \in \mathbb{R} | G_i(x) > 0\}}$, the support of the distribution G_i, is an *extreme value distribution* G_*.

◊

It is well known that multivariate extreme value distributions arise, like in the univariate case, as the limiting distributions of suitably scaled componentwise maxima of independent and identically distributed random vectors, see for example Section 5.4 of Resnick (1987, [59]). A distribution F is said to be in the *domain of attraction* of an EVD G_* if the distribution of suitably scaled componentwise maxima of independent copies of random vectors, which follow F, converges to G_*. More precisely, if for independent $X_1, \ldots, X_n$ following F there exist vectors a_n, $b_n \in \mathbb{R}^d$, $a_n > 0$, such that

$$P\left(\frac{\max_{i=1,\ldots,n} X_i - b_n}{a_n}\right) = F^n\left(a_n x + b_n\right) \to_{n\to\infty} G_*(x),$$

where all operations (including the maximum) and inequalities are meant componentwise, F is in the domain of attraction of G_*. We note this by $F \in D(G_*)$. We refer to Section 5.4 of Resnick (1987, [59]) for more mathematical details.

Definition 2.2.2
Let $X = (X_1, \ldots, X_d)$ be a random vector, which has a distribution function W with the representation

$$\begin{aligned} W(x) &= 1 + \left(\sum_{i=1}^{d} x_i\right) D\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, \ldots, \frac{x_{d-1}}{\sum_{i=1}^{d} x_i}\right) \\ &= 1 + \log\left(G\left(x_1, \ldots, x_d\right)\right), \qquad (x_1, \ldots, x_d) = x \in U \end{aligned} \tag{2.5}$$

in a neighborhood U of 0 in the relative topology of the negative quadrant $(-\infty, 0)^d$. Then X follows a *generalized Pareto distribution (GPD) with uniform margins*. D is thereby again a Pickands dependence function fulfilling (2.6) and (2.7) below.

Let G_* be again an extreme value distribution with right upper end point $u = (u_1, \ldots, u_d) = (\sup\{x \in \mathbb{R} | G_i(x) < 1\})_{i=1,\ldots,d}$, and let $X = (X_1, \ldots, X_d)$ be a random vector, which has a distribution function W_* with the representation

$$W_*(x) = 1 + \log\left(G_*\left(x\right)\right), \qquad x \in U,$$

in a neighborhood U of u in the relative topology of the quadrant $(-\infty, u_1) \times \ldots \times (-\infty, u_d)$. Then X follows a *generalized Pareto distribution*.

$\Diamond$

These definitions of GPDs are in analogy to the relation between extreme value and generalized Pareto distributions in the univariate case, where also $W_* = 1 + \log G_*$ holds, see Section 1.4 of Reiss and Thomas (2001, [57]).

Remark 2.2.3
The marginal distributions of a GPD W_* are univariate GPDs close to the right upper endpoint of the distribution. By a suitable marginal transformation we can always transfer a GPD into a GPD with uniform margins. Unless otherwise stated at special occasions, we will in the course of this manuscript always consider GPDs W with uniform margins and correspondingly EVDs with negative exponential margins.

The GP function $W = 1 + \log G$ is not a distribution function on the entire set of the negative quadrant with $W(x) \geq 0$, see Section 5.1 in Falk et al. (2004, [21]) for a counterexample with dimension $d = 3$. We will give a generalized version of this counterexample for arbitrary dimension $d \geq 3$ at the end of Section 2.3.2. However, in a neighborhood of 0 the function $W(x)$ coincides with a distribution function, see Lemma 5.1.3 in Falk et al. (2004, [21]). By suitable marginal transformations it follows that $W_*(x)$ is also a distribution function close to its upper endpoint u. Actually in the bivariate case it turns out that a GP function W is a distribution function on the entire set where $W(x) \geq 0$ (see Lemma 5.1.1 in Falk et al. (2004, [21])) but not in the higher multivariate cases. This is an example, where a property of a bivariate GPD cannot be transferred to higher dimensions. We will see other examples for this strange behavior in Sections 2.3.4 and 4.2. In contrast to the generalized Pareto function W, a multivariate extreme value distribution G is always a distribution function on the entire negative quadrant.

We have a flexible definition of a GPD, since we do not make assumptions on the structure of the GPD away from the origin. The two other definitions from Tajvidi (1996, [71]), Section 8.3 of Beirlant et al. (2005, [3]) and Rootzén and Tajvidi (2005, [60]) are in the extreme area, where all components are large, according to Section 10.1 of Reiss and Thomas (2001, [57]) and Section 8.3 of Beirlant et al. (2005, [3]) identical to ours. For the definition in Tajvidi (1996, [71])

components of observations, which fall short of given threshold lines, are set to the value of the threshold. Thus, the threshold lines which are null sets with regard to the Lebesgue measure have positive probability, and the support of the distribution is limited. Also the univariate margins are not univariate GPDs. The definition from Section 8.3 of Beirlant et al. (2005, [3]) specifies the GPD distribution as

$$W(x) = \frac{1}{-\log G(0)} \cdot \log \frac{G(x)}{G(\min(x,0))},$$

where the minimum (and the following inequalities) are to be taken componentwise. G is the corresponding extreme value distribution. This leads to $W(x) = 0$ for $x < 0$ and $W(x) = 1 - \frac{\log G(x)}{\log G(0)}$ for $x > 0$, which is our definition in the extreme area $x > 0$ after a suitable marginal transformation. Actually this representation was used in Lemma 5.1.3 of Falk et al. (2004, [21]) to show that a GP function close to the origin can be extended to a valid distribution function on the entire negative quadrant. From the above definition the structure of W in areas, where only some components are large can also be gained. Remark that due to the minimum in the definition, the distribution W might not be differentiable at the axes of the coordinate system. ◊

Before we come to the justification of the definition of a GPD we will consider the structure of a GPD a little closer.

A listing of important properties and characterizations of GPDs and especially of the Pickands dependence function D from Definitions 2.2.1 and 2.2.2 can be found in Section 4.3 of Falk et al. (2004, [21]). We only give here a short summary of those properties and characterizations, which we will need in the present text.

The Pickands dependence function can be written as

$$D(t_1, \ldots, t_{d-1}) = \int_{\overline{R_{d-1}}} \max\left(u_1 t_1, \ldots, u_{d-1} t_{d-1}, \left(1 - \sum_{i=1}^{d-1} u_i\right)\left(1 - \sum_{i=1}^{d-1} t_i\right)\right) \nu(du), \quad (2.6)$$

where ν is a measure on $\overline{R_{d-1}}$ with

$$\nu\left(\overline{R_{d-1}}\right) = d \qquad \text{and} \qquad \int_{\overline{R_{d-1}}} u_i \nu(du) = 1, \quad 1 \le i \le d-1, \quad (2.7)$$

see again for example Section 4.3 of Falk et al. (2004, [21]). ν is called the *angular measure*. The characterization of this measure by (2.7) is necessary and sufficient to define a proper Pickands dependence function, see Theorem 4.3.1 in Falk et al. (2004, [21]).

The Pickands dependence function D fulfills

$$\max\left(t_1, \ldots, t_{d-1}, 1 - \sum_{i=1}^{d-1} t_i\right) \le D\left(t_1, \ldots, t_{d-1}\right) \le 1.$$

Both bounds are Pickands dependence functions themselves, the so called cases of *complete dependence* and *independence*, see Section 2.3.1.

By

$$V(x_1, \ldots, x_d) := \left(\sum_{i=1}^{d} x_i\right) D\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, \ldots, \frac{x_{d-1}}{\sum_{i=1}^{d} x_i}\right) \quad (2.8)$$

we denote the *exponent measure function* of W. By (2.6) this exponent measure function can be represented as

$$V(x_1,\ldots,x_d) = \int_{\overline{R_{d-1}}} \min\left(u_1x_1,\ldots,u_{d-1}x_{d-1},\left(1-\sum_{i=1}^{d-1}u_i\right)x_d\right)\nu(du),$$

see again Section 4.3 of Falk et al. (2004, [21]) or Section 3.2 of Kotz and Nadarajah (2000, [47]) for the case of Fréchet margins.

A norm on $\mathbb{R}^d$ is defined by

$$||x||_D := V\left(|x_1|,\ldots,|x_d|\right),$$

see also Section 4.3 in Falk et al. (2004, [21]) and Hofmann (2006, [39]) for further details. Thus, every generalized Pareto distribution can be written as

$$W(x_1,\ldots,x_d) = 1 - ||x||_D, \tag{2.9}$$

$x \in (-\infty,0)^d$, close to 0.

The distribution function

$$L(z_1,\ldots,z_{d-1}) = \nu\left([0,z_1]\times\cdots\times[0,z_{d-1}]\right)$$

of the measure ν is called *angular distribution*. By

$$d^* := \nu\left(R_{d-1}\right) \tag{2.10}$$

we denote the mass of ν in the interior of $\overline{R_{d-1}}$; recall that by (2.7) we have $\nu\left(\overline{R_{d-1}}\right) = d$ and, thus,

$$0 \le d^* \le d.$$

If the measure ν, restricted to R_{d-1}, possesses a density we denote it with l and call it the *angular density*. We will see in later chapters that under certain regularity conditions the angular components of the Pickands coordinates of GPD distributed random vectors follow asymptotically the angular distribution, thus our choice of the name. In the literature it is also common to call the angular measure/distribution/density the *spectral* measure/distribution/density, see for example Einmahl et al. (1997, [19]) or Einmahl et al. (2001, [18]).

From (2.10) we especially obtain

$$\int_{R_{d-1}} l(z)\,dz = d^*. \tag{2.11}$$

Extreme value distributions arise both in the univariate and the multivariate case as the limiting distributions of componentwise maxima. GPDs are used for peaks-over-threshold (POT) methods in the univariate case. We will give a short justification that multivariate GPDs can also be used in multivariate POT models.

Let a d-dimensional random vector follow a distribution function F in the multivariate domain of attraction of an extreme value distribution G with margins $G_1,\ldots,G_d$ not necessarily negative exponential. Then we know by Proposition 5.15 (b) in Resnick (1987, [59]) that the marginal distributions $F_1,\ldots,F_d$ of F are in the univariate domains of attraction of the extreme value distributions $G_1,\ldots,G_d$, i.e, $F_i \in D(G_i)$, $i = 1,\ldots,d$. Define by

$$G_*(x_1,\ldots,x_d) = G\left(G_1^{-1}\left(\exp\left(-\frac{1}{x_1}\right)\right),\ldots,G_d^{-1}\left(\exp\left(-\frac{1}{x_d}\right)\right)\right)$$

the corresponding extreme value distribution with standard Fréchet margins and by

$$F_*(x_1,\dots,x_d) = F\left(F_1^{-1}\left(1-\frac{1}{x_1}\right),\dots,F_d^{-1}\left(1-\frac{1}{x_d}\right)\right)$$

the transformation of F to standard Pareto margins. Denote by

$$X_* = \left(X_*^{(1)},\dots,X_*^{(d)}\right) = \left(\frac{1}{1-F_1(X_1)},\dots,\frac{1}{1-F_d(X_d)}\right)$$

the correspondingly transformed vector X. Then we know by Proposition 5.15 (a) in Resnick (1987, [59]) that for $x=(x_1,\dots,x_d)$ with $x_i>0$, $i=1,\dots,d$, the following equation holds with the notation $\mathbf{1}=(1,\dots,1)\in\mathbb{R}^d$:

$$\lim_{t\to\infty}\frac{1-F_*(tx)}{1-F_*(t\mathbf{1})} = \frac{-\log G_*(x)}{-\log G_*(\mathbf{1})}. \tag{2.12}$$

If we restrict ourselves to $x_i>1$, $i=1,\dots,d$, we can express the above equation (2.12) in terms of conditional probabilities for X_* with the convention that inequalities for vectors hold componentwise and, thus, $X_*\not\leq x$ means $X_*^{(i)}>x_i$ for at least one index i:

$$\begin{aligned}
1+\frac{\log G_*(x)}{-\log G_*(\mathbf{1})} &= 1-\lim_{t\to\infty}\frac{1-F_*(tx)}{1-F_*(t\mathbf{1})} = 1-\lim_{t\to\infty}\frac{P\left(X_*\not\leq tx\right)}{P\left(X_*\not\leq t\mathbf{1}\right)}\\
&\overset{x\geq\mathbf{1}}{=} 1-\lim_{t\to\infty}\frac{P\left(X_*\not\leq tx, X_*\not\leq t\mathbf{1}\right)}{P\left(X_*\not\leq t\mathbf{1}\right)}\\
&= 1-\lim_{t\to\infty}P\left(X_*\not\leq tx|X_*\not\leq t\mathbf{1}\right)\\
&= \lim_{t\to\infty}P\left(X_*\leq tx|X_*\not\leq t\mathbf{1}\right).
\end{aligned} \tag{2.13}$$

Thus, we see that under the condition $X_*\not\leq t\mathbf{1}$ the distribution of the exceedances of X_* in the area, where all components exceed the threshold (since $x>\mathbf{1}$), converges to a GPD with Pareto margins with a scaling factor of $-\frac{1}{\log G_*(\mathbf{1})}>0$. After suitable marginal transformations we can thus assume that under the condition $X\not\leq u$ for some threshold u the approximation of the original distribution function F by a GPD with uniform margins is reasonable. Remark that by our definition of a GPD we only have the representation $1+\log G$ for an area around 0, i.e., where all components are large. We do not assume a specific distribution for the areas where only some components are large.

A more detailed version of the result (2.13) showing equivalence to $F\in D(G)$ can be found in Theorems 2.2 and 2.3 of Rootzén and Tajvidi (2005, [60]) for a slightly different version of a multivariate GPD, see Remark 2.2.3. In Section 5.2 of Falk et al. (2004, [21]) it is also shown that multivariate GPDs are, as in the univariate case, POT-stable. Thus altogether the GPDs are natural candidates for POT approaches in the multivariate case.

The result (2.13) was used in equation (2.9) of Smith et al. (1997, [69]) to model exceedances over high thresholds without putting it into a multivariate GPD context, which we will do in this manuscript. A variant of (2.13) leading to an extreme value approximation of exceedances over high thresholds was used in Section 2 of Ledford and Tawn (1996, [49]) to model the corresponding exceedances. Together with a limiting point process result in Coles and Tawn (1991, [10]) and Joe et al. (1992, [43]) threshold exceedances are often assumed to follow extreme value distributions. Statistical estimation in such models is done for example in Coles and Tawn (1994, [11]).

In this manuscript we will model exceedances over high thresholds as coming from GPDs and will give estimation procedures for this model assumption in Chapters 4 to 6.

Extreme value distributions arise as limiting distributions of suitably scaled maxima. Since $\min x = -\max(-x)$, the corresponding limiting distributions $\tilde{G}$ for suitably scaled minima can easily be given as

$$\tilde{G}(x) = \overline{G}(-x),$$

where $\overline{G}(x) = P(X_1 \geq x_1, \ldots, X_d \geq x_d)$ is the survivor function of an extreme value distribution G, see Section 9.1 of Reiss and Thomas (2001, [57]). One can correspondingly define generalized Pareto distributions in this case as

$$\tilde{W}(x) := 1 + \log \tilde{G}(x) = 1 + \log \overline{G}(-x).$$

They are natural candidates for the modelling of the distribution of shortfalls below a certain threshold. To our knowledge these distributions have not been investigated yet.

For more details on the definitions of GPDs and EVDs and their consequences we refer to Falk et al. (2004, [21]), Reiss and Thomas (2001, [57]), Coles (2001, [8]) and Beirlant et al. (2005, [3]).

We will show two important theorems for generalized Pareto distributions next.

Theorem 2.2.4
Let the generalized Pareto distribution W be continuously differentiable of order d. Then the corresponding angular density l fulfills

$$l\left(\frac{\frac{1}{x_1}}{\sum_{i=1}^d \frac{1}{x_i}}, \ldots, \frac{\frac{1}{x_{d-1}}}{\sum_{i=1}^d \frac{1}{x_i}}\right) = \frac{x_1^2 \cdot \ldots \cdot x_d^2}{\left(-\sum_{i=1}^d \frac{1}{x_i}\right)^{-(d+1)}} \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W(x_1, \ldots, x_d).$$

Proof:
For the angular density of an extreme value distribution with Fréchet margins it is known from Theorem 1 in Coles and Tawn (1991, [10]) or Section 3.2 in Kotz and Nadarajah (2000, [47]) that

$$\frac{\partial^d}{\partial y_1 \cdots \partial y_d} \tilde{V}(y_1, \ldots, y_d) = -\left(\sum_{i=1}^d y_i\right)^{-(d+1)} l\left(\frac{y_1}{\sum_{i=1}^d y_i}, \ldots, \frac{y_{d-1}}{\sum_{i=1}^d y_i}\right) \tag{2.14}$$

holds where $\tilde{V}$ is the exponent measure function of the corresponding extreme value distribution with Fréchet margins. Here we have a generalized Pareto distribution W with uniform margins and exponent measure function V. This passage is done with the transformation $y \mapsto -\frac{1}{y}$. We are going to investigate how (2.14) changes with the transformation of the margins.

Let $V(x_1, \ldots, x_d)$ be the exponent measure function with regard to uniform margins as defined in (2.8). Then the GPD has the representation

$$W(x_1, \ldots, x_d) = 1 - V(x_1, \ldots, x_d)$$

in a neighborhood of 0. We have

$$\tilde{V}(y_1, \ldots, y_d) = V\left(-\frac{1}{y_1}, \ldots, -\frac{1}{y_d}\right)$$

and, furthermore, by the usage of the chain rule (Theorem 165.2 in Heuser (1998, [38])),

$$\begin{aligned}\frac{\partial^d}{\partial y_1 \cdots \partial y_d}\tilde{V}(y_1,\ldots,y_d) &= \frac{\partial^d}{\partial y_1 \cdots \partial y_d}\left(V\left(-\frac{1}{y_1},\ldots,-\frac{1}{y_d}\right)\right)\\ &= \frac{1}{y_1^2 \cdot \ldots \cdot y_d^2}\left(\frac{\partial^d}{\partial x_1 \cdots \partial x_d}V\right)\left(-\frac{1}{y_1},\ldots,-\frac{1}{y_d}\right).\end{aligned}$$

This result inserted in (2.14) gives

$$\frac{1}{y_1^2 \cdot \ldots \cdot y_d^2}\left(\frac{\partial^d}{\partial x_1 \cdots \partial x_d}V\right)\left(-\frac{1}{y_1},\ldots,-\frac{1}{y_d}\right) = -\left(\sum_{i=1}^d y_i\right)^{-(d+1)} l\left(\frac{y_1}{\sum_{i=1}^d y_i},\ldots,\frac{y_{d-1}}{\sum_{i=1}^d y_i}\right).$$

By putting now again $x_i = -\frac{1}{y_i}$ for $i = 1,\ldots,d$ one arrives at

$$x_1^2 \cdot \ldots \cdot x_d^2 \cdot \frac{\partial^d}{\partial x_1 \cdots \partial x_d}V(x_1,\ldots,x_d) = -\left(-\sum_{i=1}^d \frac{1}{x_i}\right)^{-(d+1)} l\left(\frac{\frac{1}{x_1}}{\sum_{i=1}^d \frac{1}{x_i}},\ldots,\frac{\frac{1}{x_{d-1}}}{\sum_{i=1}^d \frac{1}{x_i}}\right).$$

Thus, the angular density has the representation

$$l\left(\frac{\frac{1}{x_1}}{\sum_{i=1}^d \frac{1}{x_i}},\ldots,\frac{\frac{1}{x_{d-1}}}{\sum_{i=1}^d \frac{1}{x_i}}\right) = -\frac{x_1^2 \cdot \ldots \cdot x_d^2}{\left(-\sum_{i=1}^d \frac{1}{x_i}\right)^{-(d+1)}}\frac{\partial^d}{\partial x_1 \cdots \partial x_d}V(x_1,\ldots,x_d).$$

Replacing $W = 1 - V$ we see that the assertion holds.

□

The next theorem shows the importance of the standard Pickands coordinates, since they decompose GPD distributed random vectors into two conditionally independent components.

Theorem 2.2.5
Let $(X_1,\ldots,X_d)$ follow a generalized Pareto distribution W, whose Pickands dependence function D has continuous partial derivatives of order d. Let $C := X_1 + \ldots + X_d$ and $Z := (X_1/C,\ldots,X_{d-1}/C)$ be the standard random Pickands coordinates. Put

$$\phi(z,c) := |c|^{d-1}\left(\frac{\partial^d}{\partial x_1 \cdots \partial x_d}W\right)\left(T_P^{-1}(z,c)\right). \tag{2.15}$$

Then ϕ depends only on z and, therefore, we put $\phi(z) := \phi(z,c)$.

If $\mu := \int_{R_{d-1}} \phi(z)\,dz > 0$ we have for $c_0 < 0$ close to 0 the following assertions:

(i) Under the condition $C > c_0$ the components Z and C of the Pickands coordinates are independent.

(ii) C is conditionally on $(c_0, 0)$ uniformly distributed, more precisely $P(C > c) = \mu|c|$ for $c_0 \le c \le 0$ and, thus,

$$P(C \ge c | C > c_0) = \frac{c}{c_0}, \qquad c_0 \le c \le 0.$$

(iii) Conditional on $C > c_0$ the Pickands coordinate Z has the density

$$f(z) = \frac{\phi(z)}{\mu}, \qquad z \in R_{d-1}.$$

Proof:
See Theorem 5.4.2 in Falk et al. (2004, [21]).

□

Definition 2.2.6
For a differentiable GPD W the function

$$\phi(z) := |c|^{d-1} \left(\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W \right) \left(T_P^{-1}(z, c) \right), \quad z \in R_{d-1}$$

from Theorem 2.2.5 is called the *Pickands density*.

◊

Remark that the Pickands density is the density of a probability measure only after the division by μ (in case $\mu > 0$), else it is the density of a measure, which assigns the mass μ to R_{d-1}.

2.3 Basic Examples of Generalized Pareto Distributions

In this section we present some examples of multivariate generalized Pareto distributions. We begin with the two extreme cases of independence and complete dependence in Section 2.3.1. After that we investigate the most important parametric family, the logistic model in Section 2.3.2. We also generalize a known counterexample from Falk et al. (2004, [21]) to an arbitrary dimension $d \geq 3$, showing that GP functions are not necessarily distribution functions. The nested logistic model and the asymmetric logistic model, which are generalizations of the logistic model are investigated in Sections 2.3.3 and 2.3.4. The asymmetric logistic model is thereby used to show that the angular measure of a GPD is not uniquely determined.

These models, together with a lot of others, can be found in Sections 3.4 and 3.5 of Kotz and Nadarajah (2000, [47]), described there for the case of extreme value distributions.

2.3.1 The Cases of Independence and Complete Dependence

Choose the angular measure ν such that it has only mass in the vertices of $\overline{R_{d-1}}$, i.e.,

$$\nu(\{e_i\}) = 1, \qquad i = 0, \ldots, d-1,$$

where e_i, $i = 1, \ldots, d-1$ denote the standard unit vectors of $\mathbb{R}^{d-1}$ and $e_0 := 0$. Thus e_i, $i = 0, \ldots, d-1$, are the vertices of the unit simplex $\overline{R_{d-1}}$. It is easy to see that the conditions (2.7) hold for this ν. One is also able to see by short elementary calculations that this leads to

$$D(t_1, \ldots, t_{d-1}) = 1, \qquad (t_1, \ldots, t_{d-1}) \in \overline{R_{d-1}},$$

and

$$W(x_1, \ldots, x_d) = 1 + \sum_{i=1}^{d} x_i = 1 - ||x||_1, \qquad (x_1, \ldots, x_d) \in (-\infty, 0)^d \text{ close to } 0.$$

In the case of an extreme value distribution with this angular measure ν, the corresponding random vector X has independent components $X_1, \ldots, X_d$. Therefore, this case is referred to as the *independence case*. In the generalized Pareto setup, however, the behavior is different. Actually in this case no observations fall into an area close to 0, which we show in the next theorem.

Theorem 2.3.1
Let $(X_1, \ldots, X_d)$ be distributed by $W(x)$ close to 0 with $D(t_1, \ldots, t_{d-1}) = 1$ for all $(t_1, \ldots, t_{d-1}) \in \overline{R_{d-1}}$. With $K_s := \{x \in (-\infty, 0)^d \mid ||x||_\infty < s\}$, $s > 0$ we denote the (open) cube with edge length s in the negative quadrant. Then there exists $b > 0$ close to 0 such that

$$P((X_1, \ldots, X_d) \in K_b) = 0.$$

Proof:
By definition of a generalized Pareto distribution there exists a $b > 0$ close to 0 such that $(X_1, \ldots, X_d)$ has the distribution function $W(x_1, \ldots, x_d) = 1 + \sum_{i=1}^d x_i$ on K_b. Thus by Theorem A.2.2 in Bhattacharya and Rao (1976, [4]) the density of W on K_b is

$$\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W(x_1, \ldots, x_d) = 0,$$

and the assertion follows.

□

However, it is still justified to speak of this case as the case of independence with the following rational: Let $Y = (Y_1, \ldots, Y_d)$ be a random vector with distribution function F and tail independent components Y_i. Suppose that the distribution function of Y_i is in the univariate domain of attraction of $\exp(x)$, $x \leq 0$ for each $i = 1, \ldots, d$. Then F is in the domain of attraction of $\exp(-||x||_1)$, see Proposition 5.27 in Resnick (1987, [59]). Thus by (2.13) or the results of Section 2 in Rootzén and Tajvidi (2005, [60]) we know that observations falling over a high threshold have asymptotically the distribution $1 + \log(\exp(-||x||_1)) = 1 - ||x||_1$ in the extreme area. So $W(x) = 1 - ||x||_1$ is the asymptotic exceedance distribution of random vectors with tail independent components, meaning that random vectors with tail independent components have in the limit no observations close to the origin. Because of this we, still, speak of $W(x) = 1 - ||x||_1$ as the independence case.

For practical purposes one should check ones observations for tail independence before applying a GPD model to make sure that one is not in the case of independence. Otherwise observed data are likely to show dependence even though one has tail independence, see also Section 8.4 of Coles (2001, [8]). In the case of tail independence a multivariate analysis of the tails can be reduced to the analysis of the tails of lower dimensional margins. Suggestions for tests for tail independence are given for example in Section 6.5 of Falk et al. (2004, [21]) or in Falk and Michel (2006, [24]) for EVD and related models, in Section 9.5.2 of Beirlant et al. (2005, [3]) also for other models.

In the sequel we will introduce the other extreme case, the case of complete dependence. If we choose the angular measure ν such that it has only mass in the point $(1/d, \ldots, 1/d) \in \mathbb{R}^{d-1}$, i.e.,

$$\nu(\{(1/d, \ldots, 1/d)\}) = d$$

then we get from (2.6) by elementary calculations

$$D(t_1, \ldots, t_{d-1}) = \max\left(t_1, \ldots, t_{d-1}, 1 - \sum_{i=1}^d t_i\right), \qquad (t_1, \ldots, t_{d-1}) \in \overline{R_{d-1}},$$

and

$$W(x_1,\ldots,x_d) = 1 + \min(x_1,\ldots,x_d) = 1 - \max(-x_1,\ldots,-x_d) = 1 - ||x||_\infty$$

for $(x_1,\ldots,x_d) \in (-\infty,0)^d$ close to 0. Here we have $X_1 = \ldots = X_d$ with probability 1, i.e., the random variables are completely dependent. This can be seen as follows.

Theorem 2.3.2
Let $X_1 < 0$ be uniformly distributed on $(-1,0)$, and put $X_d := X_{d-1} := \ldots := X_1$ with probability 1. Then the joint distribution function of $(X_1,\ldots,X_d)$ on the negative quadrant is $W(x) = 1 - ||x||_\infty$ for $||x||_\infty \leq 1$ and equal to 0 elsewhere.

Proof:
Choose $x = (x_1,\ldots,x_d) \in (-1,0)^d$. Then

$$\begin{aligned} W(x) &= P(X_1 \leq x_1,\ldots,X_d \leq x_d) = P(X_1 \leq x_1,\ldots,X_1 \leq x_d) = P(X_1 \leq \min(x_1,\ldots,x_d)) \\ &= 1 + \min(x_1,\ldots,x_d) = 1 - \max(|x_1|,\ldots,|x_d|) = 1 - ||x||_\infty. \end{aligned}$$

If one component of x is smaller than -1, we have obviously $W(x) = 0$.

□

In case the random variables $X_1,\ldots,X_d$ follow an extreme value distribution with this angular measure ν, we also have $X_1 = \ldots = X_d$ with probability 1. Therefore, this case is referred to as the case of *complete dependence* (see also Section 3.2 in Kotz and Nadarajah (2000, [47])).

Remark that we have shown in Theorem 2.3.2 that in the case of complete dependence the GP function is a distribution function on its entire support.

2.3.2 The Generalized Pareto Distribution of Logistic Type

We introduce in this subsection to the most important parametric family, the logistic family.

Definition 2.3.3
Let $(X_1,\ldots,X_d)$, $d \geq 2$, $X_i < 0$, $i = 1,\ldots,d$, be a multivariate random vector, whose distribution function W_λ has for $(x_1,\ldots,x_d)$ in a neighborhood of 0 the representation

$$W_\lambda(x_1,\ldots,x_d) = 1 + \left(\sum_{i=1}^d x_i\right) D_\lambda\left(\frac{x_1}{\sum_{i=1}^d x_i},\ldots,\frac{x_{d-1}}{\sum_{i=1}^d x_i}\right)$$

with

$$D_\lambda(t_1,\ldots,t_{d-1}) = \left(\sum_{i=1}^{d-1} t_i^\lambda + \left(1 - \sum_{i=1}^{d-1} t_i\right)^\lambda\right)^{1/\lambda}, \qquad \lambda \in [1,\infty).$$

The function W_λ is called the *generalized Pareto distribution of logistic type*, D_λ the *Pickands dependence function of logistic type*.

◊

The extreme value distribution of logistic type is, due to its simplicity, one of the most extensively studied and most frequently applied multivariate EVDs, see Section 3.5.1 of Kotz and Nadarajah (2000, [47]) for more information.

We start our examination of the GPD of logistic type by giving a simple representation of W_λ in showing that the corresponding norm in this case is the usual $||\cdot||_\lambda$–norm. This makes the logistic model a very natural model.

Lemma 2.3.4
We have

$$W_\lambda(x_1,\ldots,x_d) = 1 - \left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{\frac{1}{\lambda}} = 1 - ||x||_\lambda,$$

in which $x_i < 0$, $i = 1,\ldots,d$, $x = (x_1,\ldots,x_d)$ and x lies in a neighborhood of 0.

Proof:
In the following calculations we have $x_i < 0$ and, thus, $\sum_{i=1}^{d} x_i < 0$. This leads to

$$\begin{aligned}
W_\lambda(x_1,\ldots,x_d) &= 1 + \left(\sum_{j=1}^{d} x_j\right) D_\lambda\left(\frac{x_1}{\sum_{j=1}^{d} x_j},\ldots,\frac{x_{d-1}}{\sum_{j=1}^{d} x_j}\right) \\
&= 1 + \left(\sum_{j=1}^{d} x_j\right)\left[\sum_{i=1}^{d-1}\left(\frac{x_i}{\sum_{j=1}^{d} x_j}\right)^\lambda + \left(1 - \sum_{i=1}^{d-1}\frac{x_i}{\sum_{j=1}^{d} x_j}\right)^\lambda\right]^{\frac{1}{\lambda}} \\
&= 1 + \left(\sum_{j=1}^{d} x_j\right)\left[\sum_{i=1}^{d-1}\left(\frac{x_i}{\sum_{j=1}^{d} x_j}\right)^\lambda + \left(\frac{x_d}{\sum_{j=1}^{d} x_j}\right)^\lambda\right]^{\frac{1}{\lambda}} \\
&= 1 + \left(\sum_{j=1}^{d} x_j\right)\left[\sum_{i=1}^{d}\left(\frac{x_i}{\sum_{j=1}^{d} x_j}\right)^\lambda\right]^{\frac{1}{\lambda}} \\
&= 1 + \left(\sum_{j=1}^{d} x_j\right)\left[\sum_{i=1}^{d}\left(\frac{|x_i|}{|\sum_{j=1}^{d} x_j|}\right)^\lambda\right]^{\frac{1}{\lambda}} \\
&= 1 + \frac{\sum_{j=1}^{d} x_j}{|\sum_{j=1}^{d} x_j|}\left[\sum_{i=1}^{d}|x_i|^\lambda\right]^{\frac{1}{\lambda}} = 1 - \left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{\frac{1}{\lambda}} = 1 - ||x||_\lambda,
\end{aligned}$$

in a neighborhood of 0, which proves the assertion.

□

Remark 2.3.5
With this representation the case $\lambda = \infty$ of complete dependence can also be included in the logistic family, since we know from functional analysis that

$$\lim_{\lambda\to\infty} ||x||_\lambda = ||x||_\infty$$

for any $x \in \mathbb{R}^d$. This can be seen as follows. From $\sum_{i=1}^{d}|x_i|^\lambda \geq \max_{1\leq i\leq d}|x_i|^\lambda$ we conclude $||x||_\lambda \geq ||x||_\infty$ and $\liminf_{\lambda\to\infty}||x||_\lambda \geq ||x||_\infty$. On the other hand, the inequality $\sum_{i=1}^{d}|x_i|^\lambda \leq d\max_{1\leq i\leq d}|x_i|^\lambda$ implies $||x||_\lambda \leq d^{\frac{1}{\lambda}}||x||_\infty$ and, thus, $\limsup_{\lambda\to\infty}||x||_\lambda \leq ||x||_\infty$, which shows the assertion.

The parameter λ can now be interpreted as governing the dependence, since $\lambda = 1$ is the case of independence and $\lambda = \infty$ the case of complete dependence. Values of λ close to 1 imply a

great deal of independence, large values of λ a great deal of dependence. A sort of middle case is given in Lemma 2.3.11.

However, the distribution function for $\lambda = \infty$ is not differentiable, which is the reason for excluding it from some assertions like the following lemma.

$\Diamond$

We now consider the density of W_λ.

Lemma 2.3.6
For $1 \leq \lambda < \infty$ the function

$$w_\lambda(x_1, \ldots, x_d) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W_\lambda(x_1, \ldots, x_d)$$

is the density of W_λ in a neighborhood of 0. It has the representation

$$\begin{aligned} w_\lambda(x_1, \ldots, x_d) &= \prod_{i=1}^{d-1}(i\lambda - 1) \prod_{i=1}^{d}(-x_i)^{\lambda-1} \left(\sum_{i=1}^{d}(-x_i)^\lambda \right)^{-d+\frac{1}{\lambda}} \\ &= \prod_{i=1}^{d-1}(i\lambda - 1) \prod_{i=1}^{d}(-x_i)^{\lambda-1} ||x||_\lambda^{1-d\lambda} \end{aligned}$$

for $x_i < 0, i = 1, \ldots, d$, and x close to 0.

Proof:
The case $\lambda = 1$ is obvious since we know by Theorem 2.3.1 that W_1 has probability 0 close to 0 and the function w_λ reduces to 0 for $\lambda = 1$. Therefore, we assume from now on $\lambda > 1$.

First we show by induction that the equation

$$\frac{\partial^j}{\partial x_1 \ldots \partial x_j} W_\lambda(x_1, \ldots, x_d) = \prod_{i=1}^{j-1}(i\lambda - 1) \prod_{i=1}^{j}(-x_i)^{\lambda-1} \left(\sum_{i=1}^{d}(-x_i)^\lambda \right)^{-j+\frac{1}{\lambda}} \tag{2.16}$$

holds for $j = 1, \ldots, d$.

To compute the derivative we return to the representation from Lemma 2.3.4. For the zero step of the induction ($j = 1$) we have:

$$\begin{aligned} \frac{\partial}{\partial x_1} W_\lambda(x_1, \ldots, x_d) &= -\frac{1}{\lambda} \left(\sum_{i=1}^{d}(-x_i)^\lambda \right)^{\frac{1}{\lambda}-1} \lambda(-x_1)^{\lambda-1}(-1) = \left(\sum_{i=1}^{d}(-x_i)^\lambda \right)^{\frac{1}{\lambda}-1} (-x_1)^{\lambda-1} \\ &= \prod_{i=1}^{1-1}(i\lambda - 1) \prod_{i=1}^{1}(-x_i)^{\lambda-1} \left(\sum_{i=1}^{d}(-x_i)^\lambda \right)^{-1+\frac{1}{\lambda}}, \end{aligned}$$

with the convention that an empty product equals 1.

Assume that the assertion is shown for $j - 1 < d$. Then

$$\frac{\partial^j}{\partial x_1 \ldots \partial x_j} W_\lambda(x_1, \ldots, x_d)$$

$$
\begin{aligned}
&= \frac{\partial}{\partial x_j}\frac{\partial^{j-1}}{\partial x_1 \dots \partial x_{j-1}} W_\lambda(x_1,\dots,x_d) \\
&= \frac{\partial}{\partial x_j}\left[\prod_{i=1}^{j-2}(i\lambda-1)\prod_{i=1}^{j-1}(-x_i)^{\lambda-1}\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{-j+1+\frac{1}{\lambda}}\right] \\
&= \prod_{i=1}^{j-2}(i\lambda-1)\prod_{i=1}^{j-1}(-x_i)^{\lambda-1}\frac{\partial}{\partial x_j}\left[\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{-j+1+\frac{1}{\lambda}}\right] \\
&= \prod_{i=1}^{j-2}(i\lambda-1)\prod_{i=1}^{j-1}(-x_i)^{\lambda-1}\left(\frac{1}{\lambda}-j+1\right)\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{-j+\frac{1}{\lambda}}\lambda(-x_j)^{\lambda-1}(-1) \\
&= \prod_{i=1}^{j-2}(i\lambda-1)\prod_{i=1}^{j-1}(-x_i)^{\lambda-1}\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{-j+\frac{1}{\lambda}}(-x_j)^{\lambda-1}(\lambda j-\lambda-1) \\
&= \prod_{i=1}^{j-1}(i\lambda-1)\prod_{i=1}^{j}(-x_i)^{\lambda-1}\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{-j+\frac{1}{\lambda}}.
\end{aligned}
$$

Setting $j = d$ in (2.16), one gets

$$
\begin{aligned}
w_\lambda(x_1,\dots,x_d) &= \frac{\partial^d}{\partial x_1 \dots \partial x_d} W_\lambda(x_1,\dots,x_d) = \prod_{i=1}^{d-1}(i\lambda-1)\prod_{i=1}^{d}(-x_i)^{\lambda-1}\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{-d+\frac{1}{\lambda}} \\
&= \prod_{i=1}^{d-1}(i\lambda-1)\prod_{i=1}^{d}(-x_i)^{\lambda-1}||x||_\lambda^{1-d\lambda},
\end{aligned}
$$

and with Theorem A.2.2 from Bhattacharya and Rao (1976, [4]) the assertion follows.

□

We want to give the Pickands density, introduced in Theorem 2.2.5 and Definition 2.2.6, for the logistic case. This is the first main result of this manuscript.

Theorem 2.3.7
Let $1 < \lambda < \infty$. Put

$$
\phi_\lambda(z_1,\dots,z_{d-1}) := |c|^{d-1}\left(\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W_\lambda\right)(T_P^{-1}(z_1,\dots,z_{d-1},c))
$$

and $\mu_\lambda := \int_{R_{d-1}} \phi_\lambda(z)\,dz$. Then we have

$$
\phi_\lambda(z_1,\dots,z_{d-1}) = \left(\prod_{i=1}^{d-1}(i\lambda-1)\right)\frac{\left(\prod_{i=1}^{d-1} z_i\right)^{\lambda-1}\left(1-\sum_{i=1}^{d-1} z_i\right)^{\lambda-1}}{D_\lambda(z_1,\dots,z_{d-1})^{d\lambda-1}}, \tag{2.17}
$$

$\mu_\lambda > 0$ and, thus, $f_\lambda(z) := \frac{\phi_\lambda(z)}{\mu_\lambda}$ is well defined. Furthermore, ϕ_λ and f_λ respectively assume their maxima on the compact set $\overline{R_{d-1}}$ in the point $\left(\frac{1}{d},\dots,\frac{1}{d}\right)$, and they have the values

$$
m_\lambda := d^{\frac{1}{\lambda}-1}\prod_{i=1}^{d-1}(i\lambda-1) \quad \text{and} \quad \frac{m_\lambda}{\mu_\lambda}
$$

respectively.

Proof:
We show equation (2.17) by using the density

$$w_\lambda(x_1,\dots,x_d) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} W_\lambda(x_1,\dots,x_d) = \prod_{i=1}^{d-1}(i\lambda-1)\prod_{i=1}^{d}(-x_i)^{\lambda-1}\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{-d+\frac{1}{\lambda}}$$

of W_λ according to Lemma 2.3.6. Then we can compute ϕ_λ by inserting the inverse Pickands transformation (2.2).

$$\begin{aligned}
\phi_\lambda(z_1,\dots,z_{d-1}) &= |c|^{d-1} w_\lambda(T_P^{-1}(z_1,\dots,z_{d-1},c)) \\
&= |c|^{d-1}\prod_{i=1}^{d-1}(i\lambda-1)\prod_{i=1}^{d-1}(-cz_i)^{\lambda-1}\left(-c\left(1-\sum_{j=1}^{d-1}z_j\right)\right)^{\lambda-1} \\
&\quad\cdot\left[\sum_{i=1}^{d-1}(-cz_i)^\lambda + \left(-c\left(1-\sum_{j=1}^{d-1}z_j\right)\right)^\lambda\right]^{\frac{1}{\lambda}-d} \\
&= \left(\prod_{i=1}^{d-1}(i\lambda-1)\right)|c|^{d-1}(-c)^{(\lambda-1)(d-1)}\left(\prod_{i=1}^{d-1}z_i^{\lambda-1}\right)(-c)^{\lambda-1}\left(1-\sum_{j=1}^{d-1}z_j\right)^{\lambda-1} \\
&\quad\cdot\left[(-c)^\lambda\sum_{i=1}^{d-1}z_i^\lambda + (-c)^\lambda\left(1-\sum_{j=1}^{d-1}z_j\right)^\lambda\right]^{\frac{1}{\lambda}-d} \\
&= \left(\prod_{i=1}^{d-1}(i\lambda-1)\right)|c|^{d-1}|c|^{(\lambda-1)d}\left(\prod_{i=1}^{d-1}z_i^{\lambda-1}\right)\left(1-\sum_{j=1}^{d-1}z_j\right)^{\lambda-1} \\
&\quad\cdot|c|^{\lambda\left(\frac{1}{\lambda}-d\right)}\left[\sum_{i=1}^{d-1}z_i^\lambda + \left(1-\sum_{j=1}^{d-1}z_j\right)^\lambda\right]^{\frac{1}{\lambda}-d} \\
&= \left(\prod_{i=1}^{d-1}(i\lambda-1)\right)|c|^{d-1+(\lambda-1)d+1-\lambda d}\left(\prod_{i=1}^{d-1}z_i^{\lambda-1}\right)\left(1-\sum_{j=1}^{d-1}z_j\right)^{\lambda-1} \\
&\quad\cdot D_\lambda(z_1,\dots,z_{d-1})^{1-d\lambda} \\
&= \left(\prod_{i=1}^{d-1}(i\lambda-1)\right)\frac{\left(\prod_{i=1}^{d-1}z_i\right)^{\lambda-1}\left(1-\sum_{i=1}^{d-1}z_i\right)^{\lambda-1}}{D_\lambda(z_1,\dots,z_{d-1})^{d\lambda-1}}
\end{aligned}$$

Therefore, we have shown equation (2.17). Since $\phi_\lambda(z)$ is continuous, non-negative and does not vanish in the point $\left(\frac{1}{d},\dots,\frac{1}{d}\right)$ for $1<\lambda<\infty$ (see below), $\mu_\lambda>0$ follows.

Now we have to show that ϕ_λ takes its maximum on $\overline{R_{d-1}}$ in the point $\left(\frac{1}{d},\dots,\frac{1}{d}\right)$. The existence of the maximum is guaranteed, since ϕ_λ is continuous and $\overline{R_{d-1}}$ is compact. By Section 4.3 of Falk et al. (2004, [21]) the denominator in (2.17) has its minimum in the point $\left(\frac{1}{d},\dots,\frac{1}{d}\right)$. Therefore, it suffices to show that the numerator assumes its maximum also in $\left(\frac{1}{d},\dots,\frac{1}{d}\right)$. This will now be proven.

Put, therefore,

$$N(z_1,\dots,z_{d-1}) := \left(\prod_{i=1}^{d-1} z_i\right)^{\lambda-1}\left(1-\sum_{i=1}^{d-1} z_i\right)^{\lambda-1}.$$

We have obviously $N \geq 0$ on $\overline{R_{d-1}}$. Furthermore, we have $N|_{\partial\overline{R_{d-1}}} = 0$ with

$$\partial\overline{R_{d-1}} := \left\{(x_1,\dots,x_{d-1}) \in \overline{R_{d-1}} \,\middle|\, x_i = 0 \text{ for an } i \in \{1,\dots,d-1\} \text{ or } \sum_{i=1}^{d-1} x_i = 1\right\},$$

the boundary of $\overline{R_{d-1}}$. This follows immediately from inserting the corresponding boundary conditions. Thus, the maximum must lie in the interior. We compute the gradient of N and set it to 0.

$$\begin{aligned}
\frac{\partial}{\partial z_i}N(z_1,\dots,z_{d-1}) &= \left(\prod_{j=1,j\neq i}^{d-1} z_j\right)^{\lambda-1}(\lambda-1)z_i^{\lambda-2}\left(1-\sum_{j=1}^{d-1} z_j\right)^{\lambda-1} \\
&+ \left(\prod_{j=1}^{d-1} z_j\right)^{\lambda-1}(\lambda-1)\left(1-\sum_{j=1}^{d-1} z_j\right)^{\lambda-2}(-1) = 0 \\
&\iff 0 = z_i^{\lambda-2}\left(1-\sum_{j=1}^{d-1} z_j\right)^{\lambda-1} - z_i^{\lambda-1}\left(1-\sum_{j=1}^{d-1} z_j\right)^{\lambda-2} \\
&\iff 0 = 1-\sum_{j=1}^{d-1} z_j - z_i \iff 2z_i + \sum_{j=1,i\neq j}^{d-1} z_j = 1
\end{aligned}$$

In the transformation the respective terms could be cancelled, since they only vanish on the boundary. The final equation is equivalent to the system of linear equations

$$Az := \begin{pmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ 1 & 1 & 2 & & \vdots \\ \vdots & \vdots & & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 2 \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{d-1} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

The matrix A is a $(d-1)\times(d-1)$–matrix and has full rank, thus the system of equations is uniquely solvable. Insertion of $z = \left(\frac{1}{d},\dots,\frac{1}{d}\right)^T$ shows that this is a solution and, thus, the only possible one. Therefore, N takes its maximum on $\left(\frac{1}{d},\dots,\frac{1}{d}\right)$, and we can conclude that $\left(\frac{1}{d},\dots,\frac{1}{d}\right)$ is the maximum of ϕ_λ.

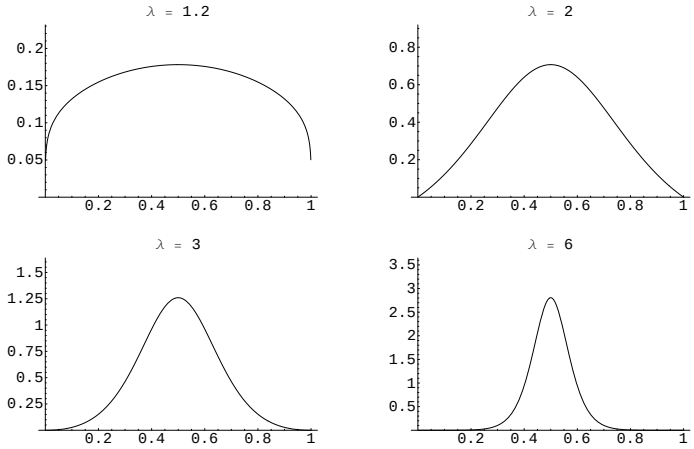
Inserting this point in (2.17) we get

$$m_\lambda = \left(\prod_{i=1}^{d-1}(i\lambda-1)\right)\frac{\left(\frac{1}{d}\right)^{d(\lambda-1)}}{\left(d\left(\frac{1}{d}\right)^\lambda\right)^{\frac{1}{\lambda}(d\lambda-1)}} = d^{\frac{1}{\lambda}-1}\prod_{i=1}^{d-1}(i\lambda-1).$$

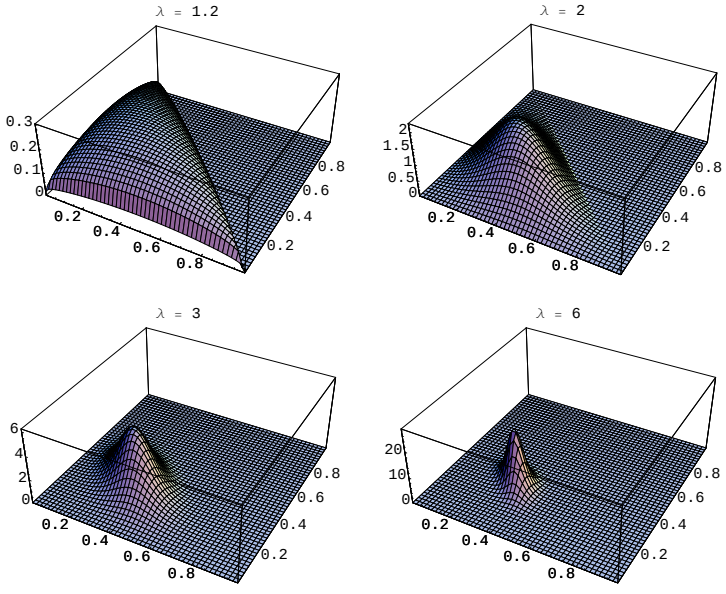
□

Observe that we can compute the Pickands density only for $\lambda < \infty$, since the case $\lambda = \infty$ is not differentiable. Furthermore, we have seen that $\mu_\lambda = \int_{R_{d-1}} \phi_\lambda(z)\, dz > 0$ if and only if $1 < \lambda < \infty$, since $\phi_1 = 0$. Therefore, the case $\lambda = 1$ also had to be excluded in Theorem 2.3.7.

Example 2.3.8
We display the Pickands density ϕ_λ of the logistic family computed in Theorem 2.3.7 for $d = 2$ and various λ,



and for $d = 3$.



◊

Remark 2.3.9
The factor $\prod_{i=1}^{d-1} (i\lambda - 1)$ appearing in Theorem 2.3.7 also has the representation

$$\prod_{i=1}^{d-1} (i\lambda - 1) = \lambda^{d-1} \frac{\Gamma\left(d - \frac{1}{\lambda}\right)}{\Gamma\left(1 - \frac{1}{\lambda}\right)},$$

where $\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)\,dt$, $x > 0$ denotes the Gamma function. Since the relation $\Gamma(x+1) = x\Gamma(x)$ holds (see Theorem 150.1 in Heuser (1998, [38])), we get

$$\begin{aligned}
\prod_{i=1}^{d-1}(i\lambda - 1) &= \lambda^{d-1}\prod_{i=1}^{d-1}\left(i - \frac{1}{\lambda}\right) \\
&= \lambda^{d-1}\left(d-1-\frac{1}{\lambda}\right)\cdot\left(d-2-\frac{1}{\lambda}\right)\cdot\left(d-3-\frac{1}{\lambda}\right)\cdot\ldots\cdot\left(2-\frac{1}{\lambda}\right)\cdot\left(1-\frac{1}{\lambda}\right) \\
&= \lambda^{d-1}\frac{\Gamma\left(d-\frac{1}{\lambda}\right)}{\Gamma\left(d-1-\frac{1}{\lambda}\right)}\cdot\frac{\Gamma\left(d-1-\frac{1}{\lambda}\right)}{\Gamma\left(d-2-\frac{1}{\lambda}\right)}\cdot\frac{\Gamma\left(d-2-\frac{1}{\lambda}\right)}{\Gamma\left(d-3-\frac{1}{\lambda}\right)}\cdot\ldots\cdot\frac{\Gamma\left(3-\frac{1}{\lambda}\right)}{\Gamma\left(2-\frac{1}{\lambda}\right)}\cdot\frac{\Gamma\left(2-\frac{1}{\lambda}\right)}{\Gamma\left(1-\frac{1}{\lambda}\right)} \\
&= \lambda^{d-1}\frac{\Gamma\left(d-\frac{1}{\lambda}\right)}{\Gamma\left(1-\frac{1}{\lambda}\right)}.
\end{aligned}$$

This second representation can be simpler for practical evaluations with the computer, if one can compute $\Gamma(x)$ efficiently. With the product representation d multiplications have to be done, which is of order $O(d)$, whereas for fixed λ a constant number of operations has to be carried out for the representation via the Gamma function. In MATHEMATICA, for instance, this new representation is superior to the product representation for very high d. This can be seen in the following table, where both representations were used to compute the factor for $\lambda = 4$.

d	10	1000	100000	10000000
time in sec. for $\lambda^{d-1}\frac{\Gamma\left(d-\frac{1}{\lambda}\right)}{\Gamma\left(1-\frac{1}{\lambda}\right)}$	0.01	0.01	0.01	0.01
time in sec. for $\prod_{i=1}^{d-1}(i\lambda-1)$	0.01	0.01	0.26	26.8

The differences are only noticeable in very high dimensions. This new representation will also be used for the computation of upper bounds of certain runtimes in Theorem 3.2.13.

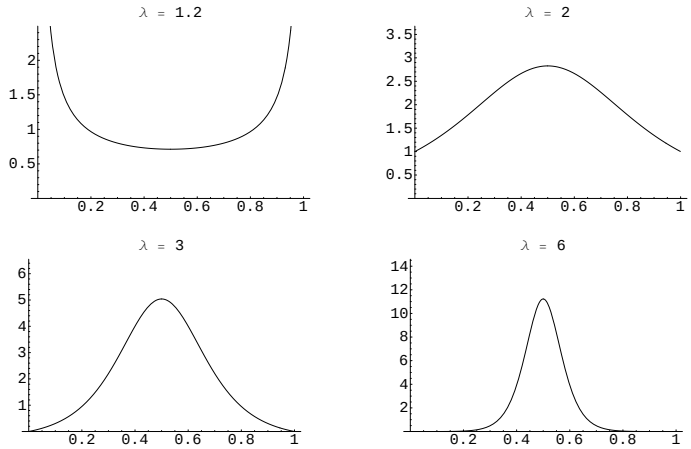
$\Diamond$

Example 2.3.10
The logistic family has, according to Section 3.5.1 in Kotz and Nadarajah (2000, [47]), for $1 \le \lambda < \infty$ the angular density

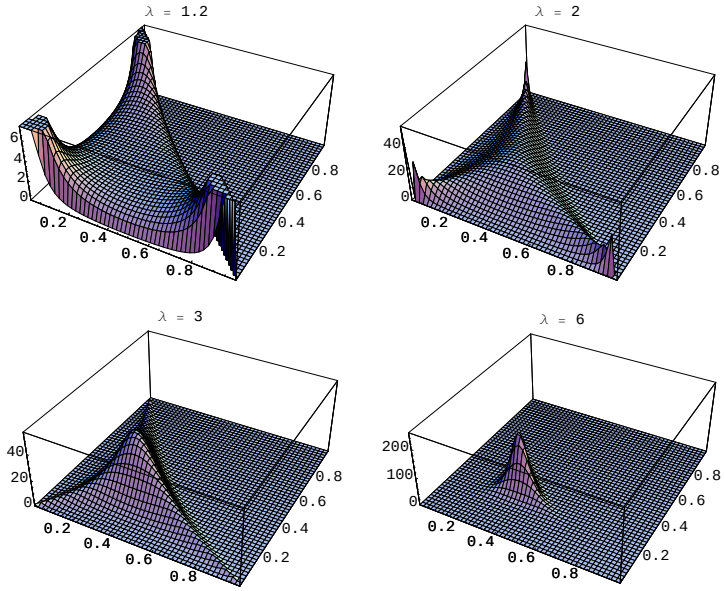
$$\begin{aligned}
&l_\lambda(z_1,\ldots,z_{d-1}) \\
&= \prod_{i=1}^{d-1}(i\lambda-1)\left(\prod_{i=1}^{d-1}z_i\right)^{-\lambda-1}\left(1-\sum_{i=1}^{d-1}z_i\right)^{-\lambda-1}\left(\sum_{i=1}^{d-1}z_i^{-\lambda}+\left(1-\sum_{i=1}^{d-1}z_i\right)^{-\lambda}\right)^{1/\lambda-d}
\end{aligned}$$

with $d_\lambda^* = d$ for $\lambda > 1$. Like the Pickands density the angular density reduces to 0 for $\lambda = 1$. It

has for $d = 2$ and miscellaneous λ the graphs



and for $d = 3$ the graphs



◊

For the investigation of the tail dependence structure the behavior of $l(z)$ especially at the vertices of $\overline{R_{d-1}}$ is of interest. A convergence of the angular density to ∞ in the vertices signifies a high degree of independence, convergence to 0 a high degree of dependence. This follows from the fact that in the case of independence the angular measure ν has all its mass in the vertices of $\overline{R_{d-1}}$, and in the case of complete dependence all its mass in the point $\left(\frac{1}{d}, \ldots, \frac{1}{d}\right)$, see Section 2.3.1.

The tail dependence structure is not distinctly visualized in such a way by the Pickands density, since all functions have a maximum in the interior and converge to 0 at the boundary in the logistic case, see Example 2.3.8. However, we will see that the Pickands density will also be an important tool in these considerations in Chapter 4.

We want to investigate the logistic case still a little closer. We are interested in a kind of boundary between complete dependence and independence. The following lemma shows that $\lambda = d$ can be considered as such a boundary, since in this case the angular density converges to a fixed real value not equal to 0 when approaching the vertices.

Lemma 2.3.11
We have for $\lambda > 1$

$$\lim_{z \to e_i} l_\lambda(z) = \begin{cases} \infty & \text{for } \lambda < d, \\ \left(\prod_{i=1}^{d-1}(i\lambda - 1)\right)(d-1)^{1/\lambda - d} & \text{for } \lambda = d, \\ 0 & \text{for } \lambda > d, \end{cases}$$

where e_i, $i = 0, \ldots, d-1$ are the vertices of $\overline{R_{d-1}}$ as above.

Proof:
We first consider the case $z \to e_0$.

$$\begin{aligned} \lim_{t \to 0} l_\lambda(t, \ldots, t) &= \lim_{t \to 0} \left(\prod_{i=1}^{d-1}(i\lambda - 1)\right) \left(t^{d-1}\right)^{-\lambda-1} (1-(d-1)t)^{-\lambda-1} \cdot \\ &\quad \cdot \left((d-1)t^{-\lambda} + (1-(d-1)t)^{-\lambda}\right)^{1/\lambda - d} \\ &= \left(\prod_{i=1}^{d-1}(i\lambda - 1)\right) \lim_{t \to 0} \left(t^{d-1}\right)^{-\lambda-1} \left((d-1)t^{-\lambda}\right)^{1/\lambda - d} \\ &= \left(\prod_{i=1}^{d-1}(i\lambda - 1)\right)(d-1)^{1/\lambda - d} \lim_{t \to 0} t^{(d-1)(-\lambda-1)+(-\lambda)(1/\lambda - d)} \\ &= \left(\prod_{i=1}^{d-1}(i\lambda - 1)\right)(d-1)^{1/\lambda - d} \lim_{t \to 0} t^{-d+\lambda} \\ &= \begin{cases} \infty & \text{for } \lambda < d, \\ \left(\prod_{i=1}^{d-1}(i\lambda - 1)\right)(d-1)^{1/\lambda - d} & \text{for } \lambda = d, \\ 0 & \text{for } \lambda > d. \end{cases} \end{aligned}$$

Since $l(z)$ is continuous, we have

$$\lim_{t \to 0} l_\lambda(t, \ldots, t) = \lim_{z \to e_0} l_\lambda(z).$$

Because of the following equalities, which are easily checked,

$$l_\lambda(t, \ldots, t) = l_\lambda(1-(d-1)t, t, \ldots, t) = l_\lambda(t, 1-(d-1)t, t, \ldots, t) = \ldots = l_\lambda(t, \ldots, t, 1-(d-1)t)$$

and again the continuity of l_λ, we have exactly the same behavior in the other vertices. Thus $\lambda = d$ can be seen as the boundary case between complete dependence and independence.

□

In Remark 2.2.3 it was stated that a GP function W is not necessarily a distribution function on its entire support in the negative quadrant. In Section 5.1 of Falk et al. (2004, [21]) an example was given for this by showing that in the trivariate case $W(x_1, x_2, x_3) = \max(1 + x_1 + x_2 + x_3, 0)$ does not define a distribution function, since it would assign the probability $-\frac{1}{2}$ to the cube

$\left(-\frac{1}{2}, 0\right]^d$. This example corresponds to the case where $D = 1$, i.e, the case of independence, which corresponds to the logistic case with $\lambda = 1$. We will show in the sequel that this counterexample also holds for dimensions $d \geq 3$, and we show additionally that this example can be transferred to cases where D is not the constant 1. This is done with the help of the logistic model. Recall that it was shown in Section 5.1 of Falk et al. (2004, [21]) that a bivariate GPD is a distribution function on its entire support in the negative quadrant. The main result we present here shows that no such assertion is possible for arbitrary dimension $d \geq 3$.

Theorem 2.3.12
The following assertions hold for any dimension $d \geq 3$:

(i) The GP function
$$W_1(x) = \max\left(1 - ||x||_1, 0\right)$$
for $x = (x_1, \ldots, x_d)$ with $x_i < 0$, $i = 1, \ldots, d$ does not define a distribution function.

(ii) There exists a $\lambda_0 > 1$, depending on d, such that the function
$$W_\lambda(x) = \max\left(1 - ||x||_\lambda, 0\right)$$
for $x = (x_1, \ldots, x_d)$ with $x_i < 0$, $i = 1, \ldots, d$ does not define a distribution function for any $1 \leq \lambda < \lambda_0$.

Proof:
We prove assertion (i) by showing that the cube $K := \left(-\frac{1}{2}, 0\right]^d$ would be assigned a negative probability if W_1 were a distribution function. By Section 4.1 of Falk et al. (2004, [21]) the probability of K under W_1 would be
$$P(K) = \sum_{m \in \{0,1\}^d} (-1)^{\left(d - \sum_{j=1}^d m_j\right)} W_1\left(0^{m_1}\left(-\frac{1}{2}\right)^{1-m_1}, \ldots, 0^{m_d}\left(-\frac{1}{2}\right)^{1-m_d}\right)$$
with the convention that $0^0 := 1$. Note that if more than one argument of W_1 is $-\frac{1}{2}$, i.e., $m_j = 0$ for more than one index $j = 1, \ldots, d$, then W_1 equals 0 at this point. Thus the above sum reduces to those summands, where at most one component of m is 0. With
$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{else,} \end{cases} \tag{2.18}$$
denoting the indicator function of a set A, the probability of K can be written as
$$\begin{aligned} P(K) &= W_1(0, \ldots, 0) + \sum_{j=1}^d (-1)^{(d-(d-1))} W_1\left(-\frac{1}{2} \cdot \mathbf{1}_{\{j\}}(1), \ldots, -\frac{1}{2} \cdot \mathbf{1}_{\{j\}}(d)\right) \\ &= 1 + \sum_{j=1}^d (-1)\left(1 - \frac{1}{2}\right) = 1 - \frac{d}{2}. \end{aligned}$$
We have $1 - \frac{d}{2} < 0$ if $d \geq 3$ and, thus, W_1 cannot define a distribution function there.

We will show assertion (ii) by using assertion (i) together with a continuity argument. Since we know from functional analysis that
$$\lim_{\lambda \to 1} ||x||_\lambda = ||x||_1$$

for any $x \in \mathbb{R}^d$, we can conclude the pointwise convergence

$$\lim_{\lambda \to 1} W_\lambda(x) = W_1(x) \tag{2.19}$$

for x in the negative quadrant.

Let $m = (m_1, \ldots, m_d) \in \{0,1\}^d$ and define the point

$$p_m := \left(0^{m_1} \left(-\frac{1}{2} \right)^{1-m_1}, \ldots, 0^{m_d} \left(-\frac{1}{2} \right)^{1-m_d} \right) \in \mathbb{R}^d$$

with the convention $0^0 = 1$ as above. Since the cardinality of $\{0,1\}^d$ is 2^d, the p_m define 2^d different points.

Let $\varepsilon > 0$ be given. By the pointwise convergence (2.19) of W_λ, there exists a $\lambda_m > 1$ for any $m \in \{0,1\}^d$ such that

$$|W_\lambda(p_m) - W_1(p_m)| < \frac{\varepsilon}{2^d}$$

for all $1 < \lambda < \lambda_m$. Put $\lambda_0 := \min_{m \in \{0,1\}^d} \lambda_m > 1$. Then we have

$$|W_\lambda(p_m) - W_1(p_m)| < \frac{\varepsilon}{2^d} \tag{2.20}$$

for all $1 < \lambda < \lambda_0$ and all $m \in \{0,1\}^d$. If we assume that W_λ is a distribution function, we have for the probability $P_\lambda(K)$ under W_λ

$$\begin{aligned}
\left| P_\lambda(K) - \left(1 - \frac{d}{2} \right) \right| &= \left| \sum_{m \in \{0,1\}^d} (-1)^{\left(d - \sum_{j=1}^d m_j\right)} W_\lambda(p_m) - \sum_{m \in \{0,1\}^d} (-1)^{\left(d - \sum_{j=1}^d m_j\right)} W_1(p_m) \right| \\
&= \left| \sum_{m \in \{0,1\}^d} (-1)^{\left(d - \sum_{j=1}^d m_j\right)} \left(W_\lambda(p_m) - W_1(p_m) \right) \right| \\
&\leq \sum_{m \in \{0,1\}^d} |W_\lambda(p_m) - W_1(p_m)| \overset{(2.20)}{<} \sum_{m \in \{0,1\}^d} \frac{\varepsilon}{2^d} = \varepsilon.
\end{aligned}$$

We know that $1 - \frac{d}{2} \leq -\frac{1}{2}$ for $d \geq 3$. If we choose $\varepsilon = \frac{1}{4}$, then we know that there exists a $\lambda_0 > 1$ such that we have $P_\lambda(K) \leq -\frac{1}{4}$ for any $1 < \lambda < \lambda_0$. Thus W_λ cannot be a distribution function for $\lambda < \lambda_0$.

□

We have shown in Theorem 2.3.12 that GP functions close to the independence case do not define a distribution function on their entire support for arbitrary dimension $d \geq 3$. However, we have seen in Theorem 2.3.2 that the GP function is a valid distribution function on its entire support in the case of complete dependence for any dimension. An open question is, how far one has to divert from the independence case or, respectively, how close one has to be to case of complete dependence such that a GP function is a valid distribution function on its entire support. Is there, for example, a λ_0 in the logistic case such that for $\lambda < \lambda_0$ the corresponding GPD is not a distribution function and for $\lambda > \lambda_0$ the corresponding GP function is a distribution function on its entire support? Or is $\lambda = \infty$ the only case, where a GP function is a distribution function on its entire support?

Theorem 2.3.12 was shown by computing the probability of the cube $\left(-\frac{1}{2}, 0\right]^d$. These considerations can surely be generalized to cubes $(-a, 0]^d$ with $a > 0$ suitably chosen. What these examples do not show, however, is the underlying mathematical structure, which results in this strange behavior for $d \geq 3$.

2.3.3 The Generalized Pareto Distribution of Nested Logistic Type

The logistic GPD from the previous section is an exchangeable model, i.e., an arbitrary permutation of the components of a corresponding random vector X has the same distribution as the original random vector itself. Such a random vector is often referred to as symmetric. In this and the next section we will introduce GPDs, whose corresponding random vectors are not exchangeable. Such models are often called asymmetric.

The idea behind the nested logistic model is to generalize the logistic model to allow different degrees of dependence between the components of the underlying random vector. In the logistic model with parameter λ the bivariate marginal distributions are again GPDs of logistic type, all with the same parameter λ. In the nested logistic model the bivariate margins will still remain GPDs of logistic type but with (possibly) different parameters.

The nested logistic distribution introduced in this section is a special case of the two-level logistic distribution from Section 3.5.2 in Kotz and Nadarajah (2000, [47]). The nested logistic model is described in its full multivariate version among others in Section 3 of Joe (1994, [42]) for the extreme value case.

Let $\lambda_1, \ldots, \lambda_{d-1} \geq 1$. Define now for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ recursively the following norm

$$||x||_{\lambda_1,\ldots,\lambda_{d-1}} := ||(x_1, \ldots, x_d)||_{\lambda_1,\ldots,\lambda_{d-1}} := \Big|\Big| \left(||(x_1, \ldots, x_{d-1})||_{\lambda_1,\ldots,\lambda_{d-2}}, x_d\right) \Big|\Big|_{\lambda_{d-1}}, \qquad (2.21)$$

where $||\cdot||_\lambda$ is the usual λ–norm with the convention that the absolute value is taken if the norm does not have an index. For $d = 2$ or $\lambda_1 = \ldots = \lambda_{d-1} = \lambda$ the norm (2.21) reduces to the λ–norm. For $d > 2$ one can easily show by induction for d that this is indeed a norm.

Definition 2.3.13
For $\lambda_1 \geq \ldots \geq \lambda_{d-1} \geq 1$ the distribution function

$$W_{\lambda_1,\ldots,\lambda_{d-1}}(x_1, \ldots, x_d) := 1 - ||(x_1, \ldots, x_d)||_{\lambda_1,\ldots,\lambda_{d-1}}$$

for $x_i < 0$, $i = 1, \ldots, d$, close to 0 is called the *generalized Pareto distribution of nested logistic type*.

$\Diamond$

Remark 2.3.14
In the case $d = 2$ or $\lambda_1 = \ldots = \lambda_{d-1}$ this is the logistic distribution from Definition 2.3.3. For $d = 3$ the distribution function is

$$W_{\lambda_1,\lambda_2}(x_1, x_2, x_3) = 1 - \left(\left((-x_1)^{\lambda_1} + (-x_2)^{\lambda_1}\right)^{\lambda_2/\lambda_1} + (-x_3)^{\lambda_2}\right)^{1/\lambda_2}.$$

One can easily see by letting the respective arguments tend to 0 that the marginal distribution of X_1 and X_2 is a logistic GPD with parameter λ_1, that the marginal distribution of X_1 and X_3 as well as the marginal distribution of X_2 and X_3 is a logistic GPD with the same parameter λ_2.

For $d = 4$ we have the distribution function

$$W_{\lambda_1,\lambda_2,\lambda_3}(x_1, x_2, x_3, x_4) = 1 - \left(\left(\left((-x_1)^{\lambda_1} + (-x_2)^{\lambda_1} \right)^{\lambda_2/\lambda_1} + (-x_3)^{\lambda_2} \right)^{\lambda_3/\lambda_2} + (-x_4)^{\lambda_3} \right)^{1/\lambda_4}.$$

Again the bivariate margins are logistic with different parameters. The extension to higher dimensions is intuitively clear, although a little tedious to notate.

Remark that not every possible combination of bivariate logistic margins can be combined in this manner to a nested logistic model. The references at the beginning of the section contain further possible generalizations with more freedom in the choice of the dependence parameters of the bivariate margins.

The condition $\lambda_1 \geq \ldots \geq \lambda_{d-1} \geq 1$ is sufficient for $W_{\lambda_1,\ldots,\lambda_{d-1}}$ to define a GPD, see Section 5 in Joe (1994, [42]), where the proof is given for extreme value distributions. It is, however, unclear if this condition is also necessary. A counterexample, given in Remark 6.9 of Hofmann (2006, [39]), shows that arbitrary $\lambda_1, \ldots, \lambda_{d-1} \geq 1$ do not always give an extreme value and, thus, a generalized Pareto distribution of nested logistic type.

The angular and the Pickands density are hard to denote in general dimension. We, therefore, only give them for dimension $d = 3$, since they will be plotted in Example 2.3.15 and used in the estimation procedures in Section 7.3.

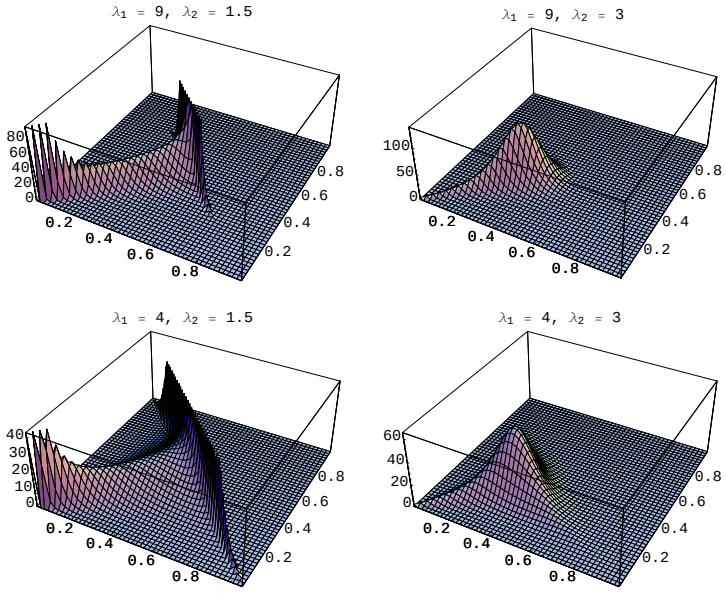
$$\begin{aligned}
l_{\lambda_1,\lambda_2}(z_1, z_2) &= (\lambda_2 - 1)\,(z_1 z_2)^{-\lambda_1 - 1}\,(1 - z_1 - z_2)^{-\lambda_2 - 1} \left(z_1^{-\lambda_1} + z_2^{-\lambda_1} \right)^{\frac{\lambda_2}{\lambda_1} - 2} \\
&\quad \cdot \left(\left(z_1^{-\lambda_1} + z_2^{-\lambda_1} \right)^{\frac{\lambda_2}{\lambda_1}} + (1 - z_1 - z_2)^{-\lambda_2} \right)^{\frac{1}{\lambda_2} - 3} \\
&\quad \cdot \left((\lambda_1 - \lambda_2)(1 - z_1 - z_2)^{-\lambda_2} + (\lambda_1 + \lambda_2 - 1) \left(z_1^{-\lambda_1} + z_2^{-\lambda_1} \right)^{\frac{\lambda_2}{\lambda_1}} \right), \\
\phi_{\lambda_1,\lambda_2}(z_1, z_2) &= (\lambda_2 - 1)\,(z_1 z_2)^{\lambda_1 - 1}\,(1 - z_1 - z_2)^{\lambda_2 - 1} \left(z_1^{\lambda_1} + z_2^{\lambda_1} \right)^{\frac{\lambda_2}{\lambda_1} - 2} \\
&\quad \cdot \left(\left(z_1^{\lambda_1} + z_2^{\lambda_1} \right)^{\frac{\lambda_2}{\lambda_1}} + (1 - z_1 - z_2)^{\lambda_2} \right)^{\frac{1}{\lambda_2} - 3} \\
&\quad \cdot \left((\lambda_1 - \lambda_2)(1 - z_1 - z_2)^{\lambda_2} + (\lambda_1 + \lambda_2 - 1) \left(z_1^{\lambda_1} + z_2^{\lambda_1} \right)^{\frac{\lambda_2}{\lambda_1}} \right).
\end{aligned}$$

In Section 4.5 of Coles and Tawn (1991, [10]) it is stated that the underlying angular measure ν has all its mass in the interior of $\overline{R_{d-1}}$, i.e., $d^*_{\lambda_1,\ldots,\lambda_{d-1}} = d$, if not $\lambda_i = 1$ for $i = 1, \ldots, d-1$.

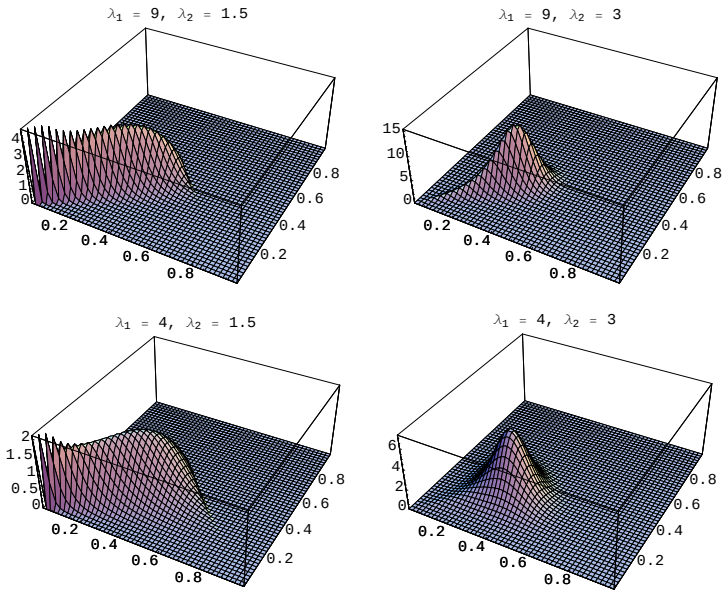
$\Diamond$

Example 2.3.15
In this example we display the angular density and the Pickands density of the nested logistic

GPD for various sets of parameters in $d = 3$. We start with the angular density



and display the Pickands density with the same parameters:



In both cases the first parameter determines the spreading along the viewing axes, the second one the spreading along the orthogonal direction. For small λ_2 the angular density converges to ∞ at the origin and the line $z_1 + z_2 = 1$.

Observe also that for small λ_2 the Pickands density is unbounded when approaching the origin, see for example the graphic with $\lambda_1 = 4$ and $\lambda_2 = 1.5$. This seems always to be the case when $\lambda_{d-1} < d - 1$ and all parameters are pairwise different. This relation is easy to see in the trivariate case but is an open problem to show in the general multivariate case.

$\Diamond$

2.3.4 The Generalized Pareto Distribution of Asymmetric Logistic Type

The family of asymmetric logistic distributions was first introduced in Tawn (1990, [73]) for the extreme value case. It is derived there as a limiting distribution of componentwise maxima of storms recorded at different locations along a coastline and is also a generalization of the logistic model but does not include the nested logistic model from the previous section.

Definition 2.3.16
Let $B := \mathcal{P}(\{1,\dots,d\})\backslash\{\emptyset\}$ be the power set of $\{1,\dots,d\}$ containing all non-empty subsets, and let $\lambda_\Gamma \geq 1$ be arbitrary numbers for every $\Gamma \in B$ with $|\Gamma| > 1$ and $\lambda_\Gamma = 1$ for $|\Gamma| = 1$. Furthermore, let $0 \leq \psi_{j,\Gamma} \leq 1$, where $\psi_{j,\Gamma} = 0$ if $j \notin \Gamma$ and the side condition $\sum_{\Gamma\in B} \psi_{j,\Gamma} = 1$ is fulfilled for $j = 1,\dots,d$. Then the distribution function

$$W_{as}(x_1,\dots,x_d) := 1 - \sum_{\Gamma\in B} \left\{ \sum_{j\in\Gamma} (-\psi_{j,\Gamma} x_j)^{\lambda_\Gamma} \right\}^{1/\lambda_\Gamma} \tag{2.22}$$

for $x_i < 0$, $i = 1,\dots,d$, close to 0 is called the *generalized Pareto distribution of asymmetric logistic type*.

$\Diamond$

Remark 2.3.17
Due to the side conditions for the $\psi_{j,\Gamma}$ we have in this model $2^{d-1}(d+2)-(2d+1)$ free parameters, $2^d - d - 1$ for the various λ_Γ and the rest for the $\psi_{j,\Gamma}$, see Section 2 in Stephenson (2003, [70]). In the case $\psi_{j,\{1,\dots,d\}} = 1$ for $j = 1,\dots,d$ and $\lambda = \lambda_\Gamma \geq 1$ we have again the (symmetric) logistic distribution.

This model has in general $d^* < d$, i.e., the angular measure ν has mass on the boundary of $\overline{R_{d-1}}$. It has, using the abbreviation $\Delta := \{1,\dots,d\}$, the angular density

$$\begin{aligned} l_{as}(z_1,\dots,z_{d-1}) &= \left(\prod_{i=1}^{d-1}(i\lambda_\Delta - 1)\right) \cdot \left(\prod_{i=1}^{d}\psi_{i,\Delta}\right)^{\lambda_\Delta} \cdot \left(\prod_{i=1}^{d-1} z_i\right)^{-\lambda_\Delta-1} \left(1 - \sum_{i=1}^{d-1} z_i\right)^{-\lambda_\Delta-1} \\ &\quad \cdot \left(\sum_{i=1}^{d-1} \left(\frac{\psi_{i,\Delta}}{z_i}\right)^{\lambda_\Delta} + \left(\frac{\psi_{d,\Delta}}{1-\sum_{i=1}^{d-1} z_i}\right)^{\lambda_\Delta} \right)^{\frac{1}{\lambda_\Delta}-d}, \end{aligned}$$

and

$$d^* = \sum_{j=1}^{d} \psi_{j,\Delta},$$

see Sections 3.4 and 3.5 in Kotz and Nadarajah (2000, [47]).

With $d = 2$ and the short notations $\psi_1 := \psi_{1,\{1,2\}}$, $\psi_2 := \psi_{2,\{1,2\}}$, $\lambda := \lambda_{\{1,2\}}$, formula (2.22) reduces to

$$W_{as}(x_1,x_2) = 1 + (1-\psi_1)x_1 + (1-\psi_2)x_2 - \Big((-\psi_1 x_1)^\lambda + (-\psi_2 x_2)^\lambda\Big)^{1/\lambda},$$

and the angular density becomes

$$\begin{aligned} l_{as}(z) &= (\lambda-1)\cdot(\psi_1\psi_2)^\lambda \cdot z^{-\lambda-1}(1-z)^{-\lambda-1} \cdot \left(\left(\frac{\psi_1}{z}\right)^\lambda + \left(\frac{\psi_2}{1-z}\right)^\lambda\right)^{\frac{1}{\lambda}-2} \\ &= (\lambda-1)\cdot(\psi_1\psi_2)^\lambda \cdot z^{\lambda-2}(1-z)^{\lambda-2} \cdot \Big((\psi_2 z)^\lambda + (\psi_1(1-z))^\lambda\Big)^{\frac{1}{\lambda}-2}. \end{aligned}$$

Here we have $d^* = \psi_1 + \psi_2$.

In the case $d = 3$ we have

$$\begin{aligned} W_{as}(x_1, x_2, x_3) &= 1 + (1 - \psi_1 - \psi_3 - \psi_7)x_1 + (1 - \psi_2 - \psi_5 - \psi_8)x_2 + (1 - \psi_4 - \psi_6 - \psi_9)x_3 \\ &\quad - \left((-\psi_1 x_1)^{\lambda_1} + (-\psi_2 x_2)^{\lambda_1}\right)^{1/\lambda_1} \\ &\quad - \left((-\psi_3 x_1)^{\lambda_2} + (-\psi_4 x_3)^{\lambda_2}\right)^{1/\lambda_2} \\ &\quad - \left((-\psi_5 x_2)^{\lambda_3} + (-\psi_6 x_3)^{\lambda_3}\right)^{1/\lambda_3} \\ &\quad - \left((-\psi_7 x_1)^{\lambda_4} + (-\psi_8 x_2)^{\lambda_4} + (-\psi_9 x_3)^{\lambda_4}\right)^{1/\lambda_4} \end{aligned}$$

with the corresponding short notations for the $\psi_{j,\Gamma}$ and λ_Γ.

$\Diamond$

Lemma 2.3.18
The function

$$w_{as}(x_1, \ldots, x_d) = \left(\prod_{i=1}^{d-1}(i\lambda_\Delta - 1)\right)\left(\prod_{i=1}^{d}\psi_{i,\Delta}\right)^{\lambda_\Delta}\left(\prod_{i=1}^{d}(-x_i)\right)^{\lambda_\Delta - 1}\left(\sum_{j=1}^{d}(-\psi_{j,\Delta}x_j)^{\lambda_\Delta}\right)^{\frac{1}{\lambda_\Delta} - d}$$

is the density of W_{as} for $x_i < 0, i = 1, \ldots, d$ and x close to 0.

Proof:
In the case $\lambda_\Delta = 1$ the derivative $\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W_{as}$ reduces to 0 as in Theorem 2.3.1, which is covered by w_{as}. This is again the independence case. Therefore, we assume from now on $\lambda_\Delta > 1$.

First we show by induction that the equation

$$\begin{aligned} &\frac{\partial^j}{\partial x_1 \ldots \partial x_j} W_{as}(x_1, \ldots, x_d) \qquad (2.23) \\ &= \sum_{\Gamma \in B, \{1,\ldots,j\} \subseteq \Gamma} \left(\prod_{i=1}^{j-1}(i\lambda_\Gamma - 1)\right)\left(\prod_{i=1}^{j}\psi_{i,\Gamma}\right)^{\lambda_\Gamma}\left(\prod_{i=1}^{j}(-x_i)\right)^{\lambda_\Gamma - 1} \cdot \left(\sum_{i \in \Gamma}(-\psi_{i,\Gamma}x_i)^{\lambda_\Gamma}\right)^{\frac{1}{\lambda_\Gamma} - j} \end{aligned}$$

holds for $j = 1, \ldots, d$.

For the zero step of the induction ($j = 1$) we have

$$\begin{aligned} &\frac{\partial}{\partial x_1} W_{as}(x_1, \ldots, x_d) \\ &= -\sum_{\Gamma \in B, \{1\} \subseteq \Gamma} \frac{1}{\lambda_\Gamma}\left(\sum_{i \in \Gamma}(-\psi_{i,\Gamma}x_i)^{\lambda_\Gamma}\right)^{\frac{1}{\lambda_\Gamma} - 1} \lambda_\Gamma \left(-\psi_{1,\Gamma}x_1\right)^{\lambda_\Gamma - 1}\left(-\psi_{1,\Gamma}\right) \\ &= \sum_{\Gamma \in B, \{1\} \subseteq \Gamma} \left(\prod_{i=1}^{1-1}(i\lambda_\Gamma - 1)\right)\left(\prod_{i=1}^{1}(\psi_{1,\Gamma})\right)^{\lambda_\Gamma}\left(\prod_{i=1}^{1}(-x_1)\right)^{\lambda_\Gamma - 1}\left(\sum_{i \in \Gamma}(-\psi_{i,\Gamma}x_i)^{\lambda_\Gamma}\right)^{\frac{1}{\lambda_\Gamma} - 1}. \end{aligned}$$

Assume that the assertion is shown for $j - 1$. Then

$$\frac{\partial^j}{\partial x_1 \ldots \partial x_j} W_{as}(x_1, \ldots, x_d)$$

$$
\begin{aligned}
&= \frac{\partial}{\partial x_j}\frac{\partial^{j-1}}{\partial x_1 \dots \partial x_{j-1}} W_{as}(x_1,\dots,x_d) \\
&= \frac{\partial}{\partial x_j}\left[\sum_{\Gamma\in B,\{1,\dots,j-1\}\subseteq\Gamma}\left(\prod_{i=1}^{j-2}(i\lambda_\Gamma-1)\right)\left(\prod_{i=1}^{j-1}\psi_{i,\Gamma}\right)^{\lambda_\Gamma}\left(\prod_{i=1}^{j-1}(-x_i)\right)^{\lambda_\Gamma-1}\right. \\
&\quad \left.\cdot\left(\sum_{i\in\Gamma}(-\psi_{i,\Gamma}x_i)^{\lambda_\Gamma}\right)^{\frac{1}{\lambda_\Gamma}-j+1}\right] \\
&= \sum_{\Gamma\in B,\{1,\dots,j\}\subseteq\Gamma}\left(\prod_{i=1}^{j-2}(i\lambda_\Gamma-1)\right)\left(\prod_{i=1}^{j-1}\psi_{i,\Gamma}\right)^{\lambda_\Gamma}\left(\prod_{i=1}^{j-1}(-x_i)\right)^{\lambda_\Gamma-1} \\
&\quad \cdot\frac{\partial}{\partial x_j}\left[\left(\sum_{i\in\Gamma}(-\psi_{i,\Gamma}x_i)^{\lambda_\Gamma}\right)^{\frac{1}{\lambda_\Gamma}-j+1}\right] \\
&= \sum_{\Gamma\in B,\{1,\dots,j\}\subseteq\Gamma}\left(\prod_{i=1}^{j-2}(i\lambda_\Gamma-1)\right)\left(\prod_{i=1}^{j-1}\psi_{i,\Gamma}\right)^{\lambda_\Gamma}\left(\prod_{i=1}^{j-1}(-x_i)\right)^{\lambda_\Gamma-1} \\
&\quad \cdot\left(\frac{1}{\lambda_\Gamma}-j+1\right)\left(\sum_{i\in\Gamma}(-\psi_{i,\Gamma}x_i)^{\lambda_\Gamma}\right)^{\frac{1}{\lambda_\Gamma}-j}\lambda_\Gamma(-\psi_{j,\Gamma}x_j)^{\lambda_\Gamma-1}(-\psi_{j,\Gamma}) \\
&= \sum_{\Gamma\in B,\{1,\dots,j\}\subseteq\Gamma}\left(\prod_{i=1}^{j-1}(i\lambda_\Gamma-1)\right)\left(\prod_{i=1}^{j}\psi_{i,\Gamma}\right)^{\lambda_\Gamma}\left(\prod_{i=1}^{j}(-x_i)\right)^{\lambda_\Gamma-1}\cdot\left(\sum_{i\in\Gamma}(-\psi_{i,\Gamma}x_i)^{\lambda_\Gamma}\right)^{\frac{1}{\lambda_\Gamma}-j}.
\end{aligned}
$$

With Theorem A.2.2 of Bhattacharya and Rao (1976, [4]) and by setting $j = d$ in (2.23), one gets the assertion, since the leading sum then consists only of one summand.

□

Note that in the density in Lemma 2.3.18 only the parameters with the set $\Delta = \{1,\dots,d\}$ in the index occur. In contrast to the extreme value case the lower hierarchical parameters do not play a role close to the origin.

The next corollary follows from Lemma 2.3.18 and is another main result of this manuscript. To our knowledge it has not been mentioned anywhere else.

Corollary 2.3.19
Let W_1 and W_2 be GPDs. The fact that there exists a neighborhood U of 0 (in the relative topology of the negative quadrant), such that

$$P_{W_1}(B) = P_{W_2}(B)$$

for all Borel sets $B \subseteq U$ does not necessarily entail that W_1 and W_2 and correspondingly ν_1 and ν_2 are identical.

Proof:
We will show the assertion by looking at the case $d = 3$. Let W_1 and W_2 be two trivariate GPDs of asymmetric logistic type with identical parameters ψ_7, ψ_8, ψ_9 and λ_4 in the notation of Remark 2.3.17 but with different parameter λ_1. Then we know by Lemma 2.3.18 that W_1 and W_2 have the same density close to the origin, i.e., $P_{W_1}(B) = P_{W_2}(B)$ for all Borel sets B close

to the origin. Let $G_i = \exp(W_i - 1)$, $i = 1, 2$ be the corresponding extreme value distributions. The angular measures ν_1 and ν_2 belonging to G_1 and G_2 and, thus, W_1 and W_2 are given in Section 3.5.1 of Kotz and Nadarajah (2000, [47]) in terms of their measure densities. These measure densities depend (on the lower boundaries of R_2) on the parameter λ_1 and are, thus, different for different λ_1, leading to $\nu_1 \neq \nu_2$.

□

When modelling exceedances with GPDs it is in some cases, thus, possible by Corollary 2.3.19 to do this with different GPDs, which lead however to the same results, since only the area around the origin is of interest. For distributions $F_i \in \mathcal{D}(G_i)$, $i = 1, 2$ from the domains of attraction of different EVDs $G_1 \neq G_2$ it may be possible to model exceedances over high thresholds of F_1 by W_2 and vice versa.

Remark that the difference between the two angular measures in the proof of Corollary 2.3.19 lay in the lower dimensional boundaries of the unit simplex. Both measures agreed in the interior and, thus, had the same angular density.

The proof of Corollary 2.3.19 can also be done with dimensions $d > 3$, since it only needed free lower hierarchical parameters. Since these exist in the asymmetric logistic case only for $d > 2$, we again see an example of the fact that things tend to get complicated or even sometimes strange, when going from dimension $d = 2$ to dimension $d = 3$.

Theorem 2.3.20
The Pickands density for the asymmetric logistic distribution is

$$\begin{aligned}\phi_{as}(z_1, \ldots, z_{d-1}) &= \left(\prod_{i=1}^{d-1}(i\lambda_\Delta - 1)\right)\left(\prod_{i=1}^{d}\psi_{i,\Delta}\right)^{\lambda_\Delta}\left(\prod_{i=1}^{d-1} z_i\right)^{\lambda_\Delta - 1}\left(1 - \sum_{i=1}^{d-1} z_i\right)^{\lambda_\Delta - 1} \\ &\quad \cdot \left(\sum_{j=1}^{d-1}(\psi_{j,\Delta} z_j)^{\lambda_\Delta} + \left(\psi_{d,\Delta}\left(1 - \sum_{j=1}^{d-1} z_j\right)\right)^{\lambda_\Delta}\right)^{\frac{1}{\lambda_\Delta} - d}.\end{aligned}$$

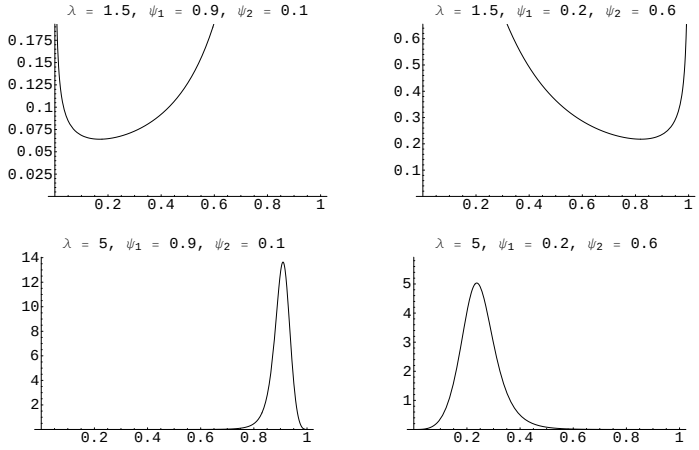
Proof:
Using the density from Lemma 2.3.18 and the definition of the Pickands density from Definition 2.2.6 we get

$$\begin{aligned}&\phi_{as}(z_1, \ldots, z_{d-1}) = |c|^{d-1} w_{as}(T_P^{-1}(z_1, \ldots, z_{d-1}, c)) \\ &= |c|^{d-1}\left(\prod_{i=1}^{d-1}(i\lambda_\Delta - 1)\right)\left(\prod_{i=1}^{d}\psi_{i,\Delta}\right)^{\lambda_\Delta}\left(\prod_{i=1}^{d-1}(-cz_i)\right)^{\lambda_\Delta - 1}\left(-c\left(1 - \sum_{i=1}^{d-1} z_i\right)\right)^{\lambda_\Delta - 1} \\ &\quad \cdot \left(\sum_{j=1}^{d-1}(-\psi_{j,\Delta} c z_j)^{\lambda_\Delta} + \left(-\psi_{d,\Delta} c\left(1 - \sum_{j=1}^{d-1} z_j\right)\right)^{\lambda_\Delta}\right)^{\frac{1}{\lambda_\Delta} - d} \\ &= |c|^{d-1+(d-1)(\lambda_\Delta - 1)+\lambda_\Delta - 1 + \lambda_\Delta\left(\frac{1}{\lambda_\Delta} - d\right)}\left(\prod_{i=1}^{d-1}(i\lambda_\Delta - 1)\right)\left(\prod_{i=1}^{d}\psi_{i,\Delta}\right)^{\lambda_\Delta}\left(\prod_{i=1}^{d-1} z_i\right)^{\lambda_\Delta - 1} \\ &\quad \cdot \left(1 - \sum_{i=1}^{d-1} z_i\right)^{\lambda_\Delta - 1}\left(\sum_{j=1}^{d-1}(\psi_{j,\Delta} z_j)^{\lambda_\Delta} + \left(\psi_{d,\Delta}\left(1 - \sum_{j=1}^{d-1} z_j\right)\right)^{\lambda_\Delta}\right)^{\frac{1}{\lambda_\Delta} - d}\end{aligned}$$

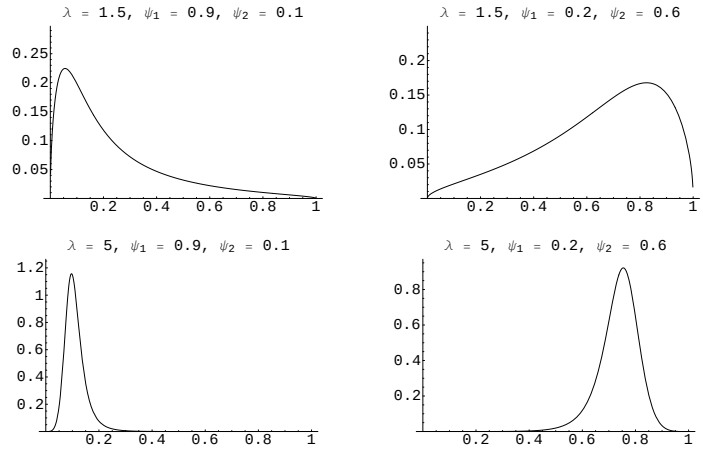
$$= \left(\prod_{i=1}^{d-1}(i\lambda_\Delta - 1)\right)\left(\prod_{i=1}^{d}\psi_{i,\Delta}\right)^{\lambda_\Delta}\left(\prod_{i=1}^{d-1}z_i\right)^{\lambda_\Delta-1}\left(1-\sum_{i=1}^{d-1}z_i\right)^{\lambda_\Delta-1}$$
$$\cdot\left(\sum_{j=1}^{d-1}(\psi_{j,\Delta}z_j)^{\lambda_\Delta} + \left(\psi_{d,\Delta}\left(1-\sum_{j=1}^{d-1}z_j\right)\right)^{\lambda_\Delta}\right)^{\frac{1}{\lambda_\Delta}-d}.$$

□

Example 2.3.21
In this example we want to illustrate the angular density and the Pickands density of the asymmetric logistic model for various sets of parameters. We start with the bivariate angular density

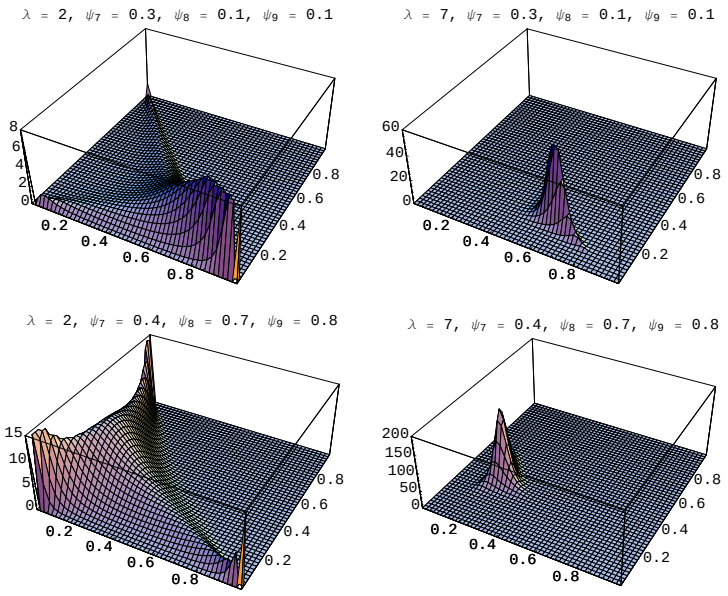


and then show the bivariate Pickands density with the same parameters.

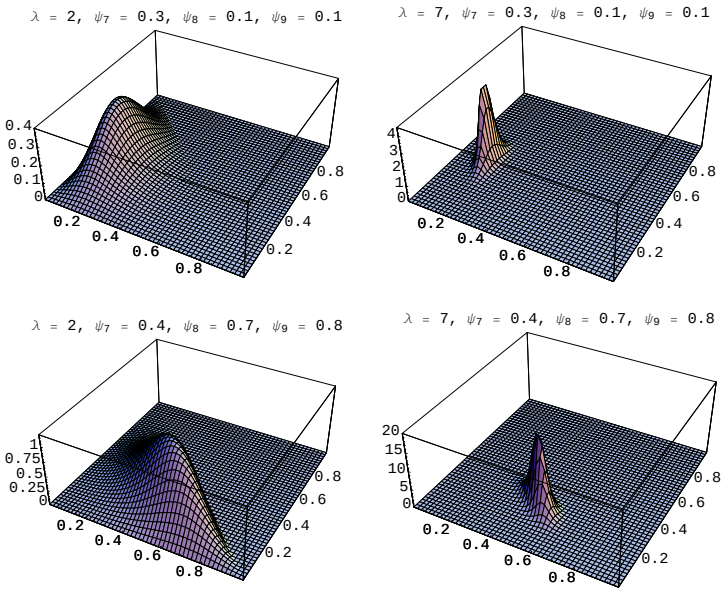


This is a non exchangeable model, which can be clearly seen in the pictures since they are not symmetric to $\frac{1}{2}$. The parameters ψ_1 and ψ_2 determine the center of the corresponding density, the parameter λ its shape.

We also give the angular densities for the dimension $d = 3$ for selected parameter values



and the corresponding Pickands densities.



◊

Chapter 3

Simulation in Multivariate Generalized Pareto Models

Many problems arising in complex practical situations cannot be solved by analytical methods. Then one has to rely on Monte-Carlo methods, which use simulations. One such example, estimating the Value-at-Risk of portfolios, can be found in Section 7 of Brommundt (2003, [5]). If one wanted to use GPD distributed random vectors for the tails of the distributions, instead of the elliptical distributions used there, one would have to generate random vectors following a multivariate GPD. The aim of this section is to complete this task.

Simulations can also be used as checks for new statistical testing or estimation procedures before applying them to real data. They are useful in this situation, since one controls the entire setup there, whereas the background of real data is commonly unknown. This will be a frequent application of the simulation algorithms derived in this chapter.

We will start in Section 3.1 with a special case, where we will present an efficient simulation algorithm for GPDs of logistic type using the Shi transformation. An analogous algorithm simulating EVDs of logistic type has been given in Stephenson (2003, [70]).

In Section 3.2 we will consider a method for a wider range of GPDs, using standard Pickands coordinates. But this method will not be as efficient as the first one. It is comparable to a method of simulating EVDs presented in Section 3 of Ghoudi et al. (1998, [32]), although we also formulate it for higher dimensions than the bivariate case. Another comparable algorithm for the EVD case is to be found in Section 2 of Nadarajah (1999, [51]) where, however, the random vectors are simulated from a limiting and not a direct representation.

Finally in Section 3.3 we will simulate relative frequencies of GPD distributed random vectors falling into certain sets. The application of this algorithm will be necessary in Section 6.3.

3.1 Simulation of Generalized Pareto Distributions of Logistic Type

The aim of this section is to simulate generalized Pareto distributions of logistic type, which were defined in Section 2.3.2. Shi (1995, [64]) proposed a transformation that we call Shi transformation, which was used by Stephenson (2003, [70]) to simulate random vectors following a multivariate EVD of logistic type. In Theorem 5.4.2 of Falk et al. (2004, [21]) the conditional

independence of the Pickands coordinates of GPDs was shown in a certain neighborhood of 0, see also Theorem 2.2.5. Both approaches will be joined to formulate a simulation algorithm for GPDs of logistic type.

In Section 3.1.1 we will introduce the Shi transformation and the Shi coordinates, which will turn out to be the crucial tool of the simulation algorithm. Afterwards we show in Section 3.1.2 the conditional independence of the Shi coordinates in logistic GPD models and give a straightforward and easy to implement simulation algorithm in Section 3.1.3.

3.1.1 The Shi Transformation

As before, the dimension d is always a number larger than 1 throughout this text.

Definition 3.1.1
The mapping $P : (0,\infty) \times \left(0, \frac{\pi}{2}\right)^{d-1} \to (0,\infty)^d$ with

$$P(r, \psi_1, \ldots, \psi_{d-1}) = r \left(\cos\psi_1, \cos\psi_2 \sin\psi_1, \ldots, \cos\psi_{d-1} \prod_{j=1}^{d-2} \sin\psi_j, \prod_{j=1}^{d-1} \sin\psi_j \right)$$

is the *polar transformation*, and its inverse defines the *polar coordinates* r, $\psi = (\psi_1, \ldots, \psi_{d-1})$ in $(0,\infty)^d$.

$\Diamond$

Remark 3.1.2
The following facts are well known (see for example Mardia et al. (1979, [50]), Section 2.4): the mapping P is one-to-one, infinitely often differentiable and

$$1 = ||P(1, \psi_1, \ldots, \psi_{d-1})||_2^2 = \sum_{i=1}^{d-1} \cos^2\psi_i \prod_{j=1}^{i-1} \sin^2\psi_j + \prod_{j=1}^{d-1} \sin^2\psi_j, \tag{3.1}$$

i.e., for $r = 1$ the function $P(1, \psi_1, \ldots, \psi_{d-1})$ is a one-to-one mapping from the cube $\left(0, \frac{\pi}{2}\right)^{d-1}$ in $\mathbb{R}^{d-1}$ onto the intersection of $(0,\infty)^d$ with the unit sphere of $\mathbb{R}^d$ with respect to the Euclidian $||\cdot||_2$–norm.

$\Diamond$

Lemma 3.1.3
The mapping $T : \left(0, \frac{\pi}{2}\right)^{d-1} \to (0,\infty)^d$ with

$$\begin{aligned} T(\psi_1, \ldots, \psi_{d-1}) &:= (T_1(\psi_1), T_2(\psi_1, \psi_2), \ldots, T_{d-1}(\psi_1, \ldots, \psi_{d-1}), T_d(\psi_1, \ldots, \psi_{d-1})) \\ &:= \left(\cos^2\psi_1, \cos^2\psi_2 \sin^2\psi_1, \ldots, \cos^2\psi_{d-1} \prod_{j=1}^{d-2} \sin^2\psi_j, \prod_{j=1}^{d-1} \sin^2\psi_j \right) \end{aligned}$$

maps the cube $\left(0, \frac{\pi}{2}\right)^{d-1}$ one-to-one and infinitely often differentiable onto the simplex

$$S_d := \left\{ x \in (0,\infty)^d \,\middle|\, \sum_{i=1}^{d} x_i = 1 \right\},$$

i.e., to the unit circle in $(0,\infty)^d$ with regard to the $||\cdot||_1$–norm.

Proof:
The function $x \mapsto x^2$ maps the interval $(0,1)$ one-to-one onto itself, thus the bijectivity and differentiability of T follow from the corresponding properties of the polar transformation in Remark 3.1.2. Let

$$\begin{aligned} x &= (x_1,\ldots,x_d) = T(\psi_1,\ldots,\psi_{d-1}) \\ &= (T_1(\psi_1), T_2(\psi_1,\psi_2),\ldots,T_{d-1}(\psi_1,\ldots,\psi_{d-1}), T_d(\psi_1,\ldots,\psi_{d-1}))\,. \end{aligned}$$

Then we have because of (3.1)

$$\begin{aligned} \sum_{i=1}^{d} x_i &= \sum_{i=1}^{d-1} T_i(\psi_1,\ldots,\psi_i) + T_d(\psi_1,\ldots,\psi_{d-1}) = \sum_{i=1}^{d-1} \cos^2\psi_i \prod_{j=1}^{i-1}\sin^2\psi_j + \prod_{j=1}^{d-1}\sin^2\psi_j \\ &= ||P(1,\psi_1,\ldots,\psi_{d-1})||_2^2 = 1. \end{aligned}$$

□

Remark 3.1.4
Let $x = (x_1,\ldots,x_d) \in (-\infty,0)^d$ and $\lambda \geq 1$. Then

$$x = -||x||_\lambda \left(\frac{-x_1}{||x||_\lambda},\ldots,\frac{-x_d}{||x||_\lambda}\right)$$

holds. We have

$$\left|\left|\left(\frac{-x_1}{||x||_\lambda},\ldots,\frac{-x_d}{||x||_\lambda}\right)\right|\right|_\lambda = \left(\sum_{i=1}^{d}\frac{(-x_i)^\lambda}{||x||_\lambda^\lambda}\right)^{1/\lambda} = \frac{1}{||x||_\lambda}\underbrace{\left(\sum_{i=1}^{d}(-x_i)^\lambda\right)^{1/\lambda}}_{=||x||_\lambda} = 1 \tag{3.2}$$

and, thus,

$$\left|\left|\left(\frac{(-x_1)^\lambda}{||x||_\lambda^\lambda},\ldots,\frac{(-x_d)^\lambda}{||x||_\lambda^\lambda}\right)\right|\right|_1 = \sum_{i=1}^{d}\frac{(-x_i)^\lambda}{||x||_\lambda^\lambda} = \left(\left(\sum_{i=1}^{d}\frac{(-x_i)^\lambda}{||x||_\lambda^\lambda}\right)^{1/\lambda}\right)^\lambda \overset{(3.2)}{=} 1^\lambda = 1.$$

It follows that

$$\left(\frac{(-x_1)^\lambda}{||x||_\lambda^\lambda},\ldots,\frac{(-x_d)^\lambda}{||x||_\lambda^\lambda}\right) \in S_d,$$

and this point has a representation with regard to the transformation T. More precisely, there exist uniquely determined $(\psi_1,\ldots,\psi_{d-1}) \in \left(0,\frac{\pi}{2}\right)^{d-1}$ with

$$\left(\frac{(-x_1)^\lambda}{||x||_\lambda^\lambda},\ldots,\frac{(-x_d)^\lambda}{||x||_\lambda^\lambda}\right) = \left(\cos^2\psi_1, \cos^2\psi_2\sin^2\psi_1,\ldots,\cos^2\psi_{d-1}\prod_{j=1}^{d-2}\sin^2\psi_j, \prod_{j=1}^{d-1}\sin^2\psi_j\right).$$

This is equivalent to

$$\left(\frac{x_1}{||x||_\lambda},\ldots,\frac{x_d}{||x||_\lambda}\right) =$$

$$= \left(-\cos^{2/\lambda}\psi_1, -\cos^{2/\lambda}\psi_2\sin^{2/\lambda}\psi_1,\ldots,-\cos^{2/\lambda}\psi_{d-1}\prod_{j=1}^{d-2}\sin^{2/\lambda}\psi_j, -\prod_{j=1}^{d-1}\sin^{2/\lambda}\psi_j\right).$$

By putting $c := ||x||_\lambda$ one arrives at

$$\begin{aligned} x &= c\left(-\cos^{2/\lambda}\psi_1, -\cos^{2/\lambda}\psi_2 \sin^{2/\lambda}\psi_1, \ldots, -\cos^{2/\lambda}\psi_{d-1}\prod_{j=1}^{d-2}\sin^{2/\lambda}\psi_j, -\prod_{j=1}^{d-1}\sin^{2/\lambda}\psi_j\right) \\ &= c\left(-T_1^{1/\lambda}(\psi_1), -T_2^{1/\lambda}(\psi_1,\psi_2), \ldots, -T_{d-1}^{1/\lambda}(\psi_1,\ldots,\psi_{d-1}), -T_d^{1/\lambda}(\psi_1,\ldots,\psi_{d-1})\right). \end{aligned}$$

$\Diamond$

Definition 3.1.5
The transformation described in Remark 3.1.4

$$\begin{aligned} x &= (x_1,\ldots,x_d) = ST_\lambda(c,\psi_1,\ldots,\psi_{d-1}) \\ &:= -c\left(\cos^{2/\lambda}\psi_1, \cos^{2/\lambda}\psi_2 \sin^{2/\lambda}\psi_1, \ldots, \cos^{2/\lambda}\psi_{d-1}\prod_{j=1}^{d-2}\sin^{2/\lambda}\psi_j, \prod_{j=1}^{d-1}\sin^{2/\lambda}\psi_j\right) \end{aligned}$$

is called the *Shi transformation* ST_λ. The transformation $ST_\lambda : (0,\infty)\times\left(0,\frac{\pi}{2}\right)^{d-1} \to (-\infty,0)^d$ is one-to-one and infinitely often differentiable.

Let $(x_1,\ldots,x_d) \in (-\infty,0)^d$. The components of the vector

$$(c,\psi_1,\ldots,\psi_{d-1}) := ST_\lambda^{-1}(x_1,\ldots,x_d)$$

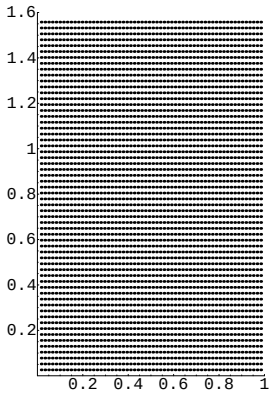
are the *Shi coordinates* of $(x_1,\ldots,x_d)$. c is called the *radial component* and $\psi := (\psi_1,\ldots,\psi_{d-1})$ is called the *angular component*.

By $(C,\Psi_1,\ldots,\Psi_{d-1}) = ST_\lambda^{-1}(X_1,\ldots,X_d)$ we denote the Shi coordinates of a random vector $(X_1,\ldots,X_d) \in (-\infty,0)^d$. Remark that $C = ||X||_\lambda$.

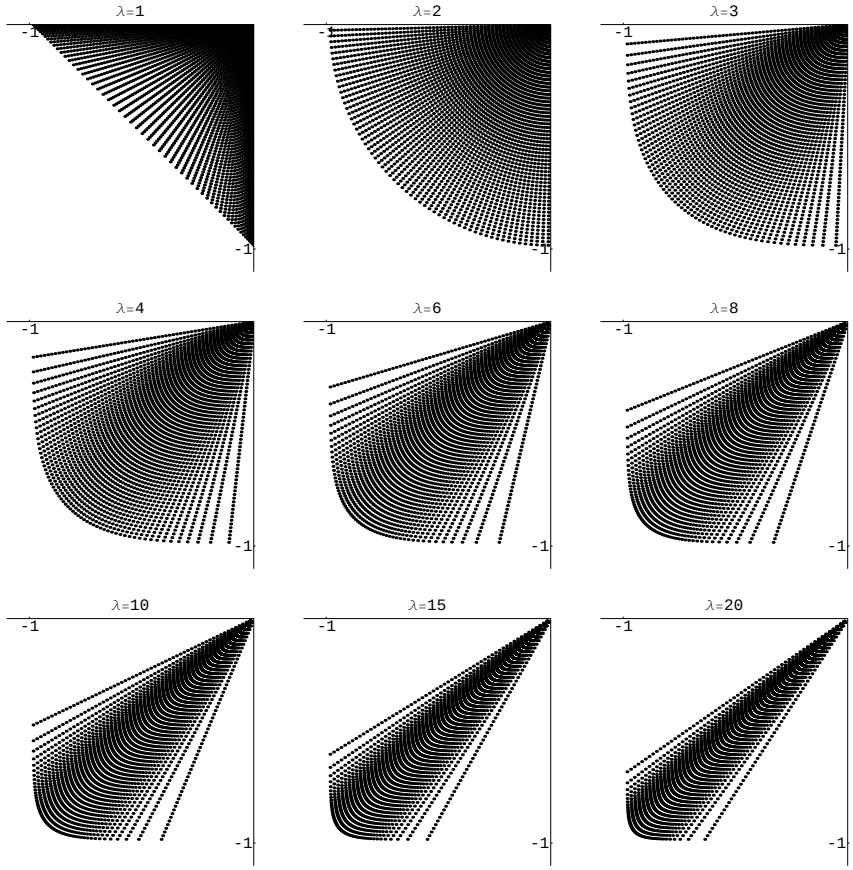
$\Diamond$

Note that in the case $\lambda = 2$ the Shi transformation is up to the sign the polar transformation from Definition 3.1.1. For $\lambda = 1$ we have a variant of the inverse Pickands transformation (2.2), where the angular component has an additional parametrization with regard to the cube $\left(0,\frac{\pi}{2}\right)^{d-1}$.

Example 3.1.6
For the illustration of the mapping properties of ST_λ some plots were created for $d = 2$. The rectangle $(0,1)\times(0,\pi/2)$, which was covered by a grid and which can be seen in the following plot,



is mapped one-to-one by ST_λ onto the unit ball with regard to the $||\cdot||_\lambda$–norm. In the following graphic, one can see the images of the grid points of the rectangle shown above under the Shi transformation for different λ:



Note that the border areas of $(0,1)\times(0,\pi/2)$ are stretched with increasing λ, whereas the interior areas are compressed.

$\Diamond$

Lemma 3.1.7
Put $\gamma_i := \cos\psi_i$, $\sigma_i := \sin\psi_i$ for $i = 1,\ldots,d-1$ and $\alpha := \frac{2}{\lambda}$. Then the Jacobian matrix of the transformation ST_λ has the following form

$$A_{1,\ldots,d} := J_{ST_\lambda}(c,\psi_1,\ldots,\psi_{d-1}) =$$

$$= \begin{pmatrix}
-\gamma_1^\alpha & \alpha c\gamma_1^\alpha\frac{\sigma_1}{\gamma_1} & 0 & \cdots & 0 \\
-\gamma_2^\alpha\sigma_1^\alpha & -\alpha c\gamma_2^\alpha\sigma_1^\alpha\frac{\gamma_1}{\sigma_1} & \alpha c\gamma_2^\alpha\sigma_1^\alpha\frac{\sigma_2}{\gamma_2} & & 0 \\
-\gamma_3^\alpha\prod_{i=1}^2\sigma_i^\alpha & -\alpha c\gamma_3^\alpha\prod_{i=1}^2\sigma_i^\alpha\frac{\gamma_1}{\sigma_1} & -\alpha c\gamma_3^\alpha\prod_{i=1}^2\sigma_i^\alpha\frac{\gamma_2}{\sigma_2} & \ddots & 0 \\
-\gamma_4^\alpha\prod_{i=1}^3\sigma_i^\alpha & -\alpha c\gamma_4^\alpha\prod_{i=1}^3\sigma_i^\alpha\frac{\gamma_1}{\sigma_1} & -\alpha c\gamma_4^\alpha\prod_{i=1}^3\sigma_i^\alpha\frac{\gamma_2}{\sigma_2} & & 0 \\
\vdots & \vdots & \vdots & & \vdots \\
\vdots & \vdots & \vdots & & 0 \\
-\gamma_{d-1}^\alpha\prod_{i=1}^{d-2}\sigma_i^\alpha & -\alpha c\gamma_{d-1}^\alpha\prod_{i=1}^{d-2}\sigma_i^\alpha\frac{\gamma_1}{\sigma_1} & -\alpha c\gamma_{d-1}^\alpha\prod_{i=1}^{d-2}\sigma_i^\alpha\frac{\gamma_2}{\sigma_2} & \cdots & \alpha c\gamma_{d-1}^\alpha\prod_{i=1}^{d-2}\sigma_i^\alpha\frac{\sigma_{d-1}}{\gamma_{d-1}} \\
-\prod_{i=1}^{d-1}\sigma_i^\alpha & -\alpha c\prod_{i=1}^{d-1}\sigma_i^\alpha\frac{\gamma_1}{\sigma_1} & -\alpha c\prod_{i=1}^{d-1}\sigma_i^\alpha\frac{\gamma_2}{\sigma_2} & \cdots & -\alpha c\prod_{i=1}^{d-1}\sigma_i^\alpha\frac{\gamma_{d-1}}{\sigma_{d-1}}
\end{pmatrix}.$$

Proof:
The Shi transformation is

$$x_i = -cT_i^{1/\lambda}(\psi_1, \ldots, \psi_i) \text{ for } i \leq d-1 \qquad \text{and} \qquad x_d = -cT_d^{1/\lambda}(\psi_1, \ldots, \psi_{d-1}). \tag{3.3}$$

For a shorter notation we will write $x_i = -cT_i^{1/\lambda}(\psi_1, \ldots, \psi_i)$ from now on as an abbreviation for (3.3), disregarding that a variable ψ_d does not exist in the case $i = d$.

Since

$$\frac{\partial x_i}{\partial c} = -T_i^{1/\lambda}(\psi_1, \ldots, \psi_i),$$

the first column of the Jacobian matrix

$$A_{1,\ldots,d} = \begin{pmatrix} \frac{\partial x_1}{\partial c} & \frac{\partial x_1}{\partial \psi_1} & \frac{\partial x_1}{\partial \psi_2} & \cdots & \frac{\partial x_1}{\partial \psi_{d-1}} \\ \frac{\partial x_2}{\partial c} & \frac{\partial x_2}{\partial \psi_1} & \frac{\partial x_2}{\partial \psi_2} & \cdots & \frac{\partial x_2}{\partial \psi_{d-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_d}{\partial c} & \frac{\partial x_d}{\partial \psi_1} & \frac{\partial x_d}{\partial \psi_2} & \cdots & \frac{\partial x_d}{\partial \psi_{d-1}} \end{pmatrix}$$

is obvious. It is also obvious that for $i < j$ one has

$$\frac{\partial x_i}{\partial \psi_j} = 0,$$

since the corresponding terms do not depend on ψ_j. For the last row we have

$$\begin{aligned} \frac{\partial x_d}{\partial \psi_j} &= \frac{\partial}{\partial \psi_j}\left(-c\prod_{i=1}^{d-1}\sin^{2/\lambda}\psi_i\right) = -c\left(\prod_{i=1,i\neq j}^{d-1}\sin^{2/\lambda}\psi_i\right)\frac{2}{\lambda}\sin^{2/\lambda-1}\psi_j\cos\psi_j \\ &= -\frac{2}{\lambda}c\left(\prod_{i=1}^{d-1}\sin^{2/\lambda}\psi_i\right)\frac{\cos\psi_j}{\sin\psi_j}. \end{aligned}$$

With the abbreviations given above one can easily see that this is the last row. For the upper secondary diagonal we compute

$$\begin{aligned} \frac{\partial x_j}{\partial \psi_j} &= \frac{\partial}{\partial \psi_j}\left(-c\cos^{2/\lambda}\psi_j\prod_{i=1}^{j-1}\sin^{2/\lambda}\psi_i\right) = -c\left(\prod_{i=1}^{j-1}\sin^{2/\lambda}\psi_i\right)\frac{2}{\lambda}\cos^{2/\lambda-1}\psi_j(-\sin\psi_j) \\ &= \frac{2}{\lambda}c\left(\prod_{i=1}^{j-1}\sin^{2/\lambda}\psi_i\right)\cos^{2/\lambda}\psi_j\frac{\sin\psi_j}{\cos\psi_j}. \end{aligned}$$

Finally for the other elements, i.e., for $k < d$ and $j = 1, \ldots, k-1$ we get

$$\begin{aligned} \frac{\partial x_k}{\partial \psi_j} &= \frac{\partial}{\partial \psi_j}\left(-c\cos^{2/\lambda}\psi_k\prod_{i=1}^{k-1}\sin^{2/\lambda}\psi_i\right) \\ &= -c\cos^{2/\lambda}\psi_k\left(\prod_{i=1,i\neq j}^{k-1}\sin^{2/\lambda}\psi_i\right)\frac{2}{\lambda}\sin^{2/\lambda-1}\psi_j(\cos\psi_j) \\ &= -\frac{2}{\lambda}c\left(\prod_{i=1}^{k-1}\sin^{2/\lambda}\psi_i\right)\cos^{2/\lambda}\psi_k\frac{\cos\psi_j}{\sin\psi_j}. \end{aligned}$$

Thus, all the entries of the Jacobian matrix have been given, and the assertion follows.

□

Lemma 3.1.8
With the abbreviations of Lemma 3.1.7 we have

$$|\det(A_{1,\dots,d})| = (\alpha c)^{d-1} \prod_{i=1}^{d-1} \gamma_i^{\alpha-1} \sigma_i^{\alpha(d-i)-1}.$$

Proof:
We prove the assertion by induction. For $d = 2$ we have the Jacobian matrix

$$A_{1,2} = \begin{pmatrix} -\gamma_1^\alpha & \alpha c \gamma_1^\alpha \frac{\sigma_1}{\gamma_1} \\ -\sigma_1^\alpha & -\alpha c \sigma_1^\alpha \frac{\gamma_1}{\sigma_1} \end{pmatrix},$$

which has the determinant

$$\begin{aligned} |\det(A_{1,2})| &= \left| \gamma_1^\alpha \alpha c \sigma_1^\alpha \frac{\gamma_1}{\sigma_1} + \sigma_1^\alpha \alpha c \gamma_1^\alpha \frac{\sigma_1}{\gamma_1} \right| = \alpha c \gamma_1^\alpha \sigma_1^\alpha \left| \frac{\gamma_1}{\sigma_1} + \frac{\sigma_1}{\gamma_1} \right| = \alpha c \gamma_1^\alpha \sigma_1^\alpha \underbrace{\left| \frac{\gamma_1^2 + \sigma_1^2}{\gamma_1 \sigma_1} \right|}_{= \frac{1}{\gamma_1 \sigma_1}} \\ &= \alpha c \gamma_1^{\alpha-1} \sigma_1^{\alpha-1} = (\alpha c)^{2-1} \prod_{i=1}^{2-1} \gamma_i^{\alpha-1} \sigma_i^{\alpha(2-i)-1}. \end{aligned}$$

Assume now that the assertion holds for all matrices of dimension $d-1$. By expanding the determinant according to the first row and the calculation rules for determinants (Chapter VI, §2 in Lang (1966, [48])), one gets

$$\begin{aligned} &\det(A_{1,\dots,d}) = \\ &= -\gamma_1^\alpha \cdot \det \begin{pmatrix} -\alpha c \gamma_2^\alpha \sigma_1^\alpha \frac{\gamma_1}{\sigma_1} & \alpha c \gamma_2^\alpha \sigma_1^\alpha \frac{\sigma_2}{\gamma_2} & 0 & 0 \\ -\alpha c \gamma_3^\alpha \prod_{i=1}^2 \sigma_i^\alpha \frac{\gamma_1}{\sigma_1} & -\alpha c \gamma_3^\alpha \prod_{i=1}^2 \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & \ddots & 0 \\ -\alpha c \gamma_4^\alpha \prod_{i=1}^3 \sigma_i^\alpha \frac{\gamma_1}{\sigma_1} & -\alpha c \gamma_4^\alpha \prod_{i=1}^3 \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & 0 \\ -\alpha c \gamma_{d-1}^\alpha \prod_{i=1}^{d-2} \sigma_i^\alpha \frac{\gamma_1}{\sigma_1} & -\alpha c \gamma_{d-1}^\alpha \prod_{i=1}^{d-2} \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & \cdots & \alpha c \gamma_{d-1}^\alpha \prod_{i=1}^{d-2} \sigma_i^\alpha \frac{\sigma_{d-1}}{\gamma_{d-1}} \\ -\alpha c \prod_{i=1}^{d-1} \sigma_i^\alpha \frac{\gamma_1}{\sigma_1} & -\alpha c \prod_{i=1}^{d-1} \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & \cdots & -\alpha c \prod_{i=1}^{d-1} \sigma_i^\alpha \frac{\gamma_{d-1}}{\sigma_{d-1}} \end{pmatrix} - \\ &\quad -\alpha c \gamma_1^\alpha \frac{\sigma_1}{\gamma_1} \cdot \det \begin{pmatrix} -\gamma_2^\alpha \sigma_1^\alpha & \alpha c \gamma_2^\alpha \sigma_1^\alpha \frac{\sigma_2}{\gamma_2} & 0 & 0 \\ -\gamma_3^\alpha \prod_{i=1}^2 \sigma_i^\alpha & -\alpha c \gamma_3^\alpha \prod_{i=1}^2 \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & \ddots & 0 \\ -\gamma_4^\alpha \prod_{i=1}^3 \sigma_i^\alpha & -\alpha c \gamma_4^\alpha \prod_{i=1}^3 \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & 0 \\ -\gamma_{d-1}^\alpha \prod_{i=1}^{d-2} \sigma_i^\alpha & -\alpha c \gamma_{d-1}^\alpha \prod_{i=1}^{d-2} \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & \cdots & \alpha c \gamma_{d-1}^\alpha \prod_{i=1}^{d-2} \sigma_i^\alpha \frac{\sigma_{d-1}}{\gamma_{d-1}} \\ -\prod_{i=1}^{d-1} \sigma_i^\alpha & -\alpha c \prod_{i=1}^{d-1} \sigma_i^\alpha \frac{\gamma_2}{\sigma_2} & \cdots & -\alpha c \prod_{i=1}^{d-1} \sigma_i^\alpha \frac{\gamma_{d-1}}{\sigma_{d-1}} \end{pmatrix} \\ &= -\gamma_1^\alpha \alpha c \frac{\gamma_1}{\sigma_1} (\sigma_1^\alpha)^{d-1} \cdot \det(A_{2,\dots,d}) - \alpha c \gamma_1^\alpha \frac{\sigma_1}{\gamma_1} (\sigma_1^\alpha)^{d-1} \cdot \det(A_{2,\dots,d}) \end{aligned}$$

$$
\begin{aligned}
&= -\alpha c \gamma_1^{\alpha} (\sigma_1^{\alpha})^{d-1} \cdot \det(A_{2,\dots,d}) \cdot \underbrace{\left(\frac{\gamma_1}{\sigma_1} + \frac{\sigma_1}{\gamma_1}\right)}_{=\frac{1}{\gamma_1 \sigma_1}} \\
&= -\alpha c \gamma_1^{\alpha-1} \sigma_1^{\alpha(d-1)-1} \cdot \det(A_{2,\dots,d}).
\end{aligned}
$$

Thus, we obtain by induction

$$
\begin{aligned}
|\det(A_{1,\dots,d})| &= \alpha c \gamma_1^{\alpha-1} \sigma_1^{\alpha(d-1)-1} \cdot |\det(A_{2,\dots,d})| \\
&= \alpha c \gamma_1^{\alpha-1} \sigma_1^{\alpha(d-1)-1} \cdot (\alpha c)^{d-2} \prod_{i=2}^{d-1} \gamma_i^{\alpha-1} \sigma_i^{\alpha(d-i)-1} \\
&= (\alpha c)^{d-1} \prod_{i=1}^{d-1} \gamma_i^{\alpha-1} \sigma_i^{\alpha(d-i)-1},
\end{aligned}
$$

which completes the proof.

□

3.1.2 Conditional Independence of the Shi Coordinates

Let

$$B_r^{\lambda} := \{x \in (-\infty, 0)^d : ||x||_{\lambda} < r\}, \qquad r > 0,$$

be the *ball* in $(-\infty, 0)^d$ of radius r with respect to the $||\cdot||_{\lambda}$–norm, centered at the origin.

Lemma 3.1.9
Let $(X_1, \dots, X_d) < 0$ be a random vector, which is distributed according to a GPD W_{λ} of logistic type. Choose a number $c_0 > 0$, such that W_{λ} has on $B_{c_0}^{\lambda}$ the representation

$$W_{\lambda}(x_1, \dots, x_d) = 1 - ||x||_{\lambda}$$

and denote by w_{λ} the density of W_{λ}. Then the inverse Shi transformation $ST_{\lambda}^{-1} : B_{c_0}^{\lambda} \to (0, c_0) \times \left(0, \frac{\pi}{2}\right)^{d-1}$ of the random vector restricted to $B_{c_0}^{\lambda}$ has a density, which is independent of the radial component c and which factorizes with regard to the angular components $\psi_1, \dots, \psi_{d-1}$. More precisely,

$$f(c, \psi_1, \dots, \psi_{d-1}) = \prod_{i=1}^{d-1} \left(2i - \frac{2}{\lambda}\right) \cdot \cos\psi_i \cdot \sin^{2(d-i)-1}\psi_i$$

is the density of $ST_{\lambda}^{-1}(X_1, \dots, X_d)$ on $(0, c_0) \times \left(0, \frac{\pi}{2}\right)^{d-1}$ under the restriction $(X_1, \dots, X_d) \in B_{c_0}^{\lambda}$.

Proof:
According to the density transformation theorem (see for example Fristedt and Gray (1997, [28]), Section 9.5), Lemma 2.3.6 and Lemma 3.1.8, one has

$$
\begin{aligned}
&f(c, \psi_1, \dots, \psi_{d-1}) \\
&= w_{\lambda}(ST_{\lambda}(c, \psi_1, \dots, \psi_{d-1})) \, |\det J_{ST_{\lambda}}(c, \psi_1, \dots, \psi_{d-1})| \\
&= \left(\prod_{i=1}^{d-1}(i\lambda - 1)\right) c^{1-d\lambda} \prod_{i=1}^{d} \left(cT_i^{1/\lambda}(\psi_1, \dots, \psi_i)\right)^{\lambda-1} \left(\frac{2}{\lambda}c\right)^{d-1} \prod_{i=1}^{d-1} \cos^{2/\lambda-1}\psi_i \cdot \sin^{2(d-i)/\lambda-1}\psi_i
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\prod_{i=1}^{d-1}\left(2i-\frac{2}{\lambda}\right)\right) c^{1-d\lambda+d-1+d\lambda-d} \prod_{i=1}^{d}\left(T_i(\psi_1,\ldots,\psi_i)\right)^{1-1/\lambda} \prod_{i=1}^{d-1}\cos^{2/\lambda-1}\psi_i \cdot \sin^{2(d-i)/\lambda-1}\psi_i \\
&= \left(\prod_{i=1}^{d-1}\left(2i-\frac{2}{\lambda}\right)\right) \prod_{i=1}^{d}\left(T_i(\psi_1,\ldots,\psi_i)\right)^{1-1/\lambda} \prod_{i=1}^{d-1}\cos^{2/\lambda-1}\psi_i \cdot \sin^{2(d-i)/\lambda-1}\psi_i.
\end{aligned}
$$

We have now

$$
\begin{aligned}
&\prod_{i=1}^{d}\left(T_i(\psi_1,\ldots,\psi_i)\right)^{1-1/\lambda} \\
&\quad= \left(\cos^2\psi_1\right)^{1-1/\lambda} \cdot \left(\cos^2\psi_2 \sin^2\psi_1\right)^{1-1/\lambda} \cdot \left(\cos^2\psi_3 \sin^2\psi_2 \sin^2\psi_1\right)^{1-1/\lambda} \cdot \ldots \\
&\qquad \ldots \cdot \left(\cos^2\psi_{d-1}\prod_{j=1}^{d-2}\sin^2\psi_j\right)^{1-1/\lambda} \cdot \left(\prod_{j=1}^{d-1}\sin^2\psi_j\right)^{1-1/\lambda} \\
&\quad= \left(\cos^2\psi_1\left(\sin^2\psi_1\right)^{d-1}\right)^{1-1/\lambda} \cdot \left(\cos^2\psi_2\left(\sin^2\psi_2\right)^{d-2}\right)^{1-1/\lambda} \cdot \ldots \\
&\qquad \ldots \cdot \left(\cos^2\psi_{d-2}\left(\sin^2\psi_{d-2}\right)^{2}\right)^{1-1/\lambda} \cdot \left(\cos^2\psi_{d-1}\sin^2\psi_{d-1}\right)^{1-1/\lambda} \\
&\quad= \prod_{i=1}^{d-1}\left(\cos^2\psi_i\left(\sin^2\psi_i\right)^{d-i}\right)^{1-1/\lambda} = \prod_{i=1}^{d-1}\left(\cos^{2-2/\lambda}\psi_i \sin^{2(d-i)-2(d-i)/\lambda}\psi_i\right)
\end{aligned}
$$

and, thus, altogether

$$
f(c,\psi_1,\ldots,\psi_{d-1}) = \prod_{i=1}^{d-1}\left(2i-\frac{2}{\lambda}\right)\cos\psi_i \cdot \sin^{2(d-i)-1}\psi_i,
$$

as asserted.

□

A corresponding result to Lemma 3.1.9 for extreme value distributions of logistic type has been shown in Shi (1995, [64]), Theorem 3. In the extreme value case the result can immediately be used to simulate the corresponding random vectors, which was done in Section 1 of Stephenson (2003, [70]). For our GPD case we still need the following Theorem 3.1.10 to be able to give the simulation algorithm, since a GPD is only explicitly defined close to the origin.

The theorem below is a main result of this manuscript and an analogy to Theorem 5.4.2 from Falk et al. (2004, [21]) for the logistic case and the Shi transformation. Whereas in Falk et al. (2004, [21]) the independence of the angular and the radial component of Pickands coordinates for general GPDs could be shown, we prove here in addition the mutual independence of the angular components. We are also able to specify their distributions more precisely. But we have to pay for that by restricting ourselves to the logistic case.

Theorem 3.1.10
Let $(X_1,\ldots,X_d)$ follow a logistic GPD with parameter $\lambda > 1$ and have the density

$$
f(\psi_1,\ldots,\psi_{d-1}) = f(c,\psi_1,\ldots,\psi_{d-1})
$$

with regard to its Shi coordinates on $B_{c_0}^{\lambda}$ as in Lemma 3.1.9. Then f has a positive mass on $\left(0, \frac{\pi}{2}\right)^{d-1}$:

$$\eta := \int_{\left(0,\frac{\pi}{2}\right)^{d-1}} f(\psi_1, \ldots, \psi_{d-1})\, d(\psi_1, \ldots, \psi_{d-1}) = \prod_{i=1}^{d-1} \left(\frac{i - \frac{1}{\lambda}}{d-i} \right) > 0.$$

Furthermore, we have, conditional on $C = ||X||_\lambda < c_0$:

(i) The Shi coordinates $C, \Psi_1, \ldots, \Psi_{d-1}$ are independent.

(ii) The random variable C is on $(0, c_0)$ uniformly distributed.

(iii) The angular component Ψ_i has the density

$$f_i(\psi_i) := (2d - 2i) \cos\psi_i \sin^{2(d-i)-1}\psi_i,$$

and, thus, the distribution function

$$F_i(\psi_i) := \int_0^{\psi_i} f_i(t)\, dt = \sin^{2(d-i)}\psi_i$$

with corresponding quantile function

$$F_i^{-1}(\psi_i) = \arcsin\left(\sqrt[2(d-i)]{\psi_i} \right),$$

for $i = 1, \ldots, d-1$.

Proof:
From Fubini's Theorem (see for example Fristedt and Gray (1997, [28]), Section 9.2) and with $\psi = (\psi_1, \ldots, \psi_{d-1})$ one gets

$$\begin{aligned}
\eta &= \int_{\left(0,\frac{\pi}{2}\right)^{d-1}} f(\psi_1, \ldots, \psi_{d-1})\, d\psi \\
&= \int_{\left(0,\frac{\pi}{2}\right)^{d-1}} \prod_{i=1}^{d-1} \left(2i - \frac{2}{\lambda}\right) \cdot \cos\psi_i \cdot \sin^{2(d-i)-1}\psi_i\, d\psi \\
&= \int_{\left(0,\frac{\pi}{2}\right)^{d-1}} \prod_{i=1}^{d-1} \left(\frac{i - \frac{1}{\lambda}}{d-i}\right) \cdot (2d-2i) \cdot \cos\psi_i \cdot \sin^{2(d-i)-1}\psi_i\, d\psi \\
&= \left(\prod_{i=1}^{d-1} \frac{i - \frac{1}{\lambda}}{d-i} \right) \left(\prod_{i=1}^{d-1} \int_0^{\frac{\pi}{2}} (2d-2i) \cdot \cos\psi_i \cdot \sin^{2(d-i)-1}\psi_i\, d\psi_i \right) \\
&= \left(\prod_{i=1}^{d-1} \frac{i - \frac{1}{\lambda}}{d-i} \right) \underbrace{\left(\prod_{i=1}^{d-1} \left[\sin^{2(d-i)}\psi_i \right]_0^{\frac{\pi}{2}} \right)}_{=1} = \prod_{i=1}^{d-1} \left(\frac{i - \frac{1}{\lambda}}{d-i} \right).
\end{aligned}$$

The condition $\eta = 0$ is equivalent to $i = \frac{1}{\lambda}$ for an $i \in \{1, \ldots, d-1\}$. Since $\lambda \geq 1$ according to Definition 2.3.3 of the logistic case this is equivalent to $\lambda = 1$. Thus

$$\eta > 0 \Longleftrightarrow \lambda > 1$$

follows.

Put $\Psi := (\Psi_1, \ldots, \Psi_{d-1})$. Consequently we have for $c \leq c_0$

$$\begin{aligned} P(C < c) &= P\left(C < c, \Psi \in \left(0, \frac{\pi}{2}\right)^{d-1}\right) \\ &= \int_0^c \int_{\left(0,\frac{\pi}{2}\right)^{d-1}} f(c, \psi_1, \ldots, \psi_{d-1})\, d\psi\, dc \\ &= \int_0^c \underbrace{\int_{\left(0,\frac{\pi}{2}\right)^{d-1}} f(\psi_1, \ldots, \psi_{d-1})\, d\psi}_{=\eta}\, dc = c \cdot \eta. \end{aligned} \tag{3.4}$$

Then for $c \in (0, c_0)$

$$P(C < c | C < c_0) = \frac{P(C < c, C < c_0)}{P(C < c_0)} = \frac{P(C < c)}{P(C < c_0)} = \frac{c \cdot \eta}{c_0 \cdot \eta} = \frac{c}{c_0} \tag{3.5}$$

holds. Thus, C is uniformly distributed on $(0, c_0)$, and we have (ii).
Let B be a Borel set from $\left(0, \frac{\pi}{2}\right)$. Notice that

$$f(c, \psi) = \prod_{i=1}^{d-1} \left(\frac{i - \frac{1}{\lambda}}{d - i}\right) f_i(\psi_i) = \eta \prod_{i=1}^{d-1} f_i(\psi_i) \tag{3.6}$$

and

$$\int_0^{\pi/2} f_i(\psi_i)\, d\psi_i = \int_0^{\pi/2} (2d - 2i) \cos\psi_i \sin^{2(d-i)-1}\psi_i\, d\psi_i = \left[\sin^{2(d-i)}\psi_i\right]_0^{\pi/2} = 1 \tag{3.7}$$

for all $1 \leq i \leq d-1$. Then we have, for all $i \in \{1, \ldots, d-1\}$, by Fubini's Theorem

$$\begin{aligned} P(\Psi_i \in B | C < c_0) &= \frac{P(\Psi_i \in B, C < c_0)}{P(C < c_0)} \\ &= \frac{P(\Psi_1 \in \left(0, \frac{\pi}{2}\right), \ldots, \Psi_{i-1} \in \left(0, \frac{\pi}{2}\right), \Psi_i \in B, \Psi_{i+1} \in \left(0, \frac{\pi}{2}\right), \ldots, \Psi_{d-1} \in \left(0, \frac{\pi}{2}\right), C < c_0)}{P(C < c_0)} \\ &\overset{(3.4)}{=} \frac{1}{c_0 \eta} \int_0^{\frac{\pi}{2}} \cdots \int_0^{\frac{\pi}{2}} \int_B \int_0^{\frac{\pi}{2}} \cdots \int_0^{\frac{\pi}{2}} \int_0^{c_0} f(c, \psi_1, \ldots, \psi_{d-1}) dc d\psi_{d-1} \ldots d\psi_{i+1} d\psi_i d\psi_{i-1} \ldots d\psi_1 \\ &\overset{(3.6),(3.7)}{=} \frac{1}{\eta c_0} c_0 \eta \int_B (2d - 2i) \cos\psi_i \sin^{2(d-i)-1}\psi_i d\psi_i \\ &= \int_B \underbrace{(2d - 2i) \cos\psi_i \sin^{2(d-i)-1}\psi_i}_{=f_i(\psi_i)} d\psi_i. \end{aligned}$$

Choosing $B = (0, \psi_i]$, the above equation reduces to

$$\begin{aligned} F_i(\psi_i) := P(\Psi_i \leq \psi_i | C < c_0) &= \int_0^{\psi_i} (2d - 2i) \cos t \cdot \sin^{2(d-i)-1} t\, dt \\ &= \left[\sin^{2(d-i)} t\right]_0^{\psi_i} = \sin^{2(d-i)}\psi_i. \end{aligned}$$

The function

$$F_i^{-1}(\psi_i) = \arcsin\left(\sqrt[2(d-i)]{\psi_i}\right)$$

is obviously the quantile function of F_i. For this reason (iii) follows.

Let $A \subset \left(0, \frac{\pi}{2}\right)^{d-1}$ be a Borel set and $0 < c < c_0$. We have, again by Fubini's Theorem,

$$\begin{aligned}
P(C < c, \Psi \in A | C < c_0) &= \frac{P(\Psi \in A, C < c)}{P(C < c_0)} \overset{(3.4)}{=} \frac{1}{c_0 \eta} \int_A \int_{(0,c)} f(t, \psi)\, dt\, d\psi \\
&= \frac{1}{\eta} \int_A f(\psi)\, d\psi \cdot \frac{c}{c_0} = \frac{1}{\eta c_0} c_0 \int_A f(\psi)\, d\psi \cdot \frac{c}{c_0} \\
&= \frac{1}{\eta c_0} \int_A \int_{(0,c_0)} f(t, \psi)\, dt\, d\psi \cdot \frac{c}{c_0} \\
&\overset{(3.4)}{=} \frac{P(\Psi \in A, C < c_0)}{P(C < c_0)} \cdot \frac{c}{c_0} \\
&\overset{(3.5)}{=} P(\Psi \in A | C < c_0) \cdot P(C < c | C < c_0),
\end{aligned}$$

thus the conditional independence of the Shi coordinates C and Ψ.

Let $A_1, \ldots, A_{d-1}$ be Borel sets in $\left(0, \frac{\pi}{2}\right)$. To prove the independence of the angular components, one has to show

$$P(\Psi_1 \in A_1, \ldots, \Psi_{d-1} \in A_{d-1} | C < c_0) = \prod_{i=1}^{d-1} P(\Psi_i \in A_i | C < c_0). \tag{3.8}$$

With an induction we show the equation

$$P(\Psi_1 \in A_1, \ldots, \Psi_j \in A_j | C < c_0) = \prod_{i=1}^{j} P(\Psi_i \in A_i | C < c_0)$$

for $j = 1, \ldots, d-1$. By putting $j = d-1$ one gets (3.8). For $j = 1$ the assertion is obvious, hence assume that the assertion is valid for j with $1 \leq j < d-1$. Then we have by Fubini's Theorem, applied several times,

$$\begin{aligned}
&P(\Psi_1 \in A_1, \ldots, \Psi_{j+1} \in A_{j+1} | C < c_0) \\
&= \frac{P(C < c_0, \Psi_1 \in A_1, \ldots, \Psi_{j+1} \in A_{j+1})}{P(C < c_0)} \\
&\overset{(3.4)}{=} \frac{1}{c_0 \eta} \int_0^{c_0} \int_{A_1 \times \ldots \times A_{j+1} \times (0,\pi/2)^{d-1-j-1}} f(c, \psi)\, d\psi dc \cdot 1 \\
&\overset{(3.6),(3.7)}{=} \frac{1}{c_0} \int_0^{c_0} \int_{A_1 \times \ldots \times A_{j+1} \times (0,\pi/2)^{d-j-2}} \prod_{i=1}^{d-1} f_i(\psi_i)\, d\psi dc \cdot \int_0^{\pi/2} f_{j+1}(\psi_{j+1})\, d\psi_{j+1} \\
&= \frac{1}{c_0} \int_0^{c_0} \int_{A_1 \times \ldots \times A_j \times (0,\pi/2)^{d-j-1}} \prod_{i=1}^{d-1} f_i(\psi_i)\, d\psi dc \cdot \int_{A_{j+1}} f_{j+1}(\psi_{j+1})\, d\psi_{j+1} \cdot 1 \\
&\overset{(3.7)}{=} \frac{1}{c_0 \eta} \int_0^{c_0} \int_{A_1 \times \ldots \times A_j \times (0,\pi/2)^{d-j-1}} \eta \prod_{i=1}^{d-1} f_i(\psi_i)\, d\psi dc \cdot \int_{A_{j+1}} f_{j+1}(\psi_{j+1})\, d\psi_{j+1} \\
&\quad \cdot \prod_{i=1, i \neq j+1}^{d-1} \int_0^{\pi/2} f_i(\psi_i)\, d\psi_i \\
&\overset{(3.6)}{=} \frac{1}{c_0 \eta} \int_0^{c_0} \int_{A_1 \times \ldots \times A_j \times (0,\pi/2)^{d-j-1}} f(c, \psi)\, d\psi dc \cdot \int_{(0,\pi/2)^j \times A_{j+1} \times (0,\pi/2)^{d-j-2}} \prod_{i=1}^{d-1} f_i(\psi_i)\, d\psi
\end{aligned}$$

$$
\begin{aligned}
&= P(\Psi_1 \in A_1, \ldots, \Psi_j \in A_j | C < c_0) \cdot \frac{1}{c_0 \eta} \int_0^{c_0} \int_{(0,\pi/2)^j \times A_{j+1} \times (0,\pi/2)^{d-j-2}} \eta \prod_{i=1}^{d-1} f_i(\psi_i)\, d\psi\, dc \\
&= \prod_{i=1}^{j} P(\Psi_i \in A_i | C < c_0) \cdot P(\Psi_{j+1} \in A_{j+1} | C < c_0) \\
&= \prod_{i=1}^{j+1} P(\Psi_i \in A_i | C < c_0).
\end{aligned}
$$

Thus the angular components are (conditionally) mutually independent and (i) follows.

□

As mentioned above, an analogy to this theorem has been shown in Theorem 3 of Shi (1995, [64]) for the extreme value case, and has been used in Algorithm 1.1 of Stephenson (2003, [70]) to simulate the corresponding random vectors. It is one of the very few results for the simulation of EVDs in general dimension. However, the distributions of the angular components are a little more complicated in the EVD case than here in the GPD case. We will apply Theorem 3.1.10 to simulate GPDs of logistic type in general dimension in the next section.

3.1.3 The Simulation Algorithm

Theorem 3.1.10 provides us with a straightforward method of simulating a random vector $(X_1, \ldots, X_d)$ following a GPD of logistic type on $B_{c_0}^{\lambda}$. The next algorithm is the only algorithm known to the author dealing with the simulation of GPDs of logistic type.

Algorithm 3.1.11

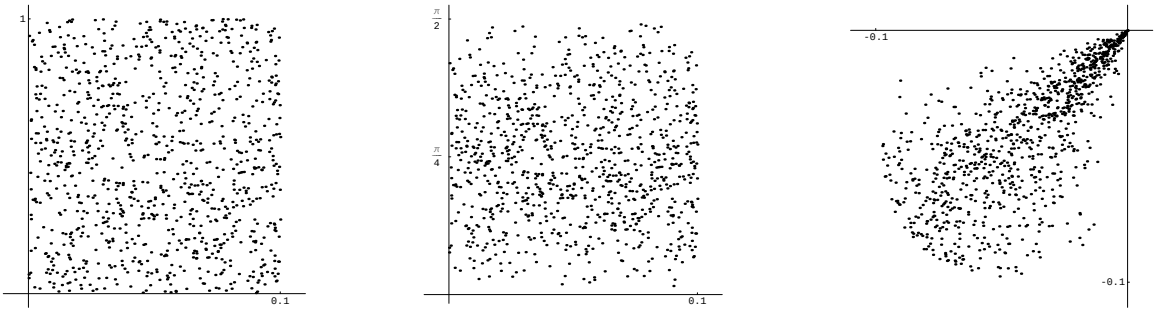
1. *Generate* U_1 *uniformly on* $(0, c_0)$ *and* $U_2, \ldots, U_d$ *uniformly on* $(0, 1)$*, all independent from each other.*
2. *Compute* $\Psi_i := F_i^{-1}(U_{i+1})$ *for* $i = 1, \ldots, d-1$.
3. *Return the vector* $(X_1, \ldots, X_d) = ST_\lambda(U_1, \Psi_1, \ldots, \Psi_{d-1})$.

◊

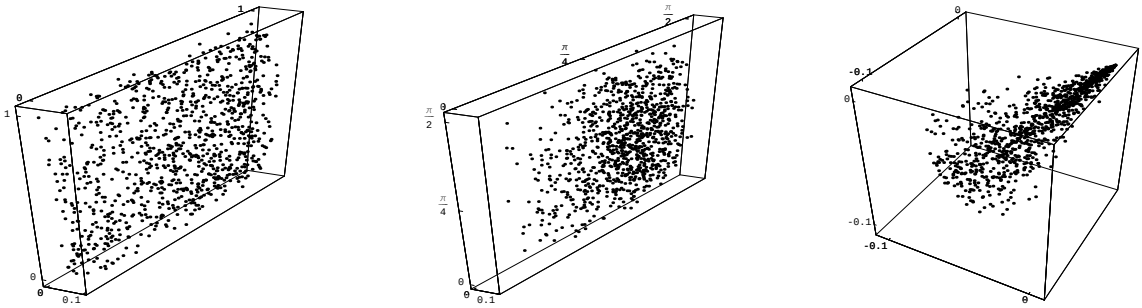
The correctness of Algorithm 3.1.11 follows immediately from Theorem 3.1.10 and the simulation of univariate random variables by the quantile transformation, see for example Corollary 1.6.4 in Falk et al. (2002, [23]).

Example 3.1.12
For illustration purposes Algorithm 3.1.11 is now used to generate 1000 vectors with the parameter values $\lambda = 4$ and $c_0 = 0.1$ for $d = 2$ and $d = 3$. First one generates the random vector (U_1, U_2) (to be seen in the first figure), then the angular component of the Shi coordinates is transformed by the quantile function F_1^{-1} (second figure), after that the Shi transformation is

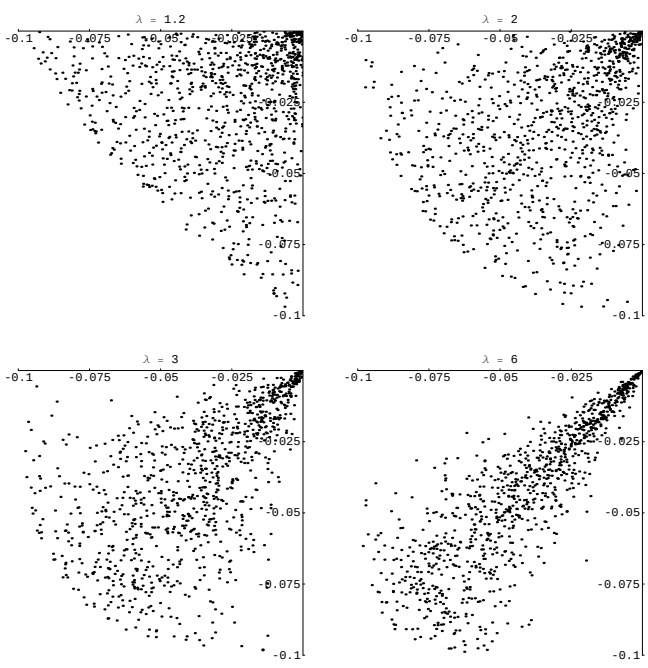
carried out (third figure).



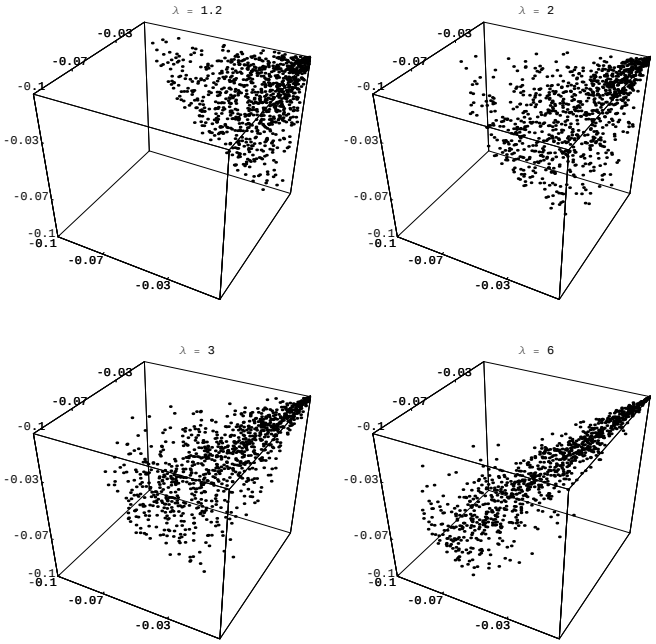
Along the same lines, one proceeds for $d = 3$, only two angular components have to be transformed this time by their quantile function (done simultaneously in the second figure).



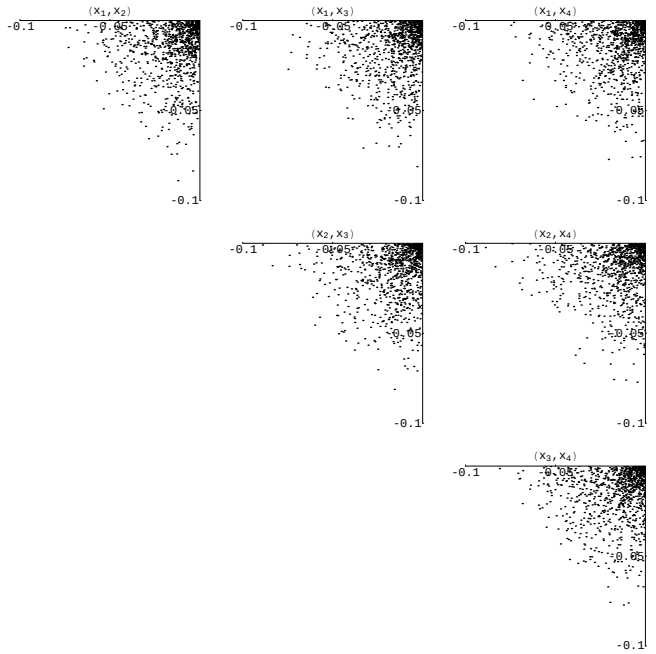
Next we look at some results for $d = 2$ and miscellaneous λ, in each plot 1000 points were generated.



The corresponding plots for $d = 3$ are presented next.



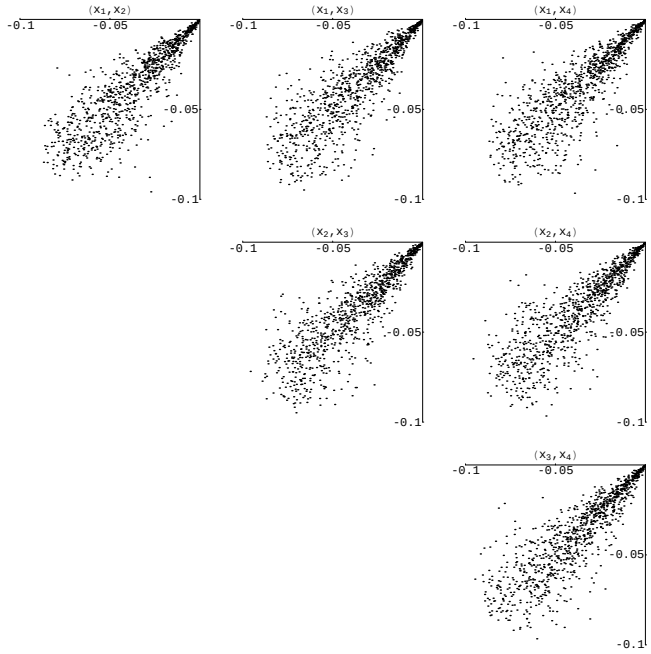
Visualizing more than three dimensions is quite difficult. A possibility are scatterplot matrices, where in each case two coordinates are plotted against each other. The next picture shows 1000 points generated by Algorithm 3.1.11 with $d = 4$ and $\lambda = 1.2$ as a scatter plot matrix.



It is easy to see by definition that the bivariate margins of GPDs of logistic type are bivariate GPDs of logistic type with the same parameter λ. This is reflected in the scatter plot matrices

above and below, where all plots show almost identical distribution patterns. The differences to the bivariate random vectors above result from the different conditions $C < c_0$ in dimensions 2 and 4.

The simulated random vectors in the next plot correspond to a parameter of $\lambda = 6$.



The generated points arrange themselves in all pictures in a sort of d–dimensional cone, whose peak lies in the origin and whose center is the line $x_i = x_j\,, i, j = 1, \ldots, d$. The lower end is naturally bounded by $||x||_\lambda = c_0$. The parameter λ describes the width of the cone. For λ close to 1 it is opened very wide, for larger λ it becomes more narrow. This is the growing dependence for $\lambda \to \infty$, see Section 2.3.2.

$\Diamond$

We will finish this section with a short runtime analysis of Algorithm 3.1.11. In the first step, d operations are done, namely the generation of d uniformly distributed random variables, so this step is $O(d)$. In the second step, a quantile function is evaluated with a constant number of operations $d-1$ times, so here we also have the runtime $O(d)$. In the final step, the Shi transformation is computed. This should be done in an efficient way. Let $(x_1, \ldots, x_d) = ST_\lambda(c, \psi_1, \ldots, \psi_{d-1})$. It is easy to see by Definition 3.1.5 that the following recursion holds

$$\begin{aligned} x_1 &= -c \cdot \cos^{2/\lambda} \psi_1, \\ x_i &= x_{i-1} \cdot \cos^{2/\lambda} \psi_i \cdot \tan^{2/\lambda} \psi_{i-1}, \qquad i = 2, \ldots, d-1, \\ x_d &= x_{d-1} \cdot \tan^{2/\lambda} \psi_{d-1}. \end{aligned}$$

With this recursion an evaluation of the Shi transformation can be done in order $O(d)$, since for each of the d components a constant number of operations has to be carried out. A straightforward implementation of the Shi transformation by Definition 3.1.5 leads to a runtime of order $O(d^2)$, which can be seen as follows. When counting the computation of $\cos^{2/\lambda}$ and $\sin^{2/\lambda}$ as one operation as well as the multiplication of those terms, we have one operation for the first

component, 3 operations for the second component, 5 operations for the third component and so on. By summing up over the d components of the computed vector we get an order of $O(d^2)$. Altogether, when using the first method, Algorithm 3.1.11 can be implemented with a runtime of order $O(d)$.

This result is backed up by the following table of runtime experiments, where 100 random vectors of dimension d with $\lambda = 4$ were generated with MATHEMATICA, version 5.2 on a Dell Notebook with an Intel Pentium M Processor with 1600 MHz.

d	2	5	10	20	50	100	200	400	800	1600
time in sec.	0.01	0.03	0.06	0.12	0.29	0.59	1.20	2.57	5.71	13.57

One can see that with a doubling of d the needed time is approximately also doubled, which means that we have a linear runtime.

We can also see that we are able to simulate very high-dimensional cases within a reasonable amount of time. This makes this algorithm interesting for practical purposes like Monte-Carlo methods in higher dimensions.

The Shi transformation used in Algorithm 3.1.11 was constructed especially for the logistic case. A question following immediately is, whether a variant of the Shi transformation exists, by which one can simulate other cases like the asymmetric or the nested logistic model. A possible generalization for the trivariate nested logistic EVD case is proposed by Shi and Zhou (1999, [65]), however, the components of that transformation are not independent like the Shi coordinates. Nevertheless, it leads to a possible simulation algorithm. To find a full generalization of the Shi transformation is, still, an open problem, which seems nontrivial, see the remarks at the end of Section 2 in Ghoudi et al. (1998, [32]). Stephenson (2003, [70]) was able to simulate the asymmetric logistic EVD case in general dimension but not by a generalization of the Shi transformation. He constructed asymmetric logistic EVD distributed random vectors from random vectors following a logistic EVD.

In the next section we will simulate GPD cases outside the logistic model, but we will use a different approach than a generalization of the Shi transformation.

3.2 Simulation of Multivariate Generalized Pareto Distributions with Bounded Pickands Density

The GPD of logistic type is the most important example of a GPD. But for practical applications it will surely be necessary to be capable of simulating random vectors outside the logistic case. That is the goal of this section. The algorithm introduced here is based on a method for the simulation of random vectors with the help of the density, the so called *rejection method*. It is described for example in Section 3.4 of Saucier (2000, [61]). Another important tool will be the transformation to Pickands coordinates, see Definition 2.1.3. The method presented here will not be as efficient as Algorithm 3.1.11 and, thus, not be applicable in high dimensions, but it has the advantage of being able to simulate a broader variety of GPDs, namely those with a bounded Pickands density. It is comparable to an algorithm simulating bivariate EVDs from Section 3 of Ghoudi et al. (1998, [32]) or Section 2 of Nadarajah (1999, [51]). But it can also go into higher dimensions.

In Section 3.2.1 we will describe the rejection method, in Section 3.2.2 we will establish a

simulation algorithm for uniform random vectors on the unit simplex which will be needed later. Finally in Section 3.2.3 we will give the simulation algorithm for GPDs with bounded Pickands density.

3.2.1 The Rejection Method for the Generation of Random Vectors

Let $F : \mathbb{R}^d \to [0,1]$ be a distribution function, which is to be simulated. We assume that it has a density $f = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F$. Furthermore, we assume that

$$A := \operatorname{supp}(f) = \overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}}$$

is a compact set in $\mathbb{R}^d$. In addition let f be bounded on A (this is always fulfilled if f is continuous on A). Then there exists $\sup_{x \in A} f(x) \in \mathbb{R}$. In the case of continuity we have a maximum instead of a supremum. Furthermore, let ξ be some real number with $\xi > 0$ and put $g(z) := \xi f(z)$. Thus $\xi \sup_{x \in A} f(x)$ is the supremum of g on A. Choose some number M with $M \geq \xi \sup_{x \in A} f(x)$. With

$$G := \left\{ (x_1, \ldots, x_{d+1}) \in \mathbb{R}^{d+1} \mid (x_1, \ldots, x_d) \in A, 0 \leq x_{d+1} \leq g(x_1, \ldots, x_d) \right\}$$

we define the set of points in $\mathbb{R}^{d+1}$, which are enclosed by the graph of g and the plane $x_{d+1} = 0$. Then we can give the following algorithm for the simulation of F.

Algorithm 3.2.1

1. *Generate a random vector $X = (X_1, \ldots, X_d)$, uniformly distributed on A.*
2. *Generate a random number Y independent of X, which is uniformly distributed on $[0, M]$.*
3. *Return X if $Y \leq g(X)$, else go to 1.*

$\Diamond$

Algorithm 3.2.1 describes the *rejection method* for the generation of random vectors. One executes this algorithm until one has the desired number of random vectors.

Lemma 3.2.2
A random vector generated by Algorithm 3.2.1 has the distribution function F.

Proof:
Let $x_1, \ldots, x_d \in \mathbb{R}$. Put $B := \{(y_1, \ldots, y_d) \in A \mid y_i \leq x_i, i = 1, \ldots, d\}$. Then we have for the distribution function $\tilde{F}$ of an accepted point

$$\begin{aligned}
\tilde{F}(x_1, \ldots, x_d) &= P((X_1, \ldots, X_d) \leq (x_1, \ldots, x_d) \mid (X_1, \ldots, X_d, Y) \in G) \\
&= \frac{P((X_1, \ldots, X_d) \leq (x_1, \ldots, x_d), (X_1, \ldots, X_d, Y) \in G)}{P((X_1, \ldots, X_d, Y) \in G)} \\
&= \frac{\frac{\int_B g(z)\, dz}{\int_{A \times [0,M]} 1\, d\zeta}}{\frac{\int_A g(z)\, dz}{\int_{A \times [0,M]} 1\, d\zeta}} = \frac{\int_B \xi f(z)\, dz}{\int_A \xi f(z)\, dz} = \frac{\int_B f(z)\, dz}{\int_A f(z)\, dz} = \int_B f(z)\, dz = F(x_1, \ldots, x_d),
\end{aligned}$$

since due to the uniform distribution of the generated points the probabilities can be expressed by corresponding volumes. Also note that $\int_A f(z)\, dz = 1$, because f is a density on A.

$\square$

A geometric interpretation of this method can be found in Section 3.4 of Saucier (2000, [61]).

Remark 3.2.3
If $\xi = 1$ in Algorithm 3.2.1 the rejection method is executed with the exact density. However, Lemma 3.2.2 also shows that the rejection method is applicable if only a constant multiple g of the density f is known with ξ possibly unknown. This is something which we will use later in Algorithm 3.2.11.

$\Diamond$

It is immediately clear that Algorithm 3.2.1 can be very inefficient, since, owing to circumstances, a lot of points might have to be generated to get one "useable" point. A measure for the efficiency of Algorithm 3.2.1 is the probability that the point (X, Y), which is uniformly distributed on $A \times [0, M]$, lies in G and is, thus, accepted.

Definition 3.2.4
The probability $\omega := P((X, Y) \in G)$ is called the *efficiency* of Algorithm 3.2.1.

$\Diamond$

Theorem 3.2.5
We have

$$\omega = \frac{\xi}{M \cdot \mathrm{vol}(A)} \leq \frac{1}{\sup_{x \in A} f(x) \cdot \mathrm{vol}(A)}$$

with equality holding if $M = \xi \sup_{x \in A} f(x)$ is chosen. Thereby $\mathrm{vol}(A) := \int_A 1 \, dz$ denotes the volume of A.

Proof:
The probability ω is the ratio of the volume under the graph of g to the volume of the set $A \times [0, M]$. Therefore, we have

$$\omega = P((X, Y) \in G) = \frac{\int_A g(z) \, dz}{\int_{A \times [0,M]} 1 \, d\zeta} = \frac{\xi}{M \cdot \int_A 1 \, dz}.$$

Thus by $M \geq \xi \sup_{x \in A} f(x)$ we have

$$\omega \leq \frac{1}{\sup_{x \in A} f(x) \cdot \mathrm{vol}(A)}.$$

By choosing $M = \xi \sup_{x \in A} f(x)$ we get

$$\omega = \frac{1}{\sup_{x \in A} f(x) \cdot \mathrm{vol}(A)}.$$

$\square$

If one wants to have n random vectors following F one has to generate

$$\frac{n}{\omega} = n \cdot \frac{M}{\xi} \cdot \mathrm{vol}(A) \geq n \cdot \sup_{x \in A} f(x) \cdot \mathrm{vol}(A)$$

points on the average. Here we see why M should be chosen close to $\xi \sup_{x \in A} f(x)$ and, thus, as small as possible. Every increase in M leads to a decrease of the efficiency and, thus, to a raise in the number of generated points.

Besides the possible inefficiency there are further problems with this method.

(i) How to generate the uniform distribution on A ?

(ii) How to compute M, the supremum of g or a proper, i.e., not to rough estimation?

These are questions, which can best be solved for a given problem. Possibilities that always work, but which are not necessarily efficient, are for example

(i) The uniform distribution on a d–dimensional cube K is easy to simulate. Thus find a cube K with $A \subseteq K$, generate a random vector, uniformly distributed on it, and accept it only if it also falls into A. This is again the rejection method.

(ii) Since the computation of the gradient and the following solution of a system of equations can often not be done exactly, one can use numerical optimization procedures such as the algorithm by Nelder-Mead (see for example Section 8.1 in Kelley (1999, [46])) to get an approximation of the maximum.

Remark 3.2.6
If $\text{supp}(f)$ is not a compact set, i.e., if it is not bounded, but if f is bounded, one defines with $\varepsilon > 0$ the function

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } f(x) > \varepsilon, \\ 0 & \text{else.} \end{cases}$$

If ε is small enough, $\tilde{f}$ is a good approximation for f and has a compact support. Thus one can carry out Algorithm 3.2.1 with $\tilde{f}$ and get random vectors which are approximately distributed by F. Thereby we have of course $\int \tilde{f}\, dx < 1$. ε should, thus, be chosen such that the integral is close to 1.

If the density f itself is unbounded, one could choose a large number M and define

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } f(x) < M, \\ M & \text{else,} \end{cases}$$

such that the integral is still close to 1. The rejection method is then applicable, and f is approximated by a density $\tilde{f}$ which corresponds to the uniform distribution in the area, where f is unbounded. Depending on f, this might be a crude approximation of the original distribution, especially if one is interested in the area, where the density is unbounded.

There exist also other variants of the rejection method, where a density g is needed, which can easily be simulated and which fulfills $\frac{f(x)}{g(x)} \leq M$ for all $x \in A$. The random vector X, which is originally distributed by g, is accepted if $Y \leq \frac{f(X)}{g(X)}$ for a uniform random number $Y \in [0, M]$, see Section 3.4 of Saucier (1999, [61]) for details.

$\Diamond$

3.2.2 Generation of Uniform Random Vectors on the Unit Simplex

For the simulation of generalized Pareto distributions with the help of the rejection method it will be important to be able to generate uniformly distributed random variables on the unit simplex R_d. We present a method in this section to complete this task as efficient as possible.

Theorem 3.2.7
Let $c > 0$ and $(U_1, \ldots, U_d)$ be uniformly distributed on $R_{c,d}$ with $R_{c,d}$ defined as in Lemma 2.1.2. Then we have that $(U_1, \ldots, U_{d-1})$ is uniformly distributed on $R_{c-u_d,d-1}$, conditional on $U_d = u_d \in (0, c)$, i.e.,

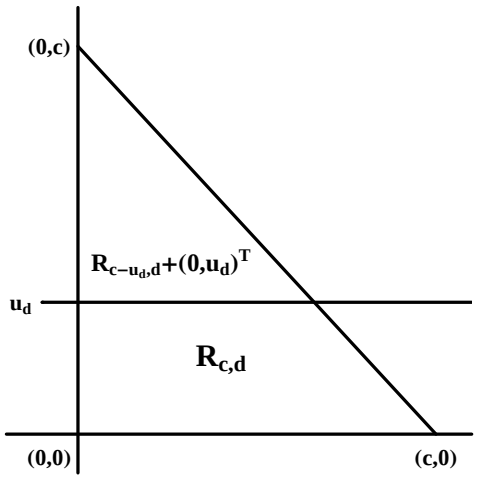
$$P(U_1 \leq u_1, \ldots, U_{d-1} \leq u_{d-1} \mid U_d = u_d) = \frac{u_1 \cdot \ldots \cdot u_{d-1}}{\mathrm{vol}\,(R_{c-u_d,d-1})},$$

where $(u_1, \ldots, u_{d-1}) \in R_{c-u_d,d-1}$ and $u_d \in (0, c)$.

Proof:
Note that U_d is not uniformly distributed on $(0, c)$, since we have by Lemma 2.1.2

$$\begin{aligned} K(u_d) &:= P(U_d \leq u_d) = 1 - P(U_d > u_d) = 1 - P(U_d > u_d, U_1 \leq c, \ldots, U_{d-1} \leq c) \\ &= 1 - \frac{\mathrm{vol}\left(R_{c,d} \cap \left\{x \in \mathbb{R}^d | x_d > u_d\right\}\right)}{\mathrm{vol}(R_{c,d})} \\ &= 1 - \frac{\mathrm{vol}\left(R_{c-u_d,d} + (0, \ldots, 0, u_d)^T\right)}{\mathrm{vol}(R_{c,d})} \\ &= 1 - \frac{\mathrm{vol}\,(R_{c-u_d,d})}{\mathrm{vol}(R_{c,d})} = 1 - \frac{(c-u_d)^d}{c^d}, \end{aligned} \tag{3.9}$$

because the intersection of $R_{c,d}$ with $\left\{x \in \mathbb{R}^d | x_d > u_d\right\}$ is a simplex of edge length $c - u_d$ shifted by the vector $(0, \ldots, 0, u_d)^T$. This is illustrated by the following figure for the case $d = 2$:



From (3.9), Lemma 2.1.2 and Example 5.2.24 from Gänssler and Stute (1977, [30]) we conclude for $(u_1, \ldots, u_{d-1}) \in R_{c-u_d,d-1}$:

$$\begin{aligned} &P(U_1 \leq u_1, \ldots, U_{d-1} \leq u_{d-1} \mid U_d = u_d) \\ &= \lim_{\varepsilon \to 0} P(U_1 \leq u_1, \ldots, U_{d-1} \leq u_{d-1} \mid U_d \in [u_d, u_d + \varepsilon]) \\ &= \lim_{\varepsilon \to 0} \frac{P(U_1 \leq u_1, \ldots, U_{d-1} \leq u_{d-1}, U_d \in [u_d, u_d + \varepsilon])}{P(U_d \in [u_d, u_d + \varepsilon])} \\ &= \lim_{\varepsilon \to 0} \frac{\varepsilon \cdot u_1 \cdot \ldots \cdot u_{d-1} / \mathrm{vol}(R_{c,d})}{\varepsilon} \cdot \frac{\varepsilon}{P(U_d \in [u_d, u_d + \varepsilon])} \\ &= \frac{u_1 \cdot \ldots \cdot u_{d-1}}{\mathrm{vol}(R_{c,d})} \cdot \lim_{\varepsilon \to 0} \frac{\varepsilon}{K(u_d + \varepsilon) - K(u_d)} \end{aligned}$$

$$\begin{aligned} &= \frac{u_1 \cdot \ldots \cdot u_{d-1}}{\mathrm{vol}(R_{c,d})} \cdot \frac{1}{K'(u_d)} \\ &= \frac{u_1 \cdot \ldots \cdot u_{d-1}}{\mathrm{vol}(R_{c,d})} \cdot \frac{1}{\frac{\partial}{\partial u_d}\left(1 - \frac{(c-u_d)^d}{c^d}\right)} \\ &= \frac{d\,!c^d}{c^d d(c-u_d)^{d-1}} \cdot u_1 \cdot \ldots \cdot u_{d-1} = \frac{u_1 \cdot \ldots \cdot u_{d-1}}{\mathrm{vol}\left(R_{c-u_d,d-1}\right)}, \end{aligned}$$

as asserted.

□

Corollary 3.2.8
Let $(U_1, \ldots, U_d)$ be uniformly distributed on R_d, and suppose that the realizations $u_{i+1}, \ldots, u_d, i \in \{1, \ldots, d\}$ of $U_{i+1}, \ldots, U_d$ are observed. Then U_i has the conditional distribution function

$$H_i(u_i) := P\left(U_i \leq u_i | U_{i+1} = u_{i+1}, \ldots, U_d = u_d\right) = 1 - \frac{(1 - u_d - \ldots - u_{i+1} - u_i)^i}{(1 - u_d - \ldots - u_{i+1})^i}$$

and the corresponding conditional quantile function

$$H_i^{-1}(x) = (1 - u_d - \ldots - u_{i+1})(1 - \sqrt[i]{1-x}).$$

Proof:
We prove the first assertion by induction, more precisely we show:

Given $u_{i+1}, \ldots, u_d$ we have that $(U_1, \ldots, U_i)$ is uniformly distributed on $R_{1-u_d-\ldots-u_{i+1},i}$ and U_i has the distribution function

$$H(u_i) = 1 - \frac{(1 - u_d - \ldots - u_{i+1} - u_i)^i}{(1 - u_d - \ldots - u_{i+1})^i}.$$

In the case $i = d$, the random vector $(U_1, \ldots, U_d)$ is uniformly distributed on R_d by assumption, and by (3.9) U_d has the distribution function

$$H(u_d) = P(U_d \leq u_d) = 1 - \frac{(1 - u_d)^d}{1^d},$$

which shows the assertion in this case.

Let the assertion hold for $i \in \{2, \ldots, d\}$. In the induction step we show the assertion for $i - 1$.

Let $(u_i, \ldots, u_d)$ be given. $(U_1, \ldots, U_i)$ is by the induction hypothesis uniformly distributed on $R_{1-u_d-\ldots-u_{i+1},i}$. Since now, in addition, u_i is given, $(U_1, \ldots, U_{i-1})$ is by Theorem 3.2.7 uniformly distributed on $R_{1-u_d-\ldots-u_{i+1}-u_i,i-1}$. By (3.9) we know that U_{i-1} has then the distribution function

$$H(u_{i-1}) = P(U_{i-1} \leq u_{i-1}) = 1 - \frac{(1 - u_d \ldots - u_{i+1} - u_i - u_{i-1})^{i-1}}{(1 - u_d - \ldots - u_{i+1} - u_i)^{i-1}}.$$

This is the assertion for $i - 1$.

Now we compute the quantile function of $H_i(u_i)$. We have

$$\begin{aligned} x \quad = \quad & H_i(u_i) = 1 - \frac{(1 - u_d - \ldots - u_{i+1} - u_i)^i}{(1 - u_d - \ldots - u_{i+1})^i} \iff \frac{(1 - u_d - \ldots - u_{i+1} - u_i)^i}{(1 - u_d - \ldots - u_{i+1})^i} = 1 - x \\ \iff \quad & 1 - u_d - \ldots - u_{i+1} - u_i = (1 - u_d - \ldots - u_{i+1})\sqrt[i]{1-x} \\ \iff \quad & u_i = (1 - u_d - \ldots - u_{i+1}) - (1 - u_d - \ldots - u_{i+1})\sqrt[i]{1-x} \\ \iff \quad & u_i = (1 - u_d - \ldots - u_{i+1})(1 - \sqrt[i]{1-x}) = H_i^{-1}(x). \end{aligned}$$

□

The known quantile transformation method, see for example Corollary 1.6.4 from Falk et al. (2002, [23]), together with Corollary 3.2.8 provides us with the following recursive algorithm for the simulation of a uniformly distributed random vector on the unit simplex R_d.

Algorithm 3.2.9
Set k:=0 and for $i = d, \ldots, 1$ *do*

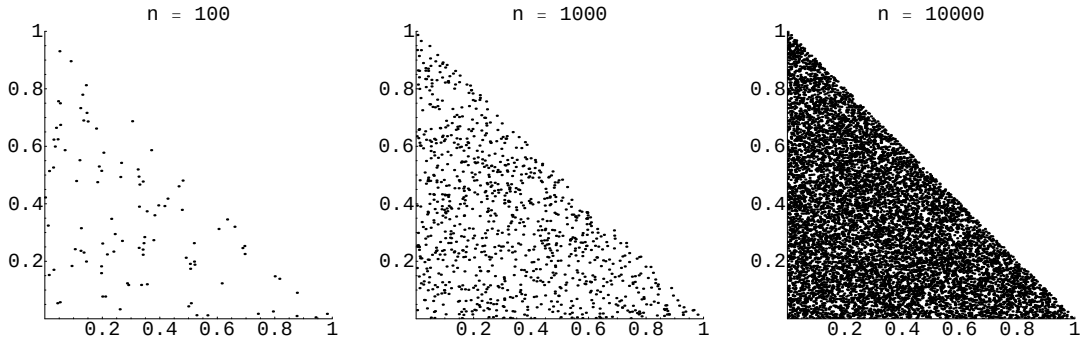
1. *Generate a random number* x_i *on* $(0,1)$ *uniformly distributed and independent of* x_j, $i{+}1 \leq j \leq d$.
2. *Compute* $u_i := (1-k)(1-\sqrt[i]{1-x_i})$.
3. *Put* $k := k + u_i$.

Return the vector $(u_1, \ldots, u_d)$.

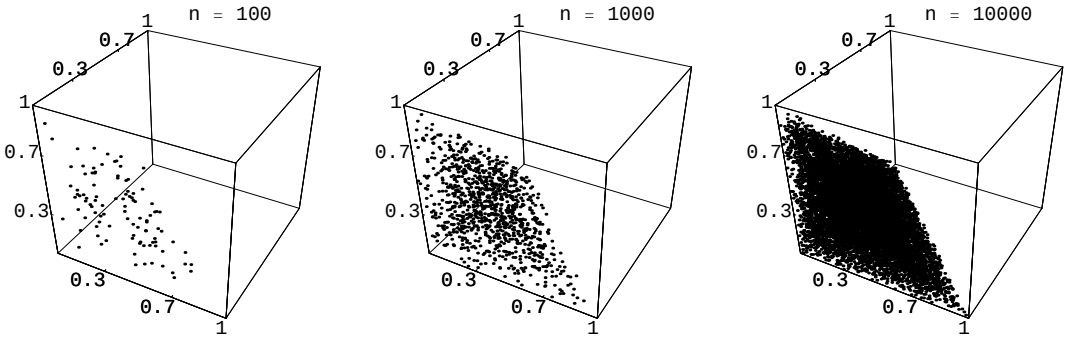
$\Diamond$

The runtime of Algorithm 3.2.9 is of order $O(d)$, since in each of the d steps a constant number of operations is performed. Anything better cannot be expected since at least the d components of the vector must be generated. If one wants to have n uniformly distributed random vectors, one has to call Algorithm 3.2.9 n times independently. Then the runtime is of order $O(nd)$, which is also optimal.

Example 3.2.10
In this example we want to take a look at a practical implementation of Algorithm 3.2.9 for the simulation of the uniform distribution on R_d. With Algorithm 3.2.9 100, 1000 and 10000 random vectors were generated for $d = 2$ and plotted in the following three graphics.



Exactly the same was done for the case $d = 3$.



In the following table we note the time in seconds which was needed on an Intel Pentium 4 Processor to generate 100 points with MATHEMATICA, version 5.2, in various dimensions d.

d	2	10	100	1000	10000
time in sec.	0.01	0.05	0.42	4.14	41.17

We see that a multiplication of the dimension by a factor of 10 leads to an increase in the runtime by a factor of approximately 10.

$\Diamond$

3.2.3 The Simulation Algorithm for Multivariate Generalized Pareto Distributions with Bounded Pickands Density

A natural scheme for the simulation of a generalized Pareto distribution W is the application of the rejection method to W or more precisely to its density $w = \frac{\partial^d}{\partial x_d \dots \partial x_1} W$. However, as can be seen for the density of the logistic case (Lemma 2.3.6), the relation

$$\begin{aligned}
\lim_{t\to 0-} w_\lambda(t,\dots,t) &= \lim_{t\to 0-} \left(\prod_{i=1}^{d-1}(i\lambda-1)\right)\left((-t)^{\lambda-1}\right)^d \left(d(-t)^\lambda\right)^{-d+\frac{1}{\lambda}} \\
&= \left(\prod_{i=1}^{d-1}(i\lambda-1)\right) d^{-d+\frac{1}{\lambda}} \lim_{t\to 0-} (-t)^{\lambda d-d-d\lambda+1} \\
&= \left(\prod_{i=1}^{d-1}(i\lambda-1)\right) d^{-d+\frac{1}{\lambda}} \lim_{t\to 0-} \left(-\frac{1}{t}\right)^{d-1} = \infty
\end{aligned}$$

holds. Thus even in the most common case the density w is not bounded. Therefore, the rejection method is not directly applicable, since we are especially interested in the area close to 0, see Remark 3.2.6. In some cases, however, a detour via Pickands coordinates can be very helpful.

The idea is to generate the Pickands coordinates Z and C of a random vector, which follows W, separately. These are by Theorem 2.2.5 independent under $C > c_0$ and, in addition, C is uniformly distributed on $(c_0, 0)$ and, thus, easy to simulate.

For the generation of Z we can apply the rejection method if its density f and, thus, the Pickands density ϕ are bounded. This is for example the case with the logistic distribution, see Theorem 2.3.7. Since Z lies in the unit simplex R_{d-1}, the uniform distribution there can be simulated by Algorithm 3.2.9. One uses the Pickands density ϕ, since for the use of f the number μ, see Theorem 2.2.5, would have to be calculated, which is possible only approximately and only with a great numerical effort. In the end one has to invert the Pickands transformation to get the desired random vector.

The following algorithm implementing these considerations is another main result of this manuscript.

Algorithm 3.2.11

1. *Generate the realization $(z_1, \dots, z_{d-1})$ of a random vector, which has density f, with Algorithm 3.2.1 applied to the Pickands density ϕ. Use the parameters $A = \overline{R_{d-1}}$, $M =$*

$\sup_{z \in R_{d-1}} \phi(z)$ *for the generation. For the uniform distribution on the unit simplex* $\overline{R_{d-1}}$ *use Algorithm 3.2.9, M can, for example, be computed with the algorithm by Nelder-Mead.*

2. *Generate, independent of* $(z_1, \ldots, z_{d-1})$*, a random number c, uniformly distributed on* $(c_0, 0)$*.*

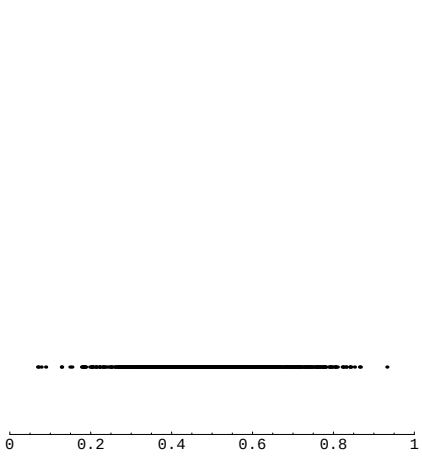
3. *Return the vector* $\left(cz_1, \ldots, cz_{d-1}, c - c\sum_{i=1}^{d-1} z_i\right)$*.*

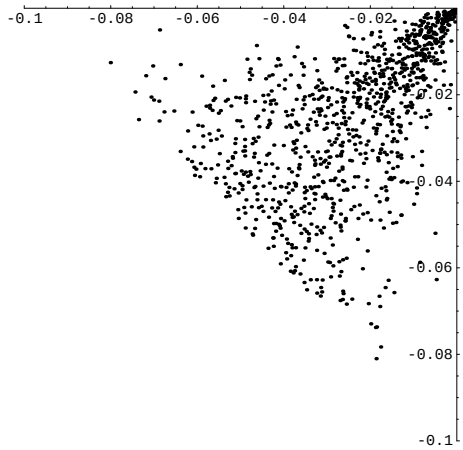
$\Diamond$

To get n random vectors one has to call Algorithm 3.2.11 n times. The efficiency of the algorithm is that of Algorithm 3.2.1, since all other steps are deterministic.

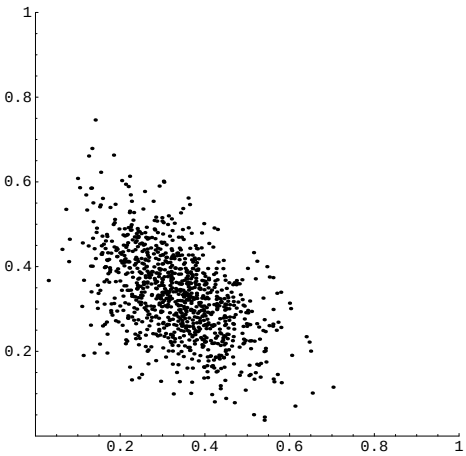
Example 3.2.12
We want to examine Algorithm 3.2.11 a little closer for the logistic case. We used again MATHEMATICA, version 5.2 and took M from Theorem 2.3.7. We begin with the parameters $d = 2$, $\lambda = 3$.

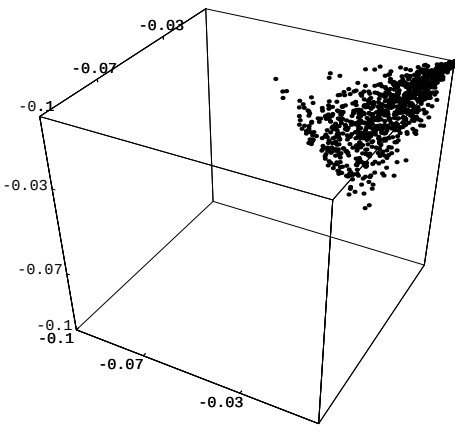
In the left graphic one can see 1000 random numbers, distributed according to f_3 and generated with the Pickands density ϕ_3, the angular components. In the right picture one can see the desired logistic random vectors. These were generated by the addition of the radial component with $c = -0.1$ to the angular components and the inversion of the Pickands transformation.





Next we show the same plots for $d = 3$.





For more graphics with vectors coming from the logistic distribution, see Example 3.1.12.

$\Diamond$

In the sequel we will do some efficiency considerations.

Theorem 3.2.13
The efficiency of Algorithm 3.2.11 in the logistic case is

$$\omega_\lambda = \frac{\mu_\lambda (d-1)!}{d^{\frac{1}{\lambda}-1} \prod_{i=1}^{d-1} (i\lambda - 1)} = \frac{\mu_\lambda (d-1)! \Gamma(1-\frac{1}{\lambda})}{d^{\frac{1}{\lambda}-1} \lambda^{d-1} \Gamma(d-\frac{1}{\lambda})} \tag{3.10}$$

with $\omega_\lambda = O\left(\frac{1}{\lambda^{d-1}}\right)$ for $\lambda \to \infty$ and $\lim_{\lambda \to 1} \omega_\lambda = 1$.

Proof:
We have by Lemma 2.1.2, Theorems 2.3.7 and 3.2.5

$$\omega_\lambda = \frac{1}{\frac{m_\lambda}{\mu_\lambda} \cdot \frac{1}{(d-1)!}} = \frac{\mu_\lambda (d-1)!}{d^{\frac{1}{\lambda}-1} \prod_{i=1}^{d-1} (i\lambda - 1)} = \frac{\mu_\lambda (d-1)! \Gamma(1-\frac{1}{\lambda})}{d^{\frac{1}{\lambda}-1} \lambda^{d-1} \Gamma(d-\frac{1}{\lambda})},$$

where the final equality is due to Remark 2.3.9. We have

$$\begin{aligned} \lim_{\lambda \to \infty} \Gamma\left(1 - \frac{1}{\lambda}\right) &= \Gamma(1) = 1, \\ \lim_{\lambda \to \infty} \Gamma\left(d - \frac{1}{\lambda}\right) &= \Gamma(d) = (d-1)!, \\ \lim_{\lambda \to \infty} d^{\frac{1}{\lambda}-1} &= d^{-1}, \\ \lim_{\lambda \to \infty} \mu_\lambda &= \frac{1}{d}, \end{aligned} \tag{3.11}$$

where (3.11) will be shown in Theorem 5.7.2. Putting the above results into (3.10) we get

$$\omega_\lambda = O\left(\frac{1}{\lambda^{d-1}}\right) \quad \text{for } \lambda \to \infty.$$

We will use the representation (3.10) and the representation of the logistic Pickands density from Theorem 2.3.7 for the second limit:

$$\begin{aligned} \lim_{\lambda \to 1} \omega_\lambda &= \lim_{\lambda \to 1} \frac{\mu_\lambda (d-1)!}{d^{\frac{1}{\lambda}-1} \prod_{i=1}^{d-1} (i\lambda - 1)} \\ &= (d-1)! \lim_{\lambda \to 1} \frac{\int_{R_{d-1}} \phi_\lambda(z)\, dz}{\prod_{i=1}^{d-1} (i\lambda - 1)} = (d-1)! \lim_{\lambda \to 1} \int_{R_{d-1}} \frac{\phi_\lambda(z)}{\prod_{i=1}^{d-1} (i\lambda - 1)}\, dz \\ &= (d-1)! \lim_{\lambda \to 1} \int_{R_{d-1}} \left(\prod_{i=1}^{d-1} z_i\right)^{\lambda-1} \left(1 - \sum_{i=1}^{d-1} z_i\right)^{\lambda-1} \left(\sum_{i=1}^{d-1} z_i^\lambda + \left(1 - \sum_{i=1}^{d-1} z_i\right)^\lambda\right)^{\frac{1}{\lambda}-d} dz. \end{aligned}$$

The function $\frac{\phi_\lambda(z)}{\prod_{i=1}^{d-1}(i\lambda-1)}$ converges pointwise to the constant 1 for $\lambda \to 1$. Since by Theorem 2.3.7 the function $\frac{\phi_\lambda(z)}{\prod_{i=1}^{d-1}(i\lambda-1)}$ is bounded by $d^{\frac{1}{\lambda}-1} \leq 1$, the dominated convergence theorem (see

for example Fristedt and Gray (1997, [28]), Section 8.2) is applicable and, thus, we get with Lemma 2.1.2

$$\begin{aligned}\lim_{\lambda\to 1}\omega_\lambda &= (d-1)!\int_{R_{d-1}}\lim_{\lambda\to 1}\left(\prod_{i=1}^{d-1}z_i\right)^{\lambda-1}\left(1-\sum_{i=1}^{d-1}z_i\right)^{\lambda-1}\left(\sum_{i=1}^{d-1}z_i^{\lambda}+\left(1-\sum_{i=1}^{d-1}z_i\right)^{\lambda}\right)^{\frac{1}{\lambda}-d}dz\\ &= (d-1)!\int_{R_{d-1}}1\,dz=\frac{(d-1)!}{(d-1)!}=1,\end{aligned}$$

which completes the proof.

□

On the average one has to generate $\frac{n}{\omega_\lambda}$ points to get n vectors distributed by f_λ. Since all deterministic operations in Algorithm 3.2.11 are of order $O(d)$ and $O(nd)$ respectively, the total runtime of Algorithm 3.2.11 is thus

$$O\Big(\underbrace{\frac{n}{\omega_\lambda}}_{\substack{\text{number of uniformly}\\ \text{distributed points}}}\cdot\underbrace{d}_{\substack{\text{runtime of generation}\\ \text{of uniform distribution}}}+\underbrace{nd}_{\substack{\text{remaining}\\ \text{operations}}}\Big)$$

on average.

Thus by Theorem 3.2.13 for a large λ this runtime is of order $O\left(nd\lambda^{d-1}\right)$, for small λ approximatively of order $O(nd)$.

Observe that the runtime is independent of the threshold c_0. The linear factor n is to be expected, since we want to generate n points. The linear factor d is also not surprising, since a higher-dimensional point also possesses more components, which have to be generated. What makes the algorithm very slow for large d and large λ, is the factor λ^{d-1}. However, this is also not unusual, since in many d–dimensional problems the dimension shows up as an exponent in the runtime. This phenomenon is also known as the "Curse of Dimension".

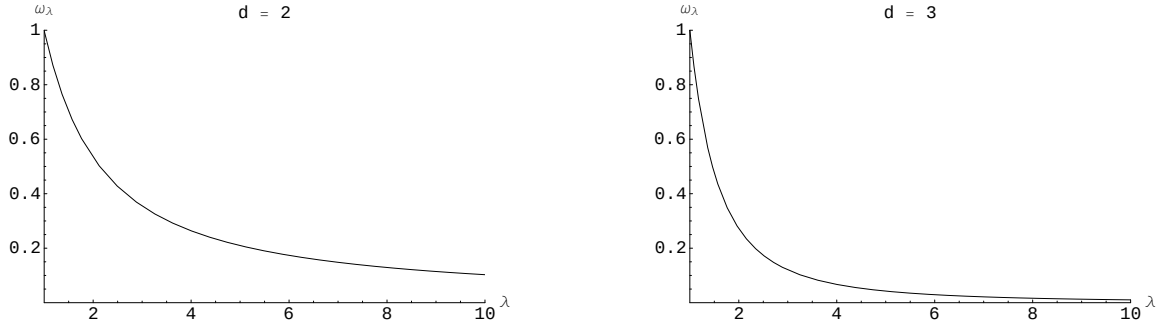
To get an empirical measure for the efficiency of Algorithm 3.2.11, one can define the *empirical efficiency* $\hat{\omega}$ by

$$\hat{\omega} := \frac{n}{\text{total number of points generated}}. \tag{3.12}$$

This converges for $n\to\infty$ by the law of large numbers (see for example Serfling (1980, [63]), Theorem 1.8B) to the theoretical efficiency ω, see Definition 3.2.4.

Example 3.2.14
First we want the show some numerical evaluations (done according to (3.10) with `NIntegrate`

from MATHEMATICA) of the efficiency ω_λ in two and three dimensions.



One sees the low efficiency for large λ and the large efficiency for small λ. As to be expected the values for $d = 2$ are much better. Only for λ very close to 1, the efficiency ω_λ is also close to 1, which means that the second convergence from Theorem 3.2.13 is very slow. This behavior strengthens when going to dimension 3 and higher.

Finally we created a table with the theoretical efficiency ω_λ as in Theorem 3.2.13, the empirical efficiency $\hat{\omega}$ according to (3.12) and the runtime of an implementation of Algorithm 3.2.11 for the logistic case in MATHEMATICA, version 5.2, for various d and λ. Each time, 100 GPD distributed points had to be generated on an Intel Pentium 4 Processor.

ω $\hat{\omega}$ time in sec.	$\lambda = 1.2$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 8$	$\lambda = 16$
$d = 2$	0.8577 0.884956 0.032	0.53284 0.487805 0.031	0.354224 0.353358 0.063	0.263868 0.263852 0.078	0.12918 0.118196 0.125	0.0636262 0.0680736 0.093
$d = 3$	0.722672 0.78125 0.044	0.271875 0.247525 0.093	0.120392 0.110011 0.188	0.0670297 0.0803983 0.297	0.016305 0.0159464 1.281	0.00399798 0.00372633 5.48
$d = 4$	0.605439 0.636943 0.057	0.137269 0.135136 0.188	0.0404847 0.0391547 0.625	0.0168861 0.0198414 1.234	0.00204753 0.00247012 9.813	0.000250514 0.000248682 97.421
$d = 5$	0.506153 0.469484 0.084	0.0690263 0.0644333 0.454	0.0135639 0.0131222 2.14	0.00423957 0.00466592 6.031	0.000256629 0.000260431 107.454	0.0000156806 0.0000140977 1993.94

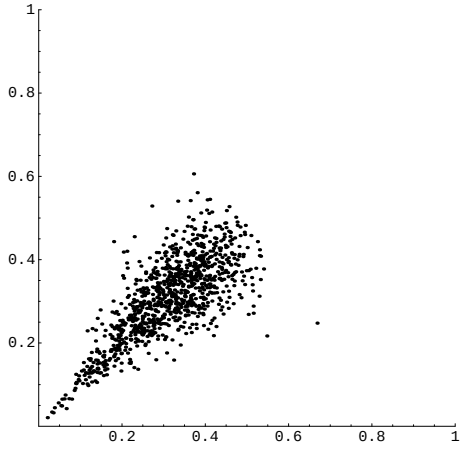
Here we see the exploding runtimes for large λ and d, as described above. The runtimes itself can be shortened by using better suited numerical programming packages like MATLAB or C++, but the overall asymptotic behavior remains unaffected.

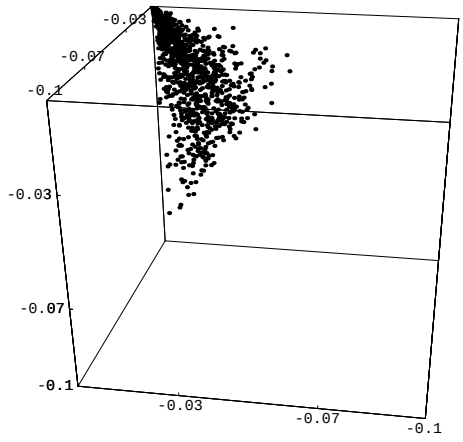
$\Diamond$

From the above runtime considerations it is clear that Algorithm 3.1.11 is superior to the algorithm considered here, with the exception of values of λ very close to 1. Another advantage of Algorithm 3.1.11 is that it is deterministic. But Algorithm 3.2.11 has the advantage that it can be used for cases other than the logistic one. We will use this property in the following two examples.

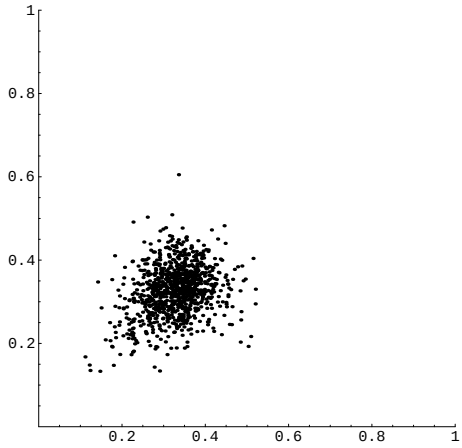
Example 3.2.15
In the following plots we want to give a visual impression of data simulating the nested logistic distribution from Definition 2.3.13. The Pickands density needed for this is given in Remark 2.3.14.

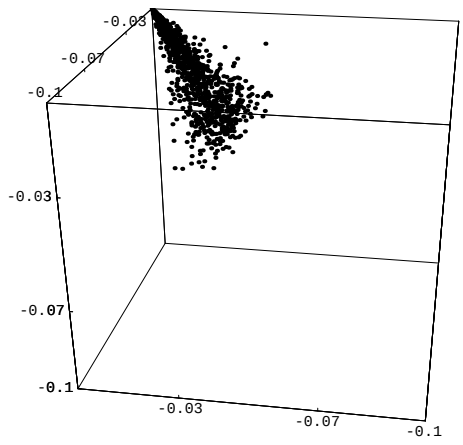
As before, the left picture shows the corresponding angular components of the Pickands coordinates, the right picture the actual observations after the inverse Pickands transformation. One has to be careful with the nested logistic model, since the Pickands density is not always bounded there, see the pictures in Example 2.3.15. For the examples shown here we have checked that the Pickands density is bounded. The following plot shows 1000 points originating from the parameter values $\lambda_1 = 8, \lambda_2 = 2$ with $c = -0.1$.



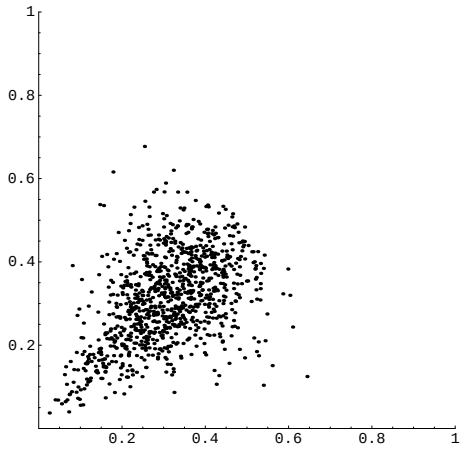


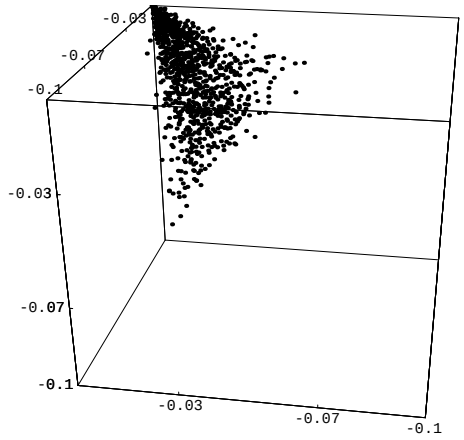
Furthermore, we have $\lambda_1 = 8, \lambda_2 = 4$.



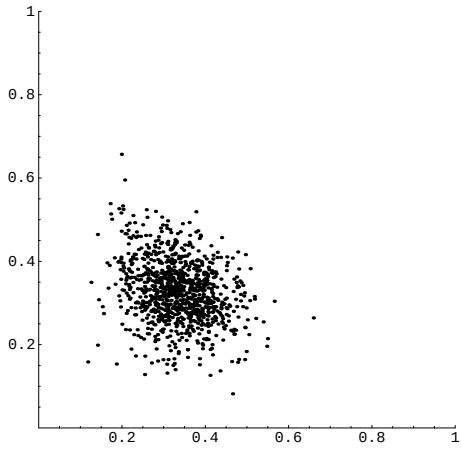


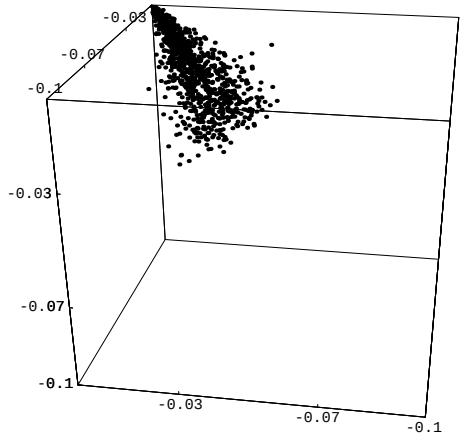
Next $\lambda_1 = 5, \lambda_2 = 2$.





And finally $\lambda_1 = 5, \lambda_2 = 4$.





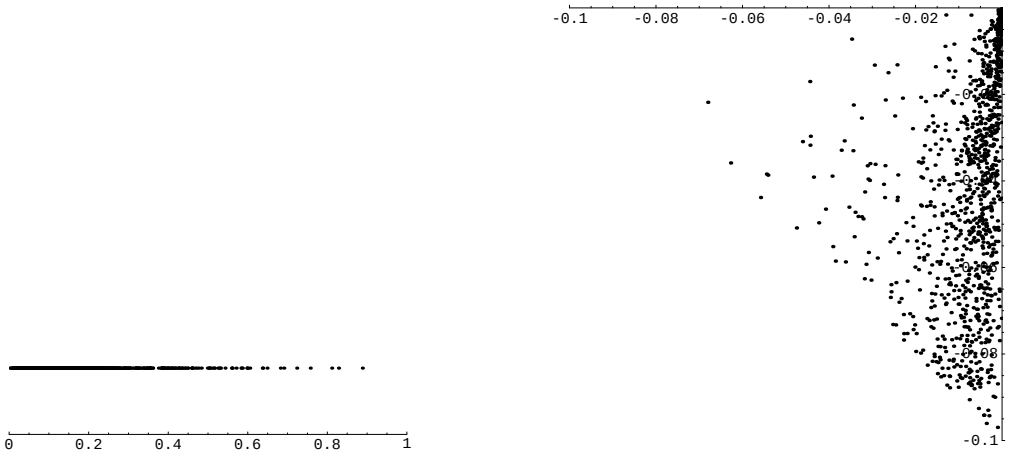
In contrast to the logistic model the observations no longer form a cone which is symmetric to the line $x_i = x_j$, $i, j = 1, \ldots, d$, see the remarks in Example 3.1.12. In the case $d = 3$, the cone is distorted in the x_3-direction, and the heaviness of the distortion is determined by the difference of the parameters. The closer the parameters are to each other, the closer one is to the logistic model, and the more symmetric the cone becomes with regard to the line $x_i = x_j$, $i, j = 1, \ldots, d$. In terms of the angular components in $d = 3$, the parameter λ_1 governs the opening angle of the angular components and λ_2 determines how closely the angular components come to the origin.

Due to the complicated nature of the model we skip a runtime analysis.

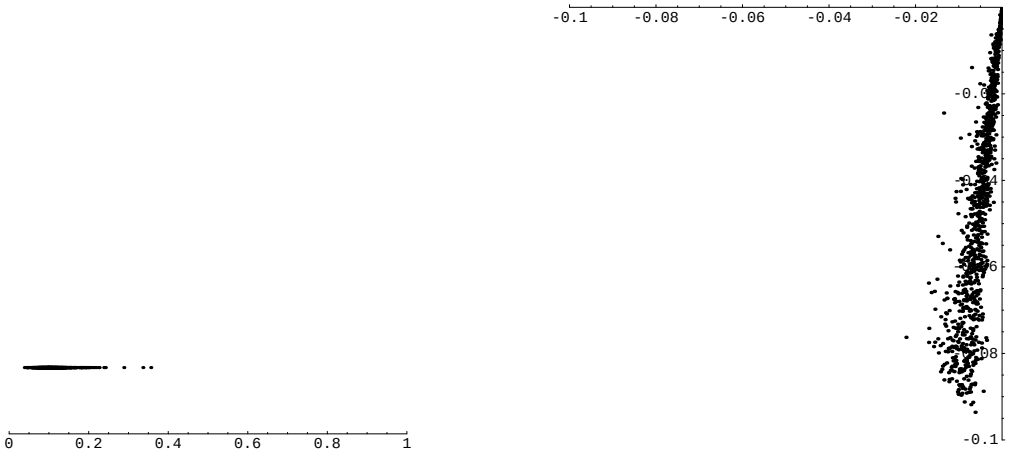
$\Diamond$

Example 3.2.16
For the rest of this section we want to visualize the asymmetric logistic model from Definition 2.3.16 with the abbreviations from Remark 2.3.17. Putting in the case $d = 2$ the parameters to $\psi_1 = 0.9$, $\psi_2 = 0.1$ and $\lambda = 2$, we get the following 1000 angular components and random

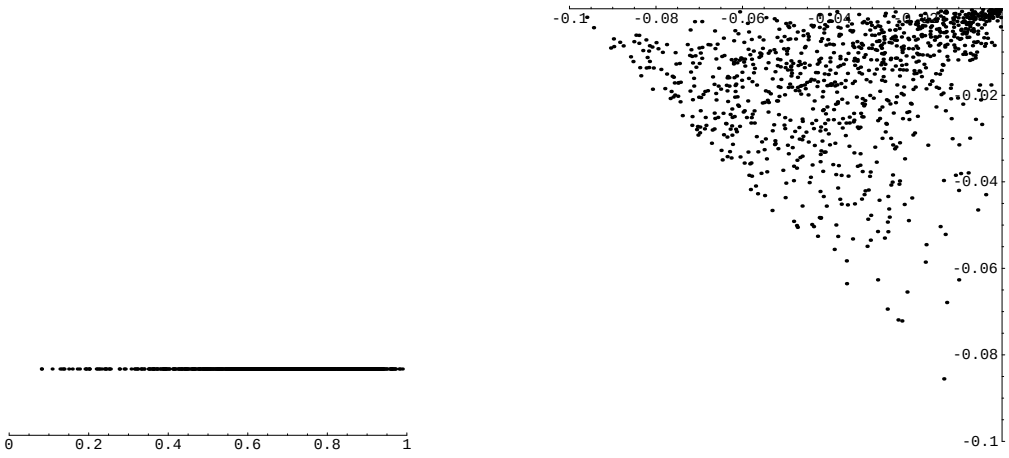
vectors again with $c = -0.1$.



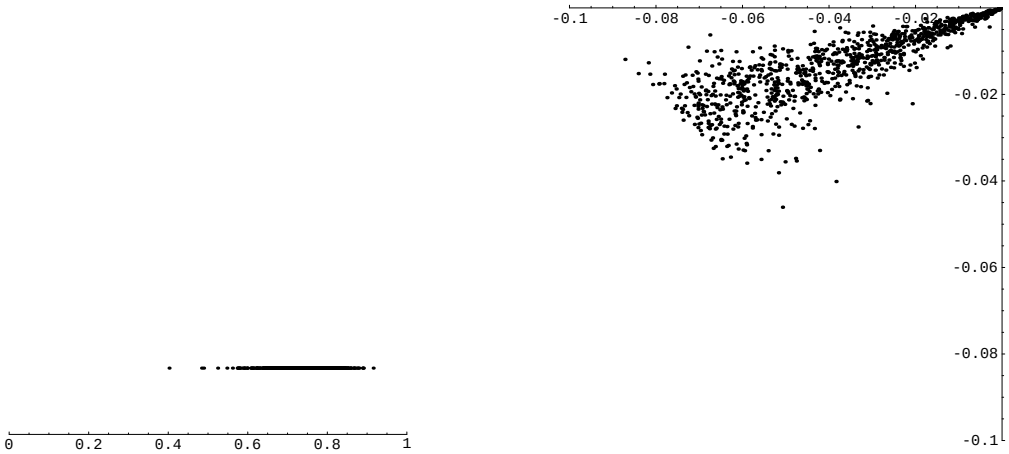
$\psi_1 = 0.9$, $\psi_2 = 0.1$ and $\lambda = 6$.



$\psi_1 = 0.2$, $\psi_2 = 0.6$ and $\lambda = 2$.



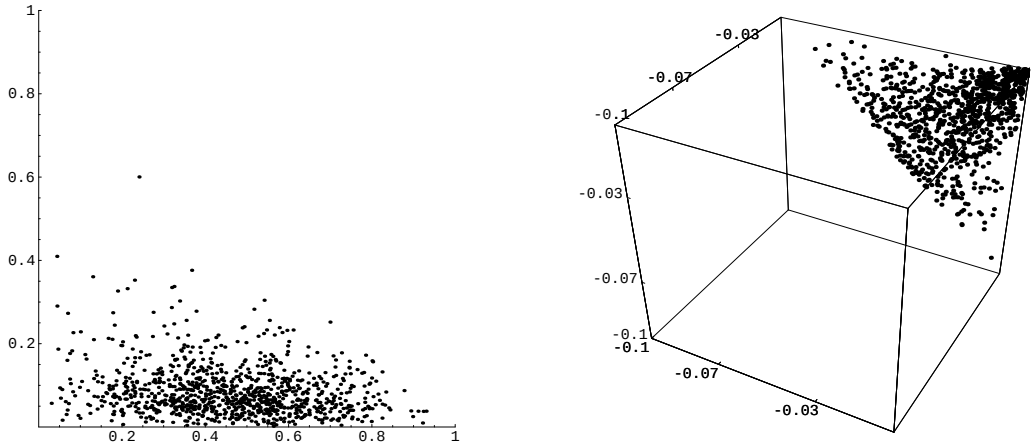
$\psi_1 = 0.2$, $\psi_2 = 0.6$ and $\lambda = 6$.



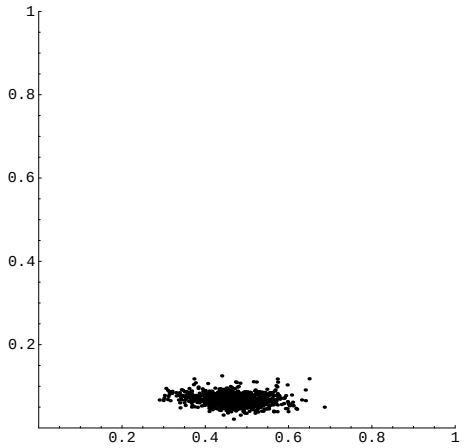
In the logistic case the points always gathered around the bisector of the third quadrant. With the parameters ψ_1, ψ_2 one can determine a gathering line different from the bisector. The parameter λ determines how closely the points stick to that line.

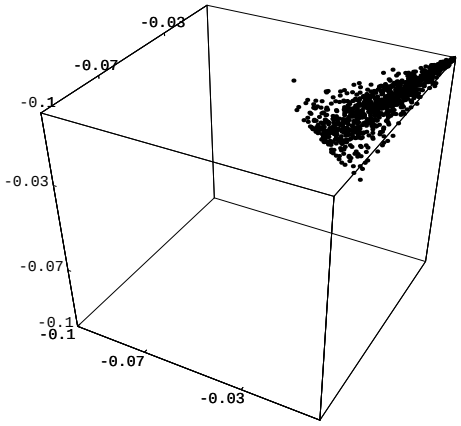
The same behavior can also be observed in the trivariate case. Although there are a large number of parameters with a lot of different possibilities for the structure of W_{as}, we restrict ourselves to two examples. But observe that only the parameters ψ_7, ψ_8, ψ_9 and λ_4 play a role according to Theorem 2.3.20.

$\lambda_4 = 2$, $\psi_8 = 0.7$, $\psi_7 = \psi_9 = 0.1$.



$\lambda_4 = 7$, $\psi_8 = 0.7$, $\psi_7 = \psi_9 = 0.1$.





The parameters ψ_7, ψ_8, ψ_9 govern the center of the angular components in R_2, λ_4 determines how close the points stick to that center.

Due to the complicated nature of the model we, again, skip a runtime analysis.

$\Diamond$

3.3 Simulation of Relative Frequencies of Generalized Pareto Distributions

In Section 6.3 we will introduce an estimation procedure, which will be based on relative frequencies of GPD distributed random vectors. We will test this estimation procedure with the help of simulated data. For that we will have to simulate relative frequencies of GPD distributed random vectors falling into certain sets. However, by Algorithms 3.1.11 and 3.2.11 we are only able to simulate conditional GPDs in a neighborhood of the origin. What we need is a way to simulate the desired relative frequencies with the help of the results of the previous sections.

In this situation, an important property of generalized Pareto distributions, namely the POT-stability, will enable us to do the desired simulations. We present in this section the algorithm together with the theoretical background.

Theorem 3.3.1
Let $X = (X_1, \ldots, X_d)$ be a random vector, which follows a generalized Pareto distribution W with dependence function D. Let $k := dD(1/d, \ldots, 1/d)$ and $t_i \in [-1/k, 0)$, $i = 1, \ldots, d$, such that $\kappa := P(X_i \geq t_i, i \leq d) > 0$ holds. Then we have for $t_i \leq \kappa s_i \leq 0$, $i = 1, \ldots, d$,

$$P(X_i \geq \kappa s_i, i \leq d | X_i \geq t_i, i \leq d) = P(X_i \geq s_i, i \leq d).$$

Proof:
See Section 5.2, page 140 in Falk et al. (2004, [21]).

$\square$

Remark 3.3.2
Since $D(z) \leq 1$ for all $z \in \overline{R_{d-1}}$, we conclude $k \leq d$ and, thus, with the choice of $t_i \in [-1/d, 0)$, the assumption in Theorem 3.3.1 is fulfilled for all generalized Pareto distributions W.

$\Diamond$

Corollary 3.3.3
Let Y be a random vector, which follows a GPD W. Put $\kappa := P(Y \geq t)$, where $t \in \operatorname{diag}\left((-1/d, 0)^d\right)$, and the inequality is meant componentwise. $\operatorname{diag}\left((-1/d, 0)^d\right)$ are thereby all elements in $(-1/d, 0)^d$ with identical components. Let $\kappa > 0$. Furthermore, let X be conditionally generalized Pareto distributed, i.e.,

$$P(X \geq x) = \frac{\overline{W}(x)}{\kappa}$$

for $x \geq t$ again meant componentwise. We denote by $\overline{W}$ the survivor function of W. Then Y and $\frac{X}{\kappa}$ are close to 0 identically distributed and, thus, $\frac{X}{\kappa}$ is generalized Pareto distributed.

Proof:
We have for x close to 0 by Theorem 3.3.1

$$\begin{aligned} P(X \geq x) &= \frac{\overline{W}(x)}{\kappa} = \frac{P(Y \geq x)}{P(Y \geq t)} = P(Y \geq x | Y \geq t) \\ &= P\left(Y \geq \frac{x}{\kappa}\kappa \middle| Y \geq t\right) = P\left(Y \geq \frac{x}{\kappa}\right) = P(\kappa Y \geq x) \end{aligned}$$

Thus X and κY are identically distributed close to 0 and, therefore, also Y and $\frac{X}{\kappa}$.

□

Corollary 3.3.3 provides a straightforward way of getting GPD distributed random vectors from conditionally GPD distributed random vectors. It suffices as in Algorithms 3.1.11 or 3.2.11 to simulate conditional GPDs. A division by κ turns them into unconditionally distributed GPD random vectors, since only the distribution close to the origin is crucial for the definition of a GPD.

Especially, the following algorithm for the simulation of the relative frequency of a set A under a generalized Pareto distribution W is implied by Corollary 3.3.3. Thereby let $t > 0$ be chosen in such a manner that

$$A \subseteq K_t := \left\{ x \in (-\infty, 0)^d \middle| \, ||x||_\infty < t \right\},$$

and the vector $(-t, \ldots, -t)$ fulfills the conditions of Theorem 3.3.1 and Corollary 3.3.3.

Algorithm 3.3.4

1. *Generate n random vectors $x_1, \ldots, x_n$ in the cube K_t, which are conditionally distributed by a generalized Pareto distribution (for example with the help of Algorithm 3.1.11 for the logistic type or Algorithm 3.2.11 for other cases).*
2. *Compute κ and $y_i := \frac{x_i}{\kappa}$, $i = 1, \ldots, n$.*
3. *Count the number m of the y_i with $y_i \in A$.*
4. *Return m/n.*

◊

Chapter 4

Nonparametric Estimation of the Angular Density in Bivariate Generalized Pareto Models

In many applications it is of great importance to have a good insight into the tail dependence structure of a given data set. Consider, for example, the analysis of a portfolio consisting, for simplicity reasons, of only two stocks (X_1, X_2) from two different companies. One is interested in the behavior of the portfolio when the value of the portfolio $X_1 + X_2$ falls short of a certain threshold c. A potential investor would surely like know if there is some kind of dependence between the two stocks, once they are in the extreme area $X_1 + X_2 < c$. If we assume that $-X_1$ and $-X_2$ follow in their upper tail a generalized Pareto distribution, we are in need of tools, which describe and visualize the extremal dependence between the two stocks. To introduce one such possible tool is the goal of this chapter. Then we may be able to get valuable insight into the extremal behavior of the two stocks. Especially in our example one would be interested if it makes economic sense to put them together into one portfolio, i.e., one does not want both stocks to fail at the same time. Since real portfolios usually contain more than two stocks, a treatment of the general multivariate case is of great practical importance and will be dealt with in Chapter 5.

When assuming the data to come from an extreme value or generalized Pareto distribution, there are a lot of objects, which can be estimated to get an insight into the distribution underlying the data. Possibilities are the distribution function itself, its density, the exponent measure, the Pickands dependence function, the angular measure, distribution or density. As we have seen in the pictures of Example 2.3.10, the angular density is a good tool to visualize extremal dependence, since it converges to ∞ at the vertices of the unit simplex close to the independence case, and to 0 close to the dependence case. Therefore, the angular density will be the target of our nonparametric estimation efforts in Chapters 4 and 5, and by it we want to gain a tool for the visualization of the extremal dependence structure. Estimating the angular density for that purpose is also popular in extreme value models, see for example Coles and Tawn (1991, [10]), Coles and Tawn (1994, [11]) or Coles et al. (1999, [9]).

There is a vast literature concerning the nonparametric estimation of the objects mentioned above in the extreme value case. Most of these are concerned with finding nonparametric estimators for the Pickands dependence function but also the angular distribution and the exponent measure. An enumeration of the most important papers can be found in the Chapter 1. Sum-

maries of the estimators in the literature are given for example in Section 3.6 of Kotz and Nadarajah (2000, [47]) or Sections 9.3 and 9.4 of Beirlant et al. (2005, [3]). If the estimators of the Pickands dependence function presented there are smooth enough, one can, in the bivariate case, derive estimators for the angular density by just twice differentiating them.

The corresponding literature for estimations in the generalized Pareto case is rare. The only known sources to the author are Section 5.2 of Falk et al. (2004, [21]) and Section 10.2 of Reiss and Thomas (2001, [57]), where the Pickands dependence function and the canonical dependence function of GPD random vectors are estimated.

We begin in Section 4.1 with the estimation in the bivariate case, which will turn out to be a unique case. We can estimate the angular density from observations, of which the exact density, a suitably scaled version of the Pickands density, is known. This is in contrast to Chapter 5 where we estimate with observations of which only the asymptotic density is known.

Section 4.1.1 will give a representation of the angular density by means of the Pickands density. Therefore, the Pickands density will be estimated nonparametrically in Section 4.1.2 with an appropriate kernel density estimator and automatic bandwidth selection. Then we can present a nonparametric estimator of the angular density in Section 4.1.3.

All these estimators will be tested with random vectors generated from Algorithm 3.1.11, and their asymptotic normality will be shown under suitable regularity conditions.

In Section 4.2 we will see that the method presented in this chapter cannot be transferred to higher dimensions. An alternative in the multivariate case will be given in Chapter 5.

An application of the method developed here to a hydrological data set will be given in Chapter 7.

4.1 The Bivariate Case

4.1.1 Representation of the Angular Density in the Bivariate Case

Theorem 4.1.1
Let (X_1, X_2) follow a bivariate generalized Pareto distribution W, which is twice differentiable. Let the quantity μ from Theorem 2.2.5 fulfill $\mu > 0$. Then we have for the angular density

$$l(z) = \frac{\phi(1-z)}{z(1-z)} = \frac{\mu \cdot f(1-z)}{z(1-z)},$$

where ϕ is the Pickands density and μ and f are defined in Theorem 2.2.5.

Proof:
From Theorem 2.2.4 we know that

$$l\left(\frac{\frac{1}{x_1}}{\frac{1}{x_1}+\frac{1}{x_2}}\right) = \frac{x_1^2 x_2^2}{\left(-\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\right)^{-3}} \frac{\partial^2}{\partial x_1 \partial x_2} W(x_1, x_2).$$

Inserting the inverse Pickands transformation (2.2) we get

$$l\left(\frac{\frac{1}{cz}}{\frac{1}{cz}+\frac{1}{c(1-z)}}\right) = -\frac{(cz)^2(c(1-z))^2}{\left(\frac{1}{cz}+\frac{1}{c(1-z)}\right)^{-3}} \cdot \frac{\partial^2}{\partial x_1 \partial x_2} W\left(T_P^{-1}(z,c)\right)$$

$$\begin{aligned}
\Leftrightarrow l\left(\frac{\frac{1}{z}}{\frac{1}{z}+\frac{1}{1-z}}\right) &= -\frac{c^4 z^2(1-z)^2 c^{-3}}{\left(\frac{1}{z}+\frac{1}{1-z}\right)^{-3}} \cdot \frac{\partial^2}{\partial x_1 \partial x_2} W\left(T_P^{-1}(z,c)\right) \\
\Leftrightarrow l\left(\frac{\frac{1}{z}}{\frac{1-z+z}{z(1-z)}}\right) &= (-c)\frac{z^2(1-z)^2}{\left(\frac{1-z+z}{z(1-z)}\right)^{-3}} \cdot \frac{\partial^2}{\partial x_1 \partial x_2} W\left(T_P^{-1}(z,c)\right) \\
\Leftrightarrow l\,(1-z) &= (-c)\frac{z^2(1-z)^2}{(z(1-z))^3} \cdot \frac{\partial^2}{\partial x_1 \partial x_2} W\left(T_P^{-1}(z,c)\right) \\
\Leftrightarrow l\,(1-z) &= \frac{|c|}{z(1-z)} \cdot \frac{\partial^2}{\partial x_1 \partial x_2} W\left(T_P^{-1}(z,c)\right).
\end{aligned}$$

With the definition of the Pickands density ϕ from Theorem 2.2.5 we get

$$l\,(1-z) = \frac{\phi(z)}{z(1-z)}.$$

Replacing z by $1-z$ and using the definition of f from Theorem 2.2.5 we finally have the assertion

$$l\,(z) = \frac{\phi(1-z)}{z(1-z)} = \frac{\mu \cdot f(1-z)}{z(1-z)}.$$

□

Remark 4.1.2
If the random variables X_1, X_2 are exchangeable, i.e., if (X_1, X_2), and (X_2, X_1) have the same distribution, then the assertion of Theorem 4.1.1 reduces to

$$l(z) = \frac{\phi(z)}{z(1-z)} = \frac{\mu \cdot f(z)}{z(1-z)}.$$

◊

4.1.2 Estimation of the Pickands Density

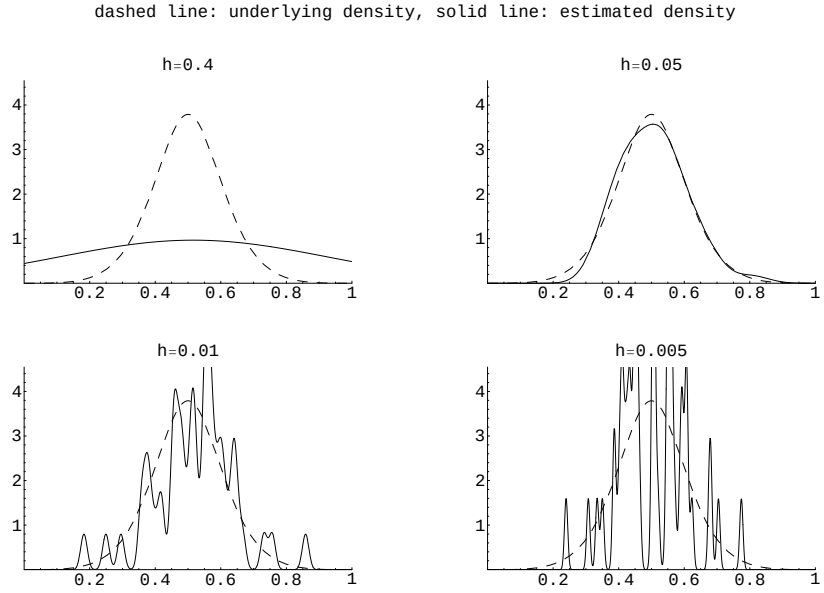
Let $(\tilde{X}_1, \tilde{X}_2) < 0$ be a bivariate random vector following a generalized Pareto distribution W. Suppose that we have n independent copies $(\tilde{X}_{1i}, \tilde{X}_{2i})$ of $(\tilde{X}_1, \tilde{X}_2)$, and denote by $\tilde{Z}_i := \tilde{X}_{1i}/(\tilde{X}_{1i} + \tilde{X}_{2i})$ and $\tilde{C}_i := \tilde{X}_{1i} + \tilde{X}_{2i}$ the corresponding bivariate Pickands coordinates, $i = 1, \ldots, n$. Fix a threshold c_0 close to 0, and consider only those observations $(\tilde{X}_{1i}, \tilde{X}_{2i})$ with $\tilde{C}_i > c_0$. Denote these by $(X_{11}, X_{21}), \ldots, (X_{1m}, X_{2m})$, where $m = \tau_n$ is the random number of observations with $\tilde{C}_i > c_0$. From Theorem 1.4.1 in Reiss (1993, [56]) we know that τ_n and the (X_{1j}, X_{2j}) are all independent random variables, that τ_n is binomial $B(n,p)$ distributed with $p = P(\tilde{C} > c_0)$, and the Z_j have by Theorem 2.2.5 the density $f(z)$.

A natural estimator of f is the kernel density estimator with kernel function k and bandwidth $h > 0$

$$\hat{f}_m(z) := \frac{1}{mh} \sum_{i=1}^{m} k\left(\frac{z - Z_i}{h}\right), \tag{4.1}$$

where $Z_i := X_{1i}/(X_{1i} + X_{2i})$, $i = 1, \ldots, m$. For basic information on kernel density estimators see Chapters 1 and 6 of Falk et al. (2002, [23]), and also the discussion in Section 5.2.

Example 4.1.3
Taking k to be the normal kernel $k(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ we did simulations of estimator (4.1) using Algorithm 3.1.11 for the simulation of GPDs of logistic type. The following figures show the results for different h with $m = 50$, $\lambda = 4$ and $c_0 = -0.1$.



As one can see from the pictures, the choice of a suitable bandwidth h is crucial, which is a common problem with kernel density estimators. This bandwidth is highly dependent on the density itself. So there is the need for an automatic generation of an appropriate bandwidth.

In the literature there is a broad variety of methods to determine an optimal bandwidth, which are also used by software packages such as SAS. Jones et al. (1996, [44]) investigate several advanced methods for the choice of h and recommend a method described in Sheather and Jones (1991, [66]). This computation of h, however, requires a certain fixed point equation to be solved and is, thus, unsuited for theoretical considerations and straightforward practical implementation.

If one assumes the data to be normally distributed and if one chooses the normal kernel, this method reduces to the choice

$$h = S_m \left(\frac{4}{3m}\right)^{1/5}, \tag{4.2}$$

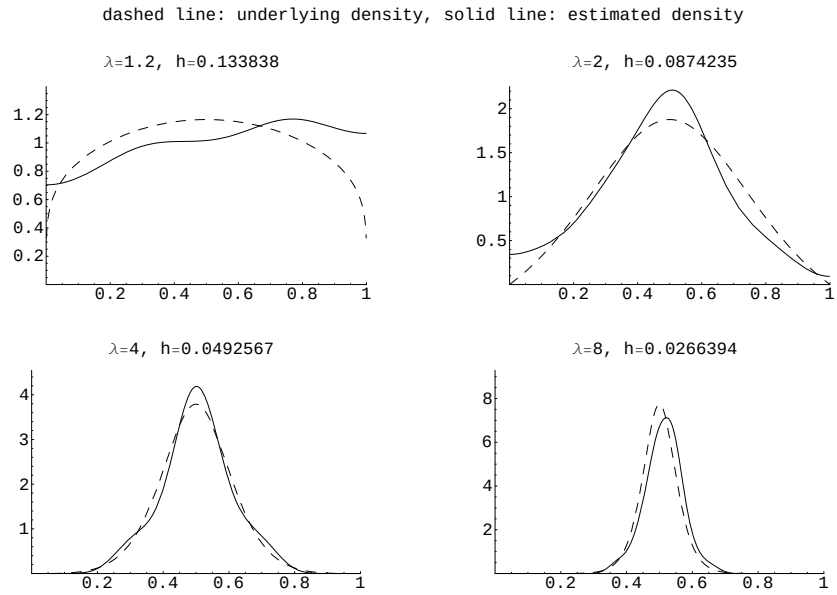
with

$$S_m := \left(\frac{1}{m-1}\sum_{i=1}^{m}(Z_i - \bar{Z}_m)^2\right)^{1/2}, \qquad \bar{Z}_m := \frac{1}{m}\sum_{i=1}^{m} Z_i\,,$$

the empirical standard deviation and the arithmetic mean of the Z_j. In our case we do not have normal data. But as can be seen from the pictures in Example 2.3.8 and those below, the densities of the Z_j have, at least in the logistic case, some similarity to the shape of a normal density. So it is natural to use the choice of h according to (4.2) and to see, whether this is a good one. A lot of simulations we did, indicate that this is actually true. We will discuss these topics of kernel choice and automatic bandwidth selection in Section 5.2 in more detail, when we come to the general multivariate case.

Since we have a bounded support of the density, which is to be estimated, one should use reflection techniques, see Section 2.1 of Reiss and Thomas (2001, [57]), to further improve the estimation. Thereby the data set is reflected at the boundaries of the support, in this case 0 and 1, and then the density estimator for the enlarged data set is computed. This is done for all following estimations, see also the general discussion in Section 5.2.

Here are some results, again for the logistic case with $m = 50$, $c_0 = -0.1$ but this time for different λ and the corresponding h is given.



◊

With the special choice of h given in (4.2) formula (4.1) becomes

$$\hat{f}_m(z) = \frac{1}{mS_m\left(\frac{4}{3m}\right)^{1/5}} \sum_{j=1}^{m} k\left(\frac{z - Z_j}{S_m\left(\frac{4}{3m}\right)^{1/5}}\right) = \frac{1}{mh'S_m} \sum_{j=1}^{m} k\left(\frac{z - Z_j}{h'S_m}\right),$$

with $h' = \left(\frac{4}{3m}\right)^{1/5}$. So for practical purposes we advocate a kernel density estimator of the form

$$\hat{f}_m^s(z) := \frac{1}{mhS_m} \sum_{j=1}^{m} k\left(\frac{z - Z_j}{hS_m}\right).$$

$\hat{f}_m^s(z)$ is a special case of a kernel density estimator with data sphering, see again the general discussion in Section 5.2 for more details.

We will shortly note a result on the theoretical behavior of the estimator (4.1), which we get from Powell (2002, [54]). For simplicity reasons we abstain from including the data sphering in our theoretical considerations. The regularity assumptions on the Pickands and, thus, the angular density and the kernel will be given later in detail in Section 5.4 for the general multivariate case. For the next theorem we just refer to the corresponding passages.

Theorem 4.1.4
Let the Pickands density ϕ have a Taylor expansion of the form (5.15) at a point $z \in (0,1)$, where $\phi(z) \neq 0$, and let the kernel k have bounded support and fulfill (5.17). If $h = o\left(m^{-1/5}\right)$, then

$$\sqrt{mh}\left(\hat{f}_m(z) - f(z)\right) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, f(z)\int k^2(u)\,du\right).$$

Proof:
See Powell (2002, [54]).

□

Remark that the normal kernel does not have bounded support and, thus, Theorem 4.1.4 does not hold there. But it has the advantage of a simple automatic bandwidth selection. For this reason, it is used throughout this manuscript, see again the detailed discussion in Section 5.2.

4.1.3 Estimation of the Angular Density

By estimating f we are now able to estimate the angular density and derive a graphical tool for the investigation of the tail dependence structure.

We obtain from Theorem 4.1.1 that

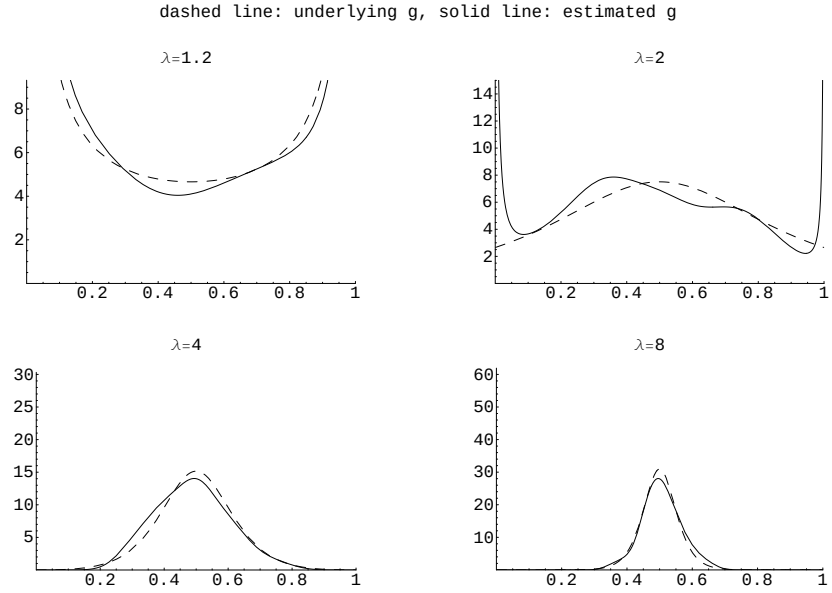
$$g(z) := \frac{f(1-z)}{z(1-z)} = \frac{l(z)}{\mu}, \quad 0 \leq z \leq 1.$$

The function g, which is a constant multiple of the angular density l, visualizes if the underlying distribution of (X_1, X_2) is closer to the case of independence or the case of dependence. A peak of $g(z)$ near 0 and 1 indicates that our observations come from a distribution which is closer to the independence case, whereas a peak in the interior of the unit interval visualizes that we are closer to the dependence case, see the remarks after Example 2.3.10. In the logistic case for λ close to 1, we have convergence to ∞ as we approach the boundary of $[0,1]$, and for large λ, we have convergence to 0 at the boundary and a single peak at $\frac{1}{2}$.

With the ability to estimate $f(z)$ by $\hat{f}_m(z)$, we have also gained the ability to estimate $g(z)$ by

$$\hat{g}_m(z) := \frac{\hat{f}_m(1-z)}{z(1-z)}. \tag{4.3}$$

Example 4.1.5
The estimator (4.3) was simulated in the following graphic for the logistic case with $m = 50$, $c_0 = -0.1$ and different λ.



Once again data sphering and reflection techniques were included for practical purposes.

For λ noticeably smaller or larger than 2, the functions g and $\hat{g}_m$ have the same behavior at the boundary. For λ close to 2, the angular density g and its estimator $\hat{g}_m$ seem to behave differently when approaching the boundary. This is due to numerical effects coming from the division by $z(1-z)$. Convergence of $\hat{g}_m$ to 0 when approaching the boundary is a clear sign of dependence. In contrast to this, one has to be careful when $\hat{g}_m$ tends to ∞ at the boundary. This can happen up to $\lambda = 3$, although $\lambda = 3$ means that the data are quite dependent. If this convergence to ∞ begins close to the boundary, we have a larger λ (see for example the figure with $\lambda = 2$), if the convergence begins away from the boundary we have a small λ (see the example with $\lambda = 1.2$) and, thus, nearly tail independence.

$\Diamond$

Theorem 4.1.6
Let the assumptions of Theorem 4.1.4 hold. Then we have

$$\sqrt{mh}\left(\hat{g}_m(z) - \frac{l(z)}{\mu}\right) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, \frac{f(1-z)}{z^2(1-z)^2}\int k^2(u)\,du\right).$$

Proof:
With Theorem 4.1.4 we get

$$\begin{aligned}\sqrt{mh}\left(\hat{g}_m(z) - \frac{l(z)}{\mu}\right) &= \sqrt{mh}\,(\hat{g}_m(z) - g(z)) = \sqrt{mh}\left(\frac{\hat{f}_m(1-z)}{z(1-z)} - \frac{f(1-z)}{z(1-z)}\right) \\ &\longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, \frac{f(1-z)}{z^2(1-z)^2}\int k^2(u)\,du\right).\end{aligned}$$

$\square$

Remark 4.1.7
Note that the distribution of the angular components of the Pickands coordinates is independent of the chosen threshold c_0. This could also be a tool for the graphical verification of the GPD model assumption by using different thresholds and comparing the resulting estimators of the angular density. If the data actually follow a GPD, then all estimators should basically give the same graphic. If the graphics differ heavily, then one can have doubts about the GPD model assumption. Such considerations are also used to check the EVD approximation in threshold models by Joe et al. (1992, [43]) and Coles and Tawn (1994, [11]).

$\Diamond$

4.2 The Problem of Generalization to the Trivariate Case

We have seen in Remark 4.1.2 that we could find a multiplicative decomposition of the angular density l into

$$l(z) = \kappa \frac{\phi(z)}{z(1-z)},$$

in the case of exchangeability, where κ is a constant depending only on l but not on z. The natural generalization to the trivariate case would be that

$$l(z_1, z_2) = \kappa \frac{\phi(z_1, z_2)}{z_1 z_2 (1 - z_1 - z_2)}$$

for exchangeable models where again κ is a constant only depending on l. We will, however, see that this equation does not hold. In fact, we will see that there is no multiplicative decomposition of a differentiable l such that

$$l(z_1, z_2) = \kappa b(z_1, z_2)\phi(z_1, z_2),$$

with κ depending only on l, and b differentiable of order 1, depending only on z_1 and z_2 but not on l.

To show this, we will use the logistic case and disprove the equation

$$l_\lambda(z_1, z_2) = \kappa(\lambda) b(z_1, z_2)\phi_\lambda(z_1, z_2).$$

Theorem 4.2.1
There does not exist a function $\kappa : (1, \infty) \to \mathbb{R}$ and a differentiable function $b(z_1, z_2) : R_2 \to \mathbb{R}$ such that for $\lambda > 1$ the decomposition

$$l_\lambda(z_1, z_2) = \kappa(\lambda) b(z_1, z_2)\phi_\lambda(z_1, z_2)$$

holds, where l_λ is the angular density and ϕ_λ the Pickands density of a logistic GPD with parameter λ.

Proof:
Suppose that there exist functions κ and b such that

$$l_\lambda(z_1, z_2) = \kappa(\lambda) b(z_1, z_2)\phi_\lambda(z_1, z_2), \quad (z_1, z_2) \in R_2.$$

From representation (2.17) in Theorem 2.3.7, we know that $\phi_\lambda > 0$ for $\lambda > 1$. Dividing by ϕ_λ we get

$$\frac{l_\lambda(z_1, z_2)}{\phi_\lambda(z_1, z_2)} = \kappa(\lambda) b(z_1, z_2).$$

Since also $l_\lambda(z_1, z_2) > 0$ by Example 2.3.10 for any $(z_1, z_2) \in R_2$, we can assume without loss of generality that $\kappa(\lambda) > 0$ and $b(z_1, z_2) > 0$. Therefore, we get

$$\log\left(\frac{l_\lambda(z_1, z_2)}{\phi_\lambda(z_1, z_2)}\right) = \log(\kappa(\lambda)) + \log(b(z_1, z_2))$$

by taking the logarithm on both sides. Computing the partial derivative with respect to z_1 (or to z_2 leading to the same results due to the exchangeability of z_1 and z_2 in the logistic case), we have

$$a(\lambda, z_1, z_2) := \frac{\partial}{\partial z_1} \log\left(\frac{l_\lambda(z_1, z_2)}{\phi_\lambda(z_1, z_2)}\right) = \frac{\frac{\partial}{\partial z_1} b(z_1, z_2)}{b(z_1, z_2)},$$

which is constant with regard to λ. This will be our contradiction since we will show that there exist $\lambda_1, \lambda_2 > 1$ and $(z_1, z_2) \in R_2$ such that

$$a(\lambda_1, z_1, z_2) \neq a(\lambda_2, z_1, z_2).$$

With the representations of l_λ and ϕ_λ from Example 2.3.10 and Theorem 2.3.7, we compute

$$\begin{aligned}
a(\lambda, z_1, z_2) &= \frac{\partial}{\partial z_1} \log\left(\frac{(z_1 z_2)^{-\lambda-1}(1-z_1-z_2)^{-\lambda-1}\left(z_1^{-\lambda} + z_2^{-\lambda} + (1-z_1-z_2)^{-\lambda}\right)^{1/\lambda-3}}{(z_1 z_2)^{\lambda-1}(1-z_1-z_2)^{\lambda-1}\left(z_1^{\lambda} + z_2^{\lambda} + (1-z_1-z_2)^{\lambda}\right)^{1/\lambda-3}}\right) \\
&= \frac{\partial}{\partial z_1} \log\left((z_1 z_2)^{-2\lambda}(1-z_1-z_2)^{-2\lambda}\left(\frac{z_1^{-\lambda} + z_2^{-\lambda} + (1-z_1-z_2)^{-\lambda}}{z_1^{\lambda} + z_2^{\lambda} + (1-z_1-z_2)^{\lambda}}\right)^{1/\lambda-3}\right) \\
&= \frac{\partial}{\partial z_1}\Big(-2\lambda \log(z_1 z_2) - 2\lambda \log(1-z_1-z_2) \\
&\quad + \left(\frac{1}{\lambda} - 3\right)\log\left(z_1^{-\lambda} + z_2^{-\lambda} + (1-z_1-z_2)^{-\lambda}\right) \\
&\quad - \left(\frac{1}{\lambda} - 3\right)\log\left(z_1^{\lambda} + z_2^{\lambda} + (1-z_1-z_2)^{\lambda}\right)\Big) \\
&= -2\lambda\frac{z_2}{z_1 z_2} - 2\lambda\frac{-1}{1-z_1-z_2} + \left(\frac{1}{\lambda} - 3\right)\frac{-\lambda z_1^{-\lambda-1} - \lambda(1-z_1-z_2)^{-\lambda-1}(-1)}{z_1^{-\lambda} + z_2^{-\lambda} + (1-z_1-z_2)^{-\lambda}} \\
&\quad - \left(\frac{1}{\lambda} - 3\right)\frac{\lambda z_1^{\lambda-1} + \lambda(1-z_1-z_2)^{\lambda-1}(-1)}{z_1^{\lambda} + z_2^{\lambda} + (1-z_1-z_2)^{\lambda}} \\
&= 2\lambda\left(\frac{1}{1-z_1-z_2} - \frac{1}{z_1}\right) + (3\lambda-1)\frac{z_1^{-\lambda-1} - (1-z_1-z_2)^{-\lambda-1}}{z_1^{-\lambda} + z_2^{-\lambda} + (1-z_1-z_2)^{-\lambda}} \\
&\quad + (3\lambda-1)\frac{z_1^{\lambda-1} - (1-z_1-z_2)^{\lambda-1}}{z_1^{\lambda} + z_2^{\lambda} + (1-z_1-z_2)^{\lambda}}.
\end{aligned}$$

By choosing $z_1 = z_2 = \frac{1}{4}$ we get

$$\begin{aligned}
a\left(\lambda, \frac{1}{4}, \frac{1}{4}\right) &= 2\lambda\left(\frac{1}{\frac{1}{2}} - \frac{1}{\frac{1}{4}}\right) + (3\lambda-1)\frac{\left(\frac{1}{4}\right)^{-\lambda-1} - \left(\frac{1}{2}\right)^{-\lambda-1}}{\left(\frac{1}{4}\right)^{-\lambda} + \left(\frac{1}{4}\right)^{-\lambda} + \left(\frac{1}{2}\right)^{-\lambda}} \\
&\quad + (3\lambda-1)\frac{\left(\frac{1}{4}\right)^{\lambda-1} - \left(\frac{1}{2}\right)^{\lambda-1}}{\left(\frac{1}{4}\right)^{\lambda} + \left(\frac{1}{4}\right)^{\lambda} + \left(\frac{1}{2}\right)^{\lambda}}
\end{aligned}$$

$$
\begin{aligned}
&= -4\lambda + (3\lambda - 1)\frac{4^{\lambda+1} - 2^{\lambda+1}}{4^{\lambda} + 4^{\lambda} + 2^{\lambda}} + (3\lambda - 1)\frac{\left(\frac{1}{4}\right)^{\lambda-1} - \left(\frac{1}{2}\right)^{\lambda-1}}{\left(\frac{1}{4}\right)^{\lambda} + \left(\frac{1}{4}\right)^{\lambda} + \left(\frac{1}{2}\right)^{\lambda}} \\
&= -4\lambda + (3\lambda - 1)\frac{2^{\lambda+1}\left(2^{\lambda+1} - 1\right)}{2^{\lambda}\left(2 \cdot 2^{\lambda} + 1\right)} + (3\lambda - 1)\frac{\left(\frac{1}{2}\right)^{\lambda-1}\left(\left(\frac{1}{2}\right)^{\lambda-1} - 1\right)}{\left(\frac{1}{2}\right)^{\lambda}\left(2\left(\frac{1}{2}\right)^{\lambda} + 1\right)} \\
&= -4\lambda + 2(3\lambda - 1)\left(\frac{2^{\lambda+1} - 1}{2^{\lambda+1} + 1} + \frac{1 - 2^{\lambda-1}}{2^{\lambda-1} + 1}\right).
\end{aligned}
$$

If we look now at

$$
a\left(2, \frac{1}{4}, \frac{1}{4}\right) = -8 + 10\left(\frac{7}{9} + \frac{-1}{3}\right) = -8 + 10 \cdot \frac{4}{9} = -\frac{32}{9} \approx -3.556
$$

and

$$
a\left(3, \frac{1}{4}, \frac{1}{4}\right) = -12 + 16 \cdot \left(\frac{15}{17} + \frac{-3}{5}\right) = -\frac{636}{85} \approx -7.482,
$$

we see that

$$
a\left(2, \frac{1}{4}, \frac{1}{4}\right) \neq a\left(3, \frac{1}{4}, \frac{1}{4}\right),
$$

which completes the proof.

□

Therefore, the results and procedures of this chapter cannot be analogously transferred to the multivariate case. The bivariate case seems insofar to be a special case. However, in Chapter 5 we will give procedures, which will approximate the results of this chapter.

Chapter 5

Nonparametric Estimation of the Angular Density in Multivariate Generalized Pareto Models

In this chapter we will generalize the results of the previous chapter to an arbitrary dimension $d \geq 2$. As we have seen in Section 4.2, this is not possible in a straightforward manner. Instead, we have to use a slightly modified approach.

In Section 5.1 we will see that the distribution of the angular component of the Pickands coordinates with regard to Fréchet margins is, under weak regularity conditions, close to the angular distribution. This will be used in Section 5.2 to estimate the angular density with kernel methods as in Chapter 4. This estimator will be an analogue of an estimator introduced in Joe et al. (1992, [43]) and in Coles and Tawn (1994, [11]) from an EVD point of view.

In Section 5.3 we will introduce δ-neighborhoods of GPDs, to which EVDs belong, and show that the estimation procedure can be carried over to these neighborhoods. Sections 5.4 to 5.6 will be concerned with the asymptotic normality of our estimator.

Finally in Section 5.7 we investigate the quantity μ from Theorem 2.2.5, which is the integral of the Pickands density. The result gained in this section is needed in Section 3.2.3 during the investigation of runtimes. It can only be shown here, since the tools necessary for its derivation will be introduced during the course of this chapter.

5.1 Distribution of the Pickands Coordinates with Regard to Fréchet Margins in Generalized Pareto Models

The goal in this section is to generate random vectors, whose density is close to the angular density. For that purpose we will need the Pickands transformation with regard to Fréchet margins. The following lemma shows that the additional transformation to Fréchet margins is crucial.

Lemma 5.1.1
Let the random vector $(X_1, \ldots, X_d)$ follow a generalized Pareto distribution with uniform margins, i.e., its distribution function W has the representation (2.5) for a neighborhood U of 0, which can be assumed to be a cube (possibly after a suitable reduction). Let the Pickands

dependence function D have continuous partial derivatives of order d. Then the Pickands transformation with regard to Fréchet margins $T_F(X_1, \ldots, X_d) = (Z, C)$ has a density $f(z, c)$ on $T_F(U)$, which can be factorized with regard to z and c. More precisely

$$f(z, c) = c^{-2} l(z), \qquad (z, c) \in T_F(U),$$

where l is the angular density of W and is continuous.

Proof:
The distribution function W possesses for $x \in U$ the density

$$\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W(x_1, \ldots, x_d),$$

see Theorem A.2.2 in Bhattacharya and Rao (1976, [4]). According to the density transformation theorem (see for example Fristedt and Gray (1997, [28]), Section 9.5), the density of Pickands coordinates with regard to Fréchet margins has the form

$$f(z, c) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z, c)\right) \left|\det J_{T_F^{-1}}(z, c)\right|, \tag{5.1}$$

where f as concatenation of continuous functions is likewise continuous.

We will now establish the connection between the density f and the angular density l. From Theorem 2.2.4 we know that the angular density l has the representation

$$l\left(\frac{\frac{1}{x_1}}{\sum_{i=1}^d \frac{1}{x_i}}, \ldots, \frac{\frac{1}{x_{d-1}}}{\sum_{i=1}^d \frac{1}{x_i}}\right) = \frac{x_1^2 \cdot \ldots \cdot x_d^2}{\left(-\sum_{i=1}^d \frac{1}{x_i}\right)^{-(d+1)}} \cdot \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W(x_1, \ldots, x_d).$$

If one inserts the Pickands coordinates with regard to Fréchet margins from Definition 2.1.4, we get

$$\begin{aligned} l(z_1, \ldots, z_{d-1}) &= \frac{\frac{1}{c^2 z_1^2} \cdot \ldots \cdot \frac{1}{c^2 z_{d-1}^2} \cdot \frac{1}{c^2\left(1 - \sum_{i=1}^{d-1} z_i\right)^2}}{(-c)^{-(d+1)}} \cdot \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z, c)\right) \\ &= (-c)^{-d+1} \frac{1}{z_1^2} \cdot \ldots \cdot \frac{1}{z_{d-1}^2} \cdot \frac{1}{\left(1 - \sum_{i=1}^{d-1} z_i\right)^2} \cdot \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z, c)\right). \end{aligned}$$

We can now show with Lemma 2.1.5 the announced relation with the density f.

$$\begin{aligned} l(z) &= \underbrace{c^2 (-c)^{-d-1} \frac{1}{z_1^2} \cdot \ldots \cdot \frac{1}{z_{d-1}^2} \cdot \frac{1}{\left(1 - \sum_{i=1}^{d-1} z_i\right)^2}}_{= \left|\det J_{T_F^{-1}(z,c)}\right|} \cdot \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z, c)\right) \\ &= \underbrace{c^2 \left|\det J_{T_F^{-1}(z,c)}\right| \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z, c)\right)}_{= f(z,c)} \\ &\overset{(5.1)}{=} c^2 f(z, c). \end{aligned}$$

Thus $l(z) : R_{d-1} \to \mathbb{R}$ is also continuous, and we have the representation

$$f(z, c) = c^{-2} l(z).$$

□

That the exponent measure function of an extreme value or generalized Pareto distribution factorizes across radial and angular components of pseudo polar coordinates is a well-known result, see for example Section 4.2 of Falk et al. (2004, [21]) or Section 5.4 of Resnick (1987, [59]). Since a GPD and its exponent measure function are closely connected by (2.5) and (2.8), the factorization of the density of a GPD in Lemma 5.1.1 is not really surprising. In the sequel, we will investigate the consequences of this factorization.

Before we proceed, we have to introduce some notations. With

$$K_s := \{x \in (-\infty, 0)^d \mid ||x||_\infty < s\}, \quad s > 0$$

we designate the (open) cube with edge length s in the negative quadrant. For r, $s > 0$, let

$$Q_{r,s} := \left\{z \in R_{d-1} \mid T_F^{-1}(z, -r) \in K_s\right\} \tag{5.2}$$

be the set of z–coordinates of the Pickands transformation with regard to Fréchet margins of the points in the cube K_s, whose c–coordinate has the value $-r$.

Lemma 5.1.2
Let $s > 0$ be fixed. Then $Q_{r,s}$ has the representation

$$Q_{r,s} = \left\{(z_1, \ldots, z_{d-1}) \in R_{d-1} \,\middle|\, z_i > \frac{1}{rs}, i = 1, \ldots, d-1, \sum_{i=1}^{d-1} z_i < 1 - \frac{1}{rs}\right\}. \tag{5.3}$$

Furthermore, we have $Q_{r,s} \neq \emptyset$ for all $r > \frac{d}{s}$, $Q_{r_1,s} \subseteq Q_{r_2,s}$ for $r_1 < r_2$ and $\lim_{r\to\infty} Q_{r,s} = R_{d-1}$.

Proof:
The assertion $z \in Q_{r,s}$ is valid according to the definition of $Q_{r,s}$ if and only if

$$s > ||T_F^{-1}(z, -r)||_\infty = \frac{1}{r} \max\left\{\frac{1}{z_1}, \ldots, \frac{1}{z_{d-1}}, \frac{1}{1 - \sum_{i=1}^{d-1} z_i}\right\},$$

thus if and only if

$$\begin{aligned} & s > \frac{1}{r} \cdot \frac{1}{z_i}, \quad i = 1, \ldots, d-1 \qquad \text{and} \qquad s > \frac{1}{r} \cdot \frac{1}{1 - \sum_{i=1}^{d-1} z_i} \\ \iff \quad & z_i > \frac{1}{rs}, \quad i = 1, \ldots, d-1 \qquad \text{and} \qquad \sum_{i=1}^{d-1} z_i < 1 - \frac{1}{rs}. \end{aligned}$$

With the representation (5.3) the other assertions follow immediately.

□

The quantity introduced next will play an important role in the asymptotic considerations, which are to follow.

Definition 5.1.3
Let l be the angular density of some GPD W. Define

$$\chi(r, s) := \int_{Q_{r,s}} l(z)\, dz \tag{5.4}$$

for $r\,, s > 0$.

◊

Lemma 5.1.4
We have

$$\lim_{r\to\infty} \chi(r,s) = d^*$$

monotone increasing with d^* from (2.11).

Proof:
Because of (2.11) and Lemma 5.1.2, the dominated convergence theorem (see for example Fristedt and Gray (1997, [28]), Section 8.2) is applicable, and, thus,

$$\lim_{r\to\infty} \chi(r,s) = \lim_{r\to\infty} \int_{Q_{r,s}} l(z)\,dz = \int_{R_{d-1}} l(z)\,dz \overset{(2.11)}{=} d^*.$$

Since $l(z) \geq 0$ the monotonicity follows.

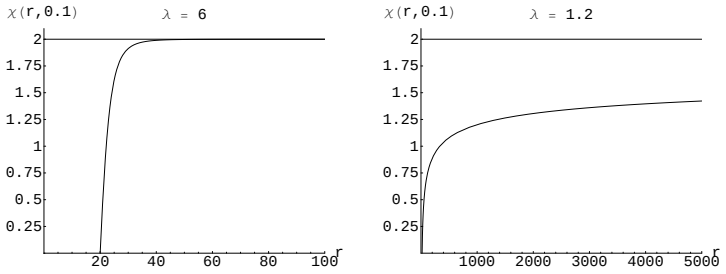
□

As we will see later, the rate of convergence of χ heavily influences the rates of convergence of our estimation procedures. For details see Sections 5.4 to 5.6. These rates of convergence can be very different, as we will show in the following example.

Example 5.1.5
To show that the rates of convergence of $\chi(r,s)$ can be very different, we consider the bivariate case

$$\chi(r,s) = \int_{\frac{1}{rs}}^{1-\frac{1}{rs}} l(z)\,dz.$$

As the underlying distribution we take the logistic distribution from Section 2.3.2. The graphs of the different angular densities in Example 2.3.10 already suggest that this rate of convergence is for higher λ faster than for lower λ. The following plots show numerical results obtained with `NIntegrate` of the software package MATHEMATICA for $\chi(r, 0.1)$ with $\lambda = 6$ and $\lambda = 1.2$, where the graphs tend to 2 for $r \to \infty$ in both cases.



One clearly sees that the rate of convergence for lower λ, thus close to the independence case, is considerably slower than for high λ, close to the dependence case.

◊

We now state the first main result of this section, with the help of which the angular density is to be estimated later.

Theorem 5.1.6
Let the random vector X follow a generalized Pareto distribution W with angular density l and $d^* > 0$. Then there exists $s_0 > 0$, such that for all $0 < s < s_0$ and all $r > 0$ as well as for any

Borel set $B \subseteq R_{d-1}$ the equation

$$P\left(Z \in B | C = -r, Z \in Q_{r,s}\right) = \frac{1}{\chi(r,s)} \int_{B \cap Q_{r,s}} l(z)\, dz$$

holds with Z and C being the random Pickands coordinates with regard to Fréchet margins. In case $Q_{r,s} = \emptyset$ both sides are set to 0. We also have for any Borel set $B \subseteq R_{d-1}$ the limit

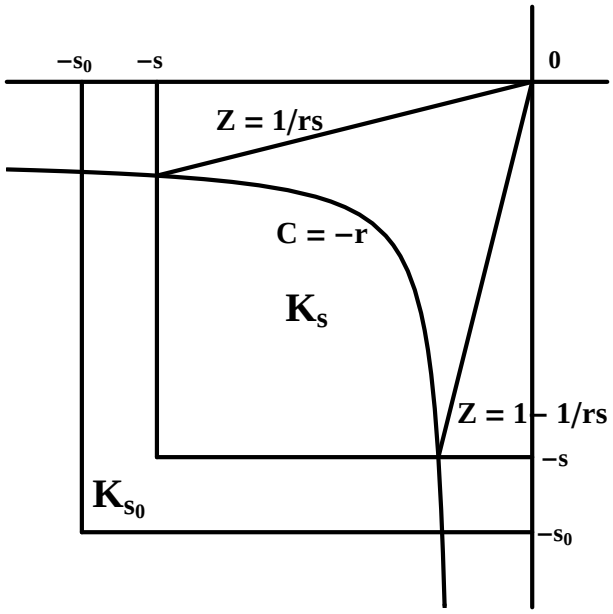
$$\lim_{r \to \infty} P\left(Z \in B | C = -r, Z \in Q_{r,s}\right) = \int_B \frac{l(z)}{d^*}\, dz.$$

Proof:
There exists an $s_0 > 0$ such that W has on K_{s_0} the representation

$$W(x) = 1 + \left(\sum_{i=1}^{d} x_i\right) D\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, \ldots, \frac{x_{d-1}}{\sum_{i=1}^{d} x_i}\right)$$

and, thus, the transformation to Pickands coordinates with regard to Fréchet margins has the density $f(z,c) = c^{-2} l(z)$ according to Lemma 5.1.1.

Consider $0 < s < s_0$ and $r > 0$ such that $Q_{r,s} \neq \emptyset$, else we have nothing to show. The situation can be illustrated with the following picture for the bivariate case, where $Q_{r,s} = \left(\frac{1}{rs}, 1 - \frac{1}{rs}\right)$.



Let $B \subseteq R_{d-1}$ be a Borel set. Then we have, see for example Gänssler and Stute (1977, [30]), Example 5.2.24,

$$\begin{aligned}
P\left(Z \in B | C = -r, Z \in Q_{r,s}\right) &= \lim_{\varepsilon \to 0} P\left(Z \in B | C \in (-r-\varepsilon, -r+\varepsilon), Z \in Q_{r,s}\right) \\
&= \lim_{\varepsilon \to 0} \frac{P\left(Z \in B \cap Q_{r,s}, C \in (-r-\varepsilon, -r+\varepsilon)\right)}{P\left(C \in (-r-\varepsilon, -r+\varepsilon), Z \in Q_{r,s}\right)} \\
&= \lim_{\varepsilon \to 0} \frac{\int_{B \cap Q_{r,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2} l(z)\, dc\, dz}{\int_{Q_{r,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2} l(z)\, dc\, dz}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\int_{B\cap Q_{r,s}} l(z)\,dz}{\int_{Q_{r,s}} l(z)\,dz} \lim_{\varepsilon\to 0} \frac{\int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\,dc}{\int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\,dc} \\
&= \frac{1}{\chi(r,s)} \int_{B\cap Q_{r,s}} l(z)\,dz,
\end{aligned}
$$

and, thus, by the dominated convergence theorem as in the proof of Lemma 5.1.4 for all $B \subseteq R_{d-1}$

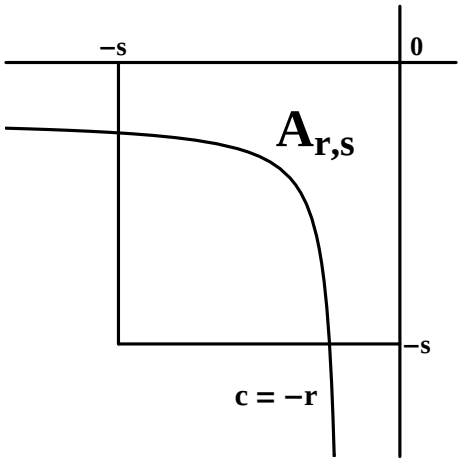
$$
\lim_{r\to\infty} P\left(Z \in B | C = -r, Z \in Q_{r,s}\right) = \lim_{r\to\infty} \frac{1}{\chi(r,s)} \int_{B\cap Q_{r,s}} l(z)\,dz = \frac{1}{d^*} \int_B l(z)\,dz = \int_B \frac{l(z)}{d^*}\,dz.
$$

□

We want to investigate the asymptotic shown in Theorem 5.1.6 a little closer. Therefore, we define the set

$$
A_{r,s} := \left\{ x = (x_1, \ldots, x_d) \in K_s \,\middle|\, c = \frac{1}{x_1} + \ldots + \frac{1}{x_d} < -r \right\}. \tag{5.5}
$$

In the bivariate case the situation can be illustrated as follows:



In the sequel we will be only interested in the observations, which fall into the set $A_{r,s}$. Therefore, the following specification of Theorem 5.1.6, which is another main result of this manuscript.

Theorem 5.1.7
Let the random vector X follow a generalized Pareto distribution W. Then

$$
\sup_{B\in\mathbb{B}_{d-1}\cap R_{d-1}} \left| P\left(Z \in B | X \in A_{r,s}\right) - \int_B \frac{l(z)}{d^*}\,dz \right| = O\left(d^* - \chi(r,s)\right)
$$

holds, where the $(d-1)$–dimensional Borel sets are noted with $\mathbb{B}_{d-1}$. I.e., the Pickands coordinate with regard to Fréchet margins Z has asymptotically (uniform) for $r \to \infty$ the conditional density $\frac{l(z)}{d^*}$.

Proof:
We define the conditional probability measure

$$
P_{r,s}(Z \in B) := P\left(Z \in B | X \in A_{r,s}\right)
$$

and note with $\mathcal{C} := P_{r,s} * C$ the distribution of the random variable C under the condition $X \in A_{r,s}$. Since $C < -r$ follows, we have

$$\int_{-\infty}^{-r} 1\,\mathcal{C}(dc) = \mathcal{C}((-\infty, -r)) = 1. \tag{5.6}$$

Then again with Example 5.2.24 in Gänssler and Stute (1977, [30]) as well as the representation of a distribution by its conditional distribution (see for example Chapter V, §2 in Rényi (1966, [58])) we get

$$\begin{aligned}
P\left(Z \in B | X \in A_{r,s}\right) &= P_{r,s}(Z \in B) \\
&= \int_{-\infty}^{-r} P_{r,s}(Z \in B | C = c)(P_{r,s} * C)(dc) \\
&= \int_{-\infty}^{-r} \lim_{\varepsilon \to 0} P_{r,s}(Z \in B | C \in (c-\varepsilon, c+\varepsilon))\,\mathcal{C}(dc) \\
&= \int_{-\infty}^{-r} \lim_{\varepsilon \to 0} \frac{P_{r,s}(Z \in B, C \in (c-\varepsilon, c+\varepsilon))}{P_{r,s}(C \in (c-\varepsilon, c+\varepsilon))}\,\mathcal{C}(dc) \\
&= \int_{-\infty}^{-r} \lim_{\varepsilon \to 0} \frac{P(Z \in B, C \in (c-\varepsilon, c+\varepsilon) | X \in A_{r,s})}{P(C \in (c-\varepsilon, c+\varepsilon) | X \in A_{r,s})}\,\mathcal{C}(dc) \\
&= \int_{-\infty}^{-r} \lim_{\varepsilon \to 0} \frac{P(Z \in B, C \in (c-\varepsilon, c+\varepsilon), X \in A_{r,s})}{P(C \in (c-\varepsilon, c+\varepsilon), X \in A_{r,s})} \cdot \frac{P(X \in A_{r,s})}{P(X \in A_{r,s})}\,\mathcal{C}(dc) \\
&= \int_{-\infty}^{-r} \lim_{\varepsilon \to 0} P(Z \in B | C \in (c-\varepsilon, c+\varepsilon), X \in A_{r,s})\,\mathcal{C}(dc) \\
&= \int_{-\infty}^{-r} P(Z \in B | C = c, X \in A_{r,s})\,\mathcal{C}(dc) \\
&= \int_{-\infty}^{-r} P(Z \in B | C = c, Z \in Q_{-c,s})\,\mathcal{C}(dc).
\end{aligned}$$

The final equality follows from the fact that, according to the definition (5.2), of $Q_{r,s}$ the conditions

$$C = c, X \in A_{r,s} \Longleftrightarrow C = c, Z \in Q_{-c,s}$$

are equivalent. With the help of Theorem 5.1.6 and the above equality it follows that

$$\begin{aligned}
&\left| P\left(Z \in B | X \in A_{r,s}\right) - \int_B \frac{l(z)}{d^*}\,dz \right| \\
&= \left| \int_{-\infty}^{-r} P\left(Z \in B | C = c, Z \in Q_{-c,s}\right)\,\mathcal{C}(dc) - \int_B \frac{l(z)}{d^*}\,dz \right| \\
&= \left| \int_{-\infty}^{-r} \int_{B \cap Q_{-c,s}} \frac{l(z)}{\chi(-c,s)}\,dz\,\mathcal{C}(dc) - \int_B \frac{l(z)}{d^*}\,dz \right| \\
&\leq \left| \int_{-\infty}^{-r} \int_B \frac{l(z)}{\chi(-c,s)}\,dz\,\mathcal{C}(dc) - \int_B \frac{l(z)}{d^*}\,dz \right| \\
&\quad + \left| \int_{-\infty}^{-r} \int_{B \cap Q_{-c,s}} \frac{l(z)}{\chi(-c,s)}\,dz\,\mathcal{C}(dc) - \int_{-\infty}^{-r} \int_B \frac{l(z)}{\chi(-c,s)}\,dz\,\mathcal{C}(dc) \right| \\
&= \underbrace{\left(\int_B \frac{l(z)}{d^*}\,dz \right)}_{\leq 1} \left| \int_{-\infty}^{-r} \frac{d^*}{\chi(-c,s)}\,\mathcal{C}(dc) - 1 \right| + \int_{-\infty}^{-r} \int_{B \setminus Q_{-c,s}} \frac{l(z)}{\chi(-c,s)}\,dz\,\mathcal{C}(dc)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(5.6)}{\leq} \left| \int_{-\infty}^{-r} \frac{d^*}{\chi(-c,s)} - 1 \, \mathcal{C}(dc) \right| + \int_{-\infty}^{-r} \int_{R_{d-1} \setminus Q_{r,s}} \frac{l(z)}{\chi(r,s)} \, dz \, \mathcal{C}(dc) \\
&= \int_{-\infty}^{-r} \frac{d^* - \chi(-c,s)}{\chi(-c,s)} \, \mathcal{C}(dc) + \int_{R_{d-1} \setminus Q_{r,s}} \frac{l(z)}{\chi(r,s)} \, dz \int_{-\infty}^{-r} 1 \, \mathcal{C}(dc) \\
&\overset{(5.6)}{\leq} \int_{-\infty}^{-r} \frac{d^* - \chi(r,s)}{\chi(r,s)} \, \mathcal{C}(dc) + \frac{1}{\chi(r,s)} \left(\int_{R_{d-1}} l(z) \, dz - \int_{Q_{r,s}} l(z) \, dz \right) \\
&= \frac{d^* - \chi(r,s)}{\chi(r,s)} \int_{-\infty}^{-r} 1 \, \mathcal{C}(dc) + \frac{d^* - \chi(r,s)}{\chi(r,s)} \overset{(5.6)}{=} O(d^* - \chi(r,s)).
\end{aligned}
$$

The right hand side does not depend on B and, thus, we have the uniform convergence

$$
\sup_{B \in \mathbb{B}_{d-1} \cap R_{d-1}} \left| P\left(Z \in B | X \in A_{r,s}\right) - \int_B \frac{l(z)}{d^*} \, dz \right| = O\left(d^* - \chi(r,s)\right).
$$

□

5.2 Nonparametric Estimation of the Angular Density for Generalized Pareto Distributions

The goal of this section is to give an estimation procedure for the angular density with the help of the results found in Section 5.1. We will proceed analogously to Section 4.1 and use again kernel density estimators but this time for arbitrary dimension $d \geq 2$.

Assume that we have n independent copies $\tilde{X}^{(1)}, \ldots, \tilde{X}^{(n)}$ of a random vector X, which follows a generalized Pareto distribution according to representation (2.5) with $U = K_s$. Denote by $\tilde{Z}^{(i)}$ and $\tilde{C}^{(i)}$ the corresponding Pickands coordinates with regard to Fréchet margins, $i = 1, \ldots, n$. Choose a large threshold $r > 0$, and consider only those observations $\tilde{X}^{(i)}$ with $\tilde{X}^{(i)} \in A_{r,s}$, i.e., $\tilde{X}^{(i)} \in K_s$ and $\tilde{C}^{(i)} < -r$. We denote these by $X^{(1)}, \ldots, X^{(m)}$. They are independent of the binomially distributed random variable $m = \tau_n$, independent from each other and identically distributed, see Theorem 1.4.1 in Reiss (1993, [56]). According to Theorem 5.1.7 the $Z^{(i)}$ have the density $l(z)/d^*$ for $r \to \infty$.

A natural estimator for l is, thus, a kernel density estimator with kernel k, bandwidth $h > 0$ and data sphering

$$
\hat{l}_{m,r}(z) = d \cdot \frac{1}{(\det S_m)^{1/2} \, m h^{d-1}} \sum_{i=1}^{m} k\left(\frac{S_m^{-1/2} \left(z - Z^{(i)}\right)}{h} \right). \tag{5.7}
$$

Thereby S_m denotes the covariance matrix of the $Z^{(i)}$ and $S_m^{-1/2}$ its inverse symmetric root. For a practical introduction on multivariate kernel density estimators we again refer to Falk et al. (2002, [23]), Chapter 6 or Chapter 4 of Simonoff (1996, [67]). A more theoretical introduction can be found in Chapters 2 and 3 of Prakasa Rao (1983, [55]) or Chapter 4 of Wertz (1978, [74]).

Data sphering is a concept from multivariate analysis, where the data are first multiplied by the inverse symmetric root of their covariance matrix, then the density is estimated, and afterwards transformed back, thus using the so called Mahalanobis distance. In the univariate case, this reduces to dividing the data by their empirical standard deviation. For more information on data sphering, see for example Falk et al. (2002, [23]), Chapter 6.

The factor d in the estimator (5.7) is included to get an asymptotically unbiased estimator of $l(z)$ in the case $d^* = d$. To get an unbiased estimator in an arbitrary case, we would have to put d^* here. But since d^* depends on the angular density which is to be estimated, d^* is usually unknown. So in the case $d^* < d$ we can only estimate a constant multiple of $l(z)$. The asymptotic results underlying these statements will be shown in Sections 5.4 to 5.6.

Regarding the choice of the kernel, there are a number of popular variants. One is the normal kernel

$$k_N(x) = (2\pi)^{-(d-1)/2} \exp\left(-\frac{1}{2}x^T x\right),$$

another one the Epanechnikov kernel

$$k_E(x) = \max\left\{0, \frac{\left(1+\frac{d-1}{2}\right)(d-1)\Gamma\left(\frac{d-1}{2}\right)}{2\pi^{(d-1)/2}}\left(1-x^T x\right)\right\},$$

where Γ denotes the Gamma function, see also Remark 2.3.9. As experience shows, the choice of the kernel does not play an overly important role in practice, see, for example, Chapter 4 in Wertz (1978, [74]). Thus, also other kernels can be used, see, for example, Section 6.4 in Falk et al. (2002, [23])) for further possibilities.

The asymptotic theory in Sections 5.4 to 5.6 will be developed for kernels with bounded support, which also fulfill conditions, which are stated in (5.17). The Epanechnikov kernel fulfills all these conditions, the normal kernel does this with one exception, the bounded support. Thus the Epanechnikov kernel seems a more natural choice. However, the normal kernel has one advantage in practical applications, and that is a simple choice of the bandwidth. Simonoff (1996, [67]) gives in Section 4 a generalization of the optimal bandwidth (4.2) in the general multivariate case for normally distributed data and the normal kernel. With our notations this bandwidth reads

$$h = \left(\frac{4}{m(d+1)}\right)^{1/(d+3)}. \tag{5.8}$$

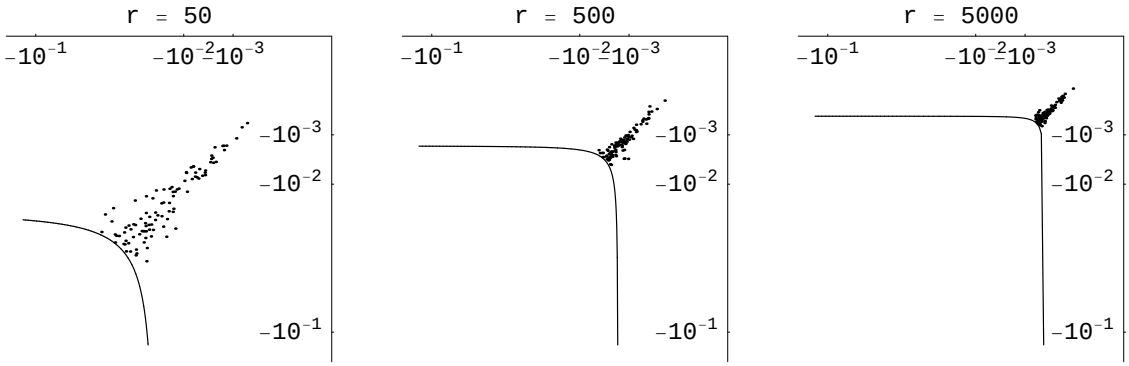
Of course we do not have normal data here, in contrast to the Pickands densities in Section 4.1.2, the angular densities of logistic type close to the independence case do not resemble the normal distribution. Nevertheless, simulations show that with the bandwidth given above, quite good results can be achieved and that this is a reasonable approach.

Due to this simple automatic choice of a bandwidth, we restrict ourselves, thus, in the course of this manuscript for practical purposes to the normal kernel and to (5.8). Of course there are again advanced methods for the choice of h, see the references in Section 4.1.2, which should be used for other choices of kernels such as the Epanechnikov kernel, but which are much harder to implement.

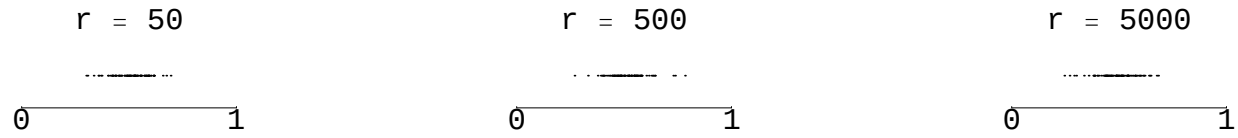
Since we have a bounded support of the angular density, one should use reflection techniques, see Sections 2.1 and 8.2 of Reiss and Thomas (2001, [57]), to further improve the estimation. Thereby the data set is reflected at the boundaries of the support. In our case these were the $(d-2)$-dimensional coordinate planes $x_i = 0$, $i = 1, \ldots, d-1$ and the plane $\sum_{i=1}^{d-1} x_i = 1$. Then the density estimator for the enlarged data set is computed. This especially improves the behavior of the kernel density estimator at the boundary of the support, and was used in all following estimations.

Example 5.2.1
We want to evaluate the estimator $\hat{l}_{m,r}(z)$ for the angular density found in (5.7) with simulated data. The logistic random vectors underlying the estimator, which are used in the following,

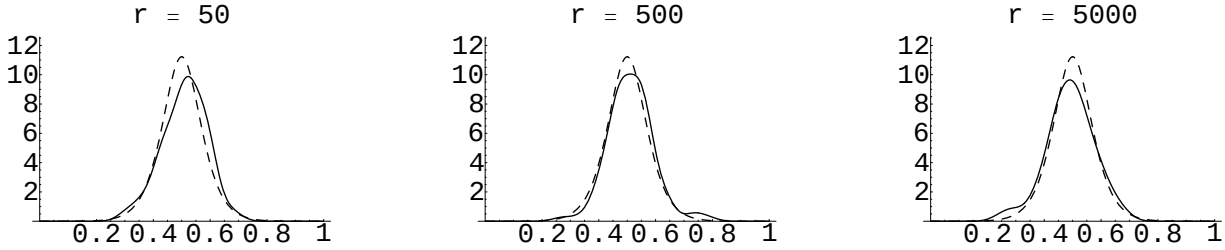
were all generated by Algorithm 3.1.11. Remark that in the logistic case $d^* = d$ holds for $\lambda > 1$. We begin with $\lambda = 6$ and $d = 2$. We choose $s = 0.1$, and in each case 100 data points $X^{(i)}$ are generated, which fulfill $C^{(i)} < -r$ for $r = 50$, 500, 5000 and $||X^{(i)}||_\infty < s$, i.e., $X^{(i)} \in A_{r,s}$. The following plots show the resulting data together with the corresponding threshold line. For illustration purposes a logarithmic scale is used.



Subsequently the angular components $Z^{(i)}$ are computed and can be seen in the following graphics.

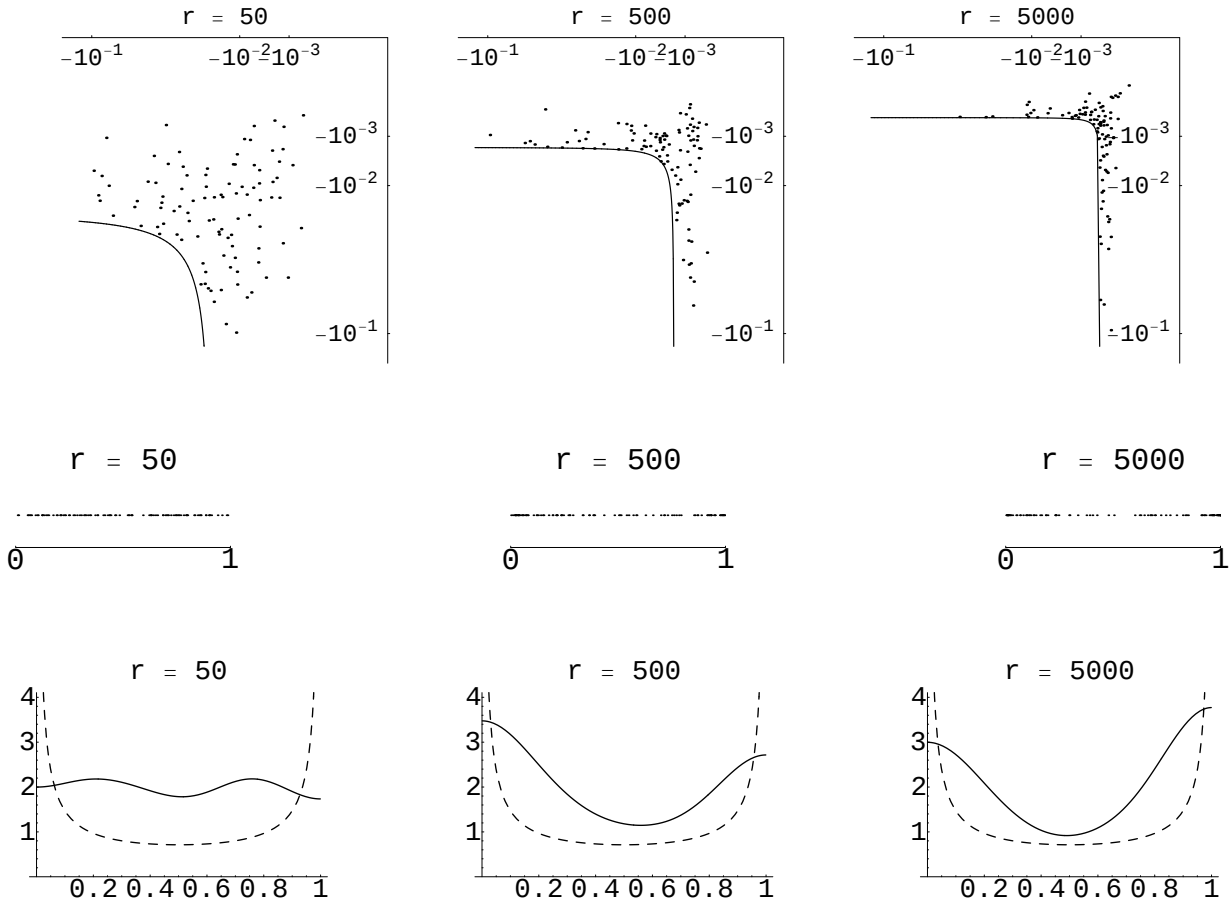


One now estimates the angular density by using a kernel density estimator for the $Z^{(i)}$. The dashed line represents the underlying angular density (see Example 2.3.10).



Exactly the same graphics were also created for the choice of $\lambda = 1.2$. The other parameters

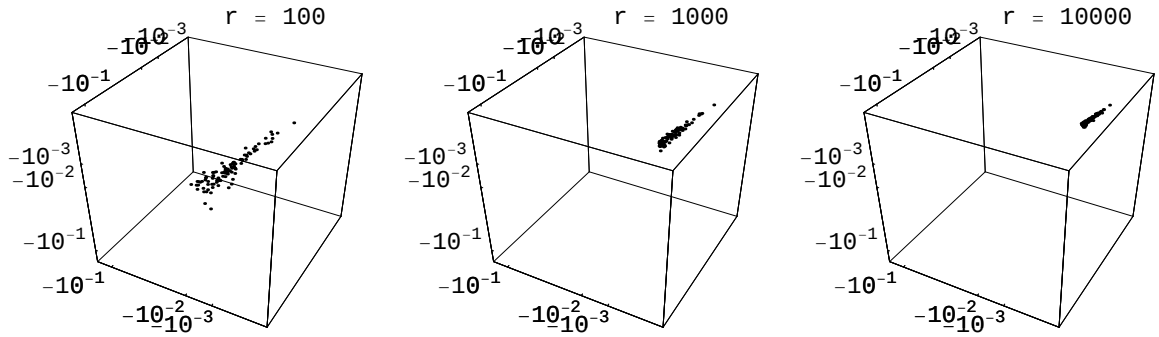
were thereby left unchanged.



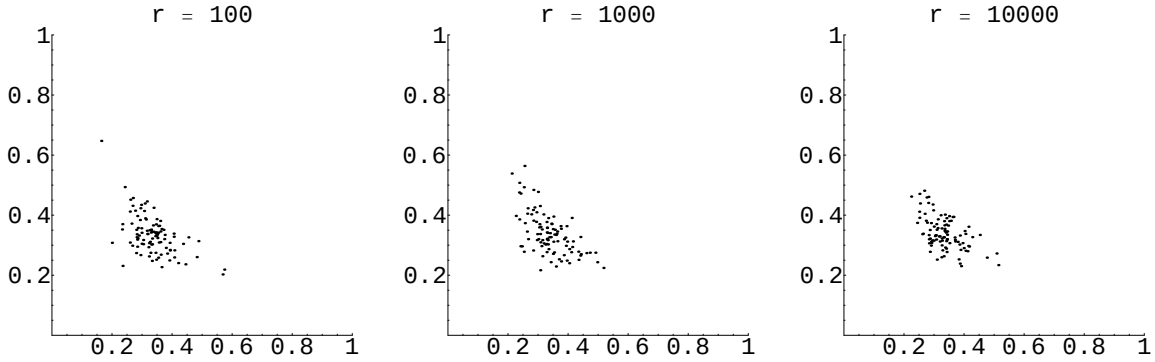
In the case of a high degree of dependence ($\lambda = 6$), one gets always good results, independent of the choices of r. However, in the case of a low degree of dependence ($\lambda = 1.2$), a clear difference in the pictures is visible for different r. The rate of convergence underlying the distribution of the used data depends, according to Theorem 5.1.7, on how fast $\chi(r, s)$ converges to d^*. This rate is slower for lower λ according to Example 5.1.5. Exact mathematical results of the influence of $\chi(r, s)$ on the convergence will be shown in Sections 5.4 to 5.6. That such nonparametric estimators depend heavily on the threshold and have difficulties close to the independence case is a common phenomenon in extreme value analysis, see for example Section 4.4 of Coles et al. (1999, [9]). A proposal for the practical choice of r with real data can be found in Section 5 in Joe et al. (1992, [43]) and in Section 6 in Coles and Tawn (1994, [11]), where similar estimators for the extreme value case are given.

The three dimensional case can also be visualized quite well. We begin again with $\lambda = 6$. For the values of $r = 100$, 1000, 10000 and $s = 0.1$, the following 100 data were generated and again

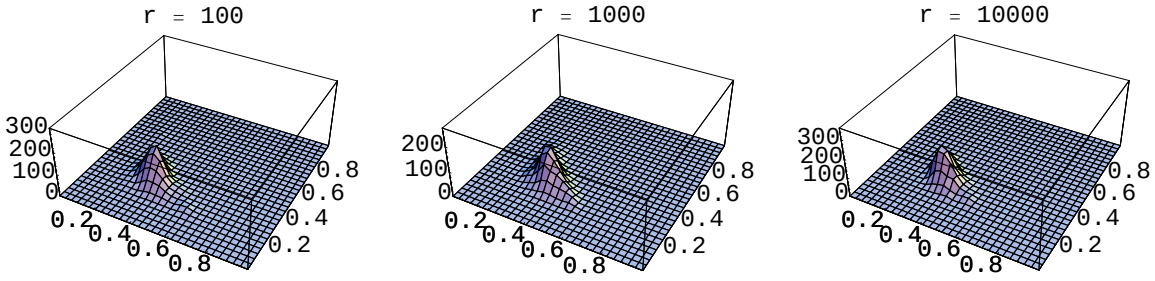
plotted on a logarithmic scale.



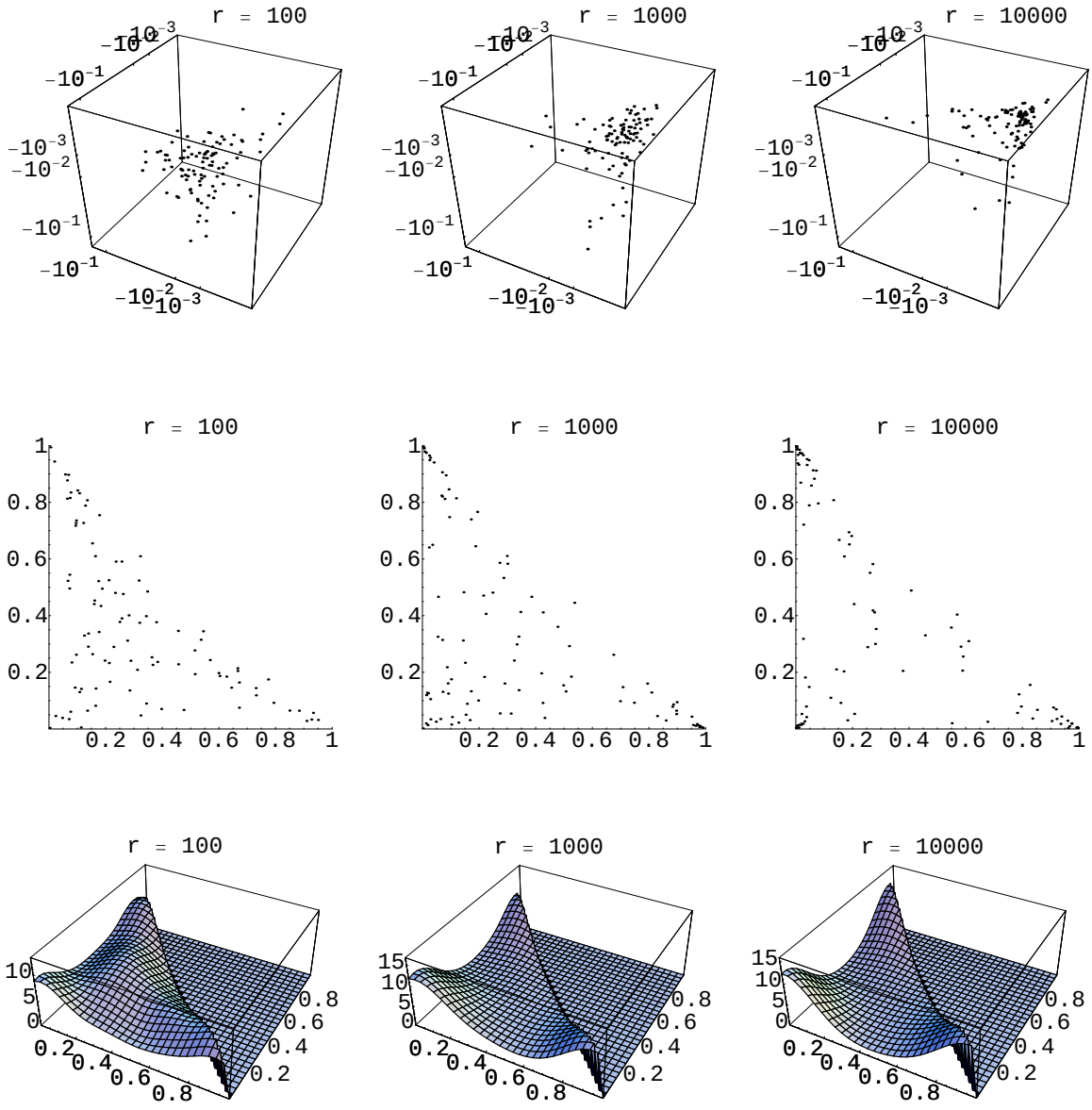
We compute the angular components of these points



and are again able to estimate the angular density. A picture of the underlying angular density can be found in Example 2.3.10.



One proceeds in exactly the same way for $\lambda = 1.2$.



The behavior of the estimator is roughly the same as in the bivariate case. One can see again in the case close to the independence that the angular components accumulate only for high r in the vertices of R_2.

For dimensions $d \geq 4$ a complete graphical representation is not possible. The data and the angular components can be visualized, for example, by a scatterplot matrix, see Example 3.1.12. Also the angular density and its estimator are only partially presentable. To investigate the tail dependence structure, the behavior of $l(z)$ and $\hat{l}_{m,r}(z)$ respectively is of interest especially in the vertices of R_{d-1}, see Example 2.3.10. So one should try to visualize these areas.

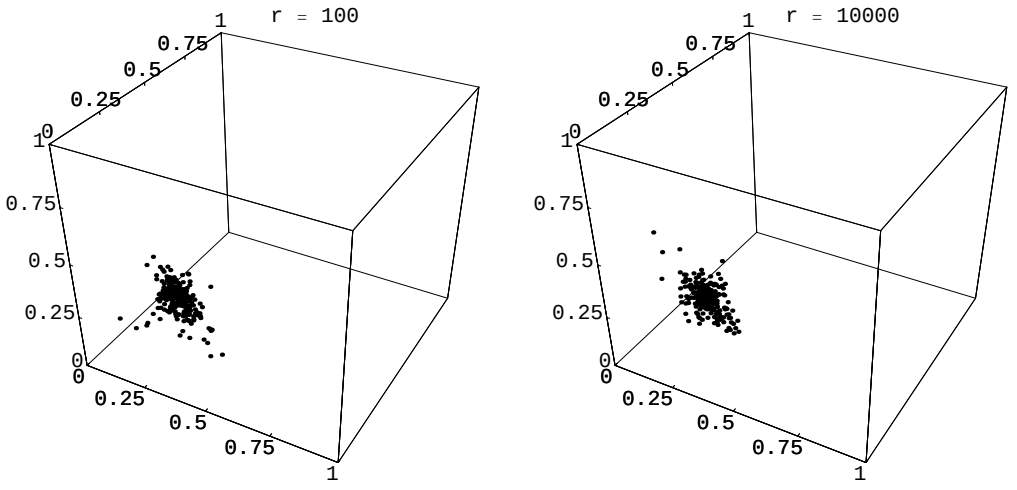
Let e_i, $i = 0, \ldots, d-1$ be the vertices of the unit simplex R_{d-1}, as defined at the beginning of

Section 2.3.1. We consider the lines

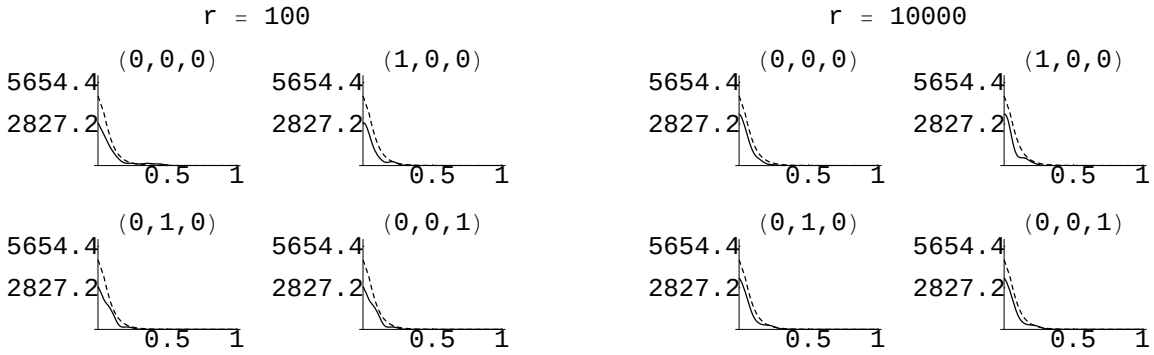
$$g_i(t) := \begin{pmatrix} 1/d \\ \vdots \\ 1/d \end{pmatrix} + t \left(e_i - \begin{pmatrix} 1/d \\ \vdots \\ 1/d \end{pmatrix} \right), \qquad t \in [0,1), \quad i = 0, \ldots, d-1.$$

They originate at $(1/d, \ldots, 1/d)$ and end at the vertices e_i. Through the corresponding graph of $l\left(g_i(t)\right)$ and $\hat{l}_{m,r}\left(g_i(t)\right)$ respectively, one can draw conclusions about the behavior of the estimator at the boundary. We want to show this for the case $d = 4$.

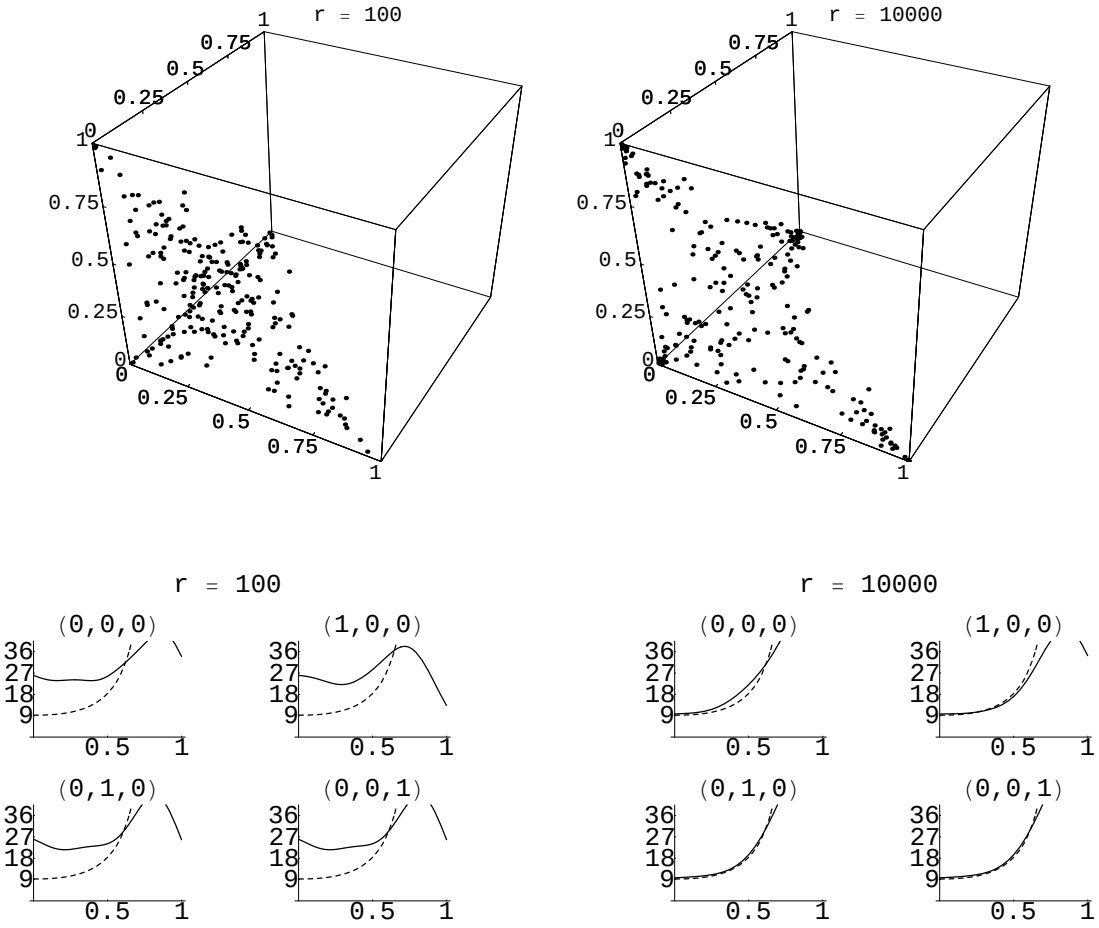
s was again chosen as 0.1, r was set to 100 and 10000 respectively and 250 points were generated, which exceeded the corresponding thresholds. Beginning with $\lambda = 6$, we plot the angular components



and subsequently parts of the angular density (dashed) together with its estimator as described above.



Exactly the same parameters were also used in the case $\lambda = 1.2$.



One sees again the strong dependence of the estimator on r, close to the independence case.

To use kernel density estimators in dimensions higher than 4 or 5 is practically not very reasonable. Because of the so-called "Problem of Empty Space" an unrealistically high number of data is necessary. Further information on problems with kernel density estimators in high dimensions can be found in Section 4 of Simonoff (1996, [67]).

$\Diamond$

5.3 Differentiable δ–Neighborhoods of Pickands Transforms with Regard to Fréchet Margins

For practical applications, the assumption of random vectors following exactly a GPD might not be the most realistic one. By the POT approach, one more often encounters random vectors whose distribution function is, in a certain sense, close to a GPD. In the univariate case this leads to the definition of δ-neighborhoods, see Section 2.2 of Falk et al. (2004, [21]). The problem of defining appropriate multivariate versions of δ-neighborhoods is dealt with in Kaufmann and Reiss (1995, [45]) and Section 5.3 of Falk et al. (2004, [21]). In Section 5.4 of Falk et al. (2004, [21]), the concept of differentiable δ-neighborhoods with the help of the standard

Pickands transformation is introduced. This is also a possible way to define a multivariate version of δ-neighborhoods. It is used to extend certain results shown for GPDs to a broader class of distribution functions. In this section, we want to give an analogue of these differentiable δ-neighborhoods for the case of the Pickands transform with regard to Fréchet margins. It helps us to find a wider class of distribution functions, to which the estimator defined in the previous section is applicable.

Definition 5.3.1
Let $X = (X_1, \ldots, X_d) \in (-\infty, 0)^d$ be a random vector with a distribution function H. For $\delta > 0$ the distribution function H is said to be in the *differentiable δ–neighborhood (with regard to Fréchet margins)* of the generalized Pareto distribution W with angular density l, if a neighborhood U of 0 in the relative topology of the negative quadrant $(-\infty, 0)^d$ exists, such that the Pickands transformation with regard to Fréchet margins $T_F(X) = (Z, C)$ has on $T_F(U)$ a density $h(z, c)$ with the uniform representation

$$h(z,c) = c^{-2}\left(l(z) + q(z)O\left(\left|\frac{1}{c}\right|^{\delta}\right)\right)$$

for $c \to -\infty$, where $q(z) \geq 0$. Thereby we demand that a number $s > 0$ exists, such that for all $v > 0$ the integral

$$\alpha(v,s) := \int_{Q_{v,s}} q(z)\, dz$$

exists and is finite.

◊

Remark 5.3.2
In contrast to the limit of $\chi(r, s)$ the limit $\lim_{v\to\infty} \alpha(v, s)$ may not exist. For the targeted estimation procedure to work, some conditions on the rate of convergence of $v \to \infty$ must be met. Then we can prove extensions of Theorems 5.1.6 and 5.1.7 for distribution functions from the δ–neighborhood of generalized Pareto distributions.

◊

Example 5.3.3
In this example we want to show that an extreme value distribution falls into the 1-neighborhood of the corresponding GPD. An extreme value distribution is defined by

$$G(x_1, \ldots, x_d) = \exp\left(\left(\sum_{i=1}^{d} x_i\right) D\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, \ldots, \frac{x_{d-1}}{\sum_{i=1}^{d} x_i}\right)\right),$$

see Definition 2.2.1. As was shown in Falk and Reiss (2005, [27]) after the proof of Theorem 5.3, we have

$$\frac{\partial^d}{\partial x_1 \cdots \partial x_d} G(x_1, \ldots, x_d) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W(x_1, \ldots, x_d) + O\left(\frac{1}{|x_1 + \ldots + x_d|^{d-2}}\right)$$

close to 0.

Then we have, by inserting the inverse Pickands transformation with regard to Fréchet margins (2.3),

$$\frac{\partial^d}{\partial x_1 \cdots \partial x_d} G\left(T_F^{-1}(z, c)\right)$$

$$
\begin{aligned}
&= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z,c)\right) + O\left(\frac{1}{\frac{1}{|c|^{d-2}}\left(\frac{1}{z_1}+\ldots+\frac{1}{z_{d-1}}+\frac{1}{1-\sum_{i=1}^d z_i}\right)^{d-2}}\right) \\
&= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z,c)\right) + O\left(|c|^{d-2}\right),
\end{aligned}
$$

since the function

$$
(z_1,\ldots,z_{d-1}) \to \frac{1}{\left(\frac{1}{z_1}+\ldots+\frac{1}{z_{d-1}}+\frac{1}{1-\sum_{i=1}^d z_i}\right)^{d-2}}
$$

is bounded on R_{d-1}. Thus, we have for the density g of G in terms of the Pickands coordinates with regard to Fréchet margins by Lemma 5.1.1, the density transformation theorem and Lemma 2.1.5

$$
\begin{aligned}
g(z,c) &= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} G\left(T_F^{-1}(z,c)\right) |\det J_{T_F^{-1}}(z,c)| \\
&= \left(\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\left(T_F^{-1}(z,c)\right) + O\left(|c|^{d-2}\right)\right) |\det J_{T_F^{-1}}(z,c)| \\
&= f(z,c) + |\det J_{T_F^{-1}}(z,c)| O\left(|c|^{d-2}\right) \\
&= c^{-2} l(z) + \frac{1}{|c|^{d+1}} \cdot \underbrace{\frac{1}{z_1^2 \cdot z_2^2 \cdot \ldots \cdot z_{d-1}^2 \left(1-\sum_{i=1}^{d-1} z_i\right)^2}}_{:=q(z)} O\left(|c|^{d-2}\right) \\
&= c^{-2} l(z) + q(z) O\left(|c|^{-3}\right) \\
&= c^{-2}\left(l(z) + q(z) O\left(\left|\frac{1}{c}\right|\right)\right).
\end{aligned}
$$

Since $q(z) \geq 0$ is continuous on R_{d-1}, the function is integrable for every $v > 0$ on $Q_{v,s}$. Thus the extreme value distribution G lies in the 1–neighborhood of the generalized Pareto distribution W. Note, however, that

$$
\lim_{v\to\infty} \alpha(v,s) = \lim_{v\to\infty} \int_{Q_{v,s}} q(z)\, dz = \infty
$$

for the special choice of

$$
q(z) = \frac{1}{z_1^2 \cdot z_2^2 \cdot \ldots \cdot z_{d-1}^2 \left(1-\sum_{i=1}^{d-1} z_i\right)^2}.
$$

$\Diamond$

We can now show analogues of Theorems 5.1.6 and 5.1.7 for distribution functions in δ-neighborhoods of GPDs.

Theorem 5.3.4
Let the random vector $X \in (-\infty,0)^d$ follow a distribution function H, which lies in the δ–neighborhood of a generalized Pareto distribution W with angular density $l(z)$ and $d^* > 0$.

Then there exists $s_0 > 0$ such that for all $0 < s < s_0$ and all $0 < v \leq r$ and any Borel set $B \subseteq R_{d-1}$ the equation

$$P\left(Z \in B | C = -r, Z \in Q_{v,s}\right) = \frac{1}{\chi(v,s)} \int_{B \cap Q_{v,s}} l(z)\, dz + O\left(\alpha(v,s) r^{-\delta-1}\right)$$

holds. If

$$r^{-\delta-1}\alpha(v,s) \to 0 \tag{5.9}$$

for $r \to \infty$ and $v \to \infty$, then

$$\lim_{r\to\infty, v\to\infty} P\left(Z \in B | C = -r, Z \in Q_{v,s}\right) = \int_B \frac{l(z)}{d^*}\, dz.$$

Proof:
Before we begin with the actual proof, we need an auxiliary result. First we have for $\delta, \varepsilon > 0$ and $r > 0$ with a proper constant $M > 0$

$$\begin{aligned}
&\left| \int_{-r-\varepsilon}^{-r+\varepsilon} O\left(|c|^{-\delta-2}\right)\, dc \right| \leq \int_{-r-\varepsilon}^{-r+\varepsilon} M\,|c|^{-\delta-2}\, dc = M\left[\frac{|c|^{-\delta-1}}{\delta+1}\right]_{-r-\varepsilon}^{-r+\varepsilon} \\
&= \frac{M}{\delta+1}\left(|-r+\varepsilon|^{-\delta-1} - |-r-\varepsilon|^{-\delta-1}\right) = O\left(r^{-\delta-1}\right).
\end{aligned} \tag{5.10}$$

Now we prove the main result. By assumption an $s_0 > 0$ exists, such that H possesses on K_{s_0} in the Pickands coordinates with regard to Fréchet margins the density

$$h(z,c) = c^{-2} l(z) + q(z) O\left(|c|^{-\delta-2}\right).$$

Consider $0 < s < s_0$. Let $B \subseteq R_{d-1}$. We get as in Theorem 5.1.6 and with the calculation rules of the O-notation

$$\begin{aligned}
&P\left(Z \in B | C = -r, Z \in Q_{v,s}\right) \\
&= \lim_{\varepsilon\to 0} P\left(Z \in B | C \in (-r-\varepsilon, -r+\varepsilon), Z \in Q_{v,s}\right) \\
&= \lim_{\varepsilon\to 0} \frac{P\left(Z \in B \cap Q_{v,s}, C \in (-r-\varepsilon, -r+\varepsilon)\right)}{P\left(C \in (-r-\varepsilon, -r+\varepsilon), Z \in Q_{v,s}\right)} \\
&= \lim_{\varepsilon\to 0} \frac{\int_{B\cap Q_{v,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\left(l(z) + q(z) O\left(|c|^{-\delta}\right)\right)\, dc\, dz}{\int_{Q_{v,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\left(l(z) + q(z) O\left(|c|^{-\delta}\right)\right)\, dc\, dz} \\
&= \lim_{\varepsilon\to 0} \frac{\int_{B\cap Q_{v,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2} l(z)\, dc\, dz + \int_{B\cap Q_{v,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} q(z) O\left(|c|^{-\delta-2}\right)\, dc\, dz}{\int_{Q_{v,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2} l(z)\, dc\, dz + \int_{Q_{v,s}} \int_{-r-\varepsilon}^{-r+\varepsilon} q(z) O\left(|c|^{-\delta-2}\right)\, dc\, dz} \\
&= \lim_{\varepsilon\to 0} \frac{\int_{B\cap Q_{v,s}} l(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\, dc + \int_{B\cap Q_{v,s}} q(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} O\left(|c|^{-\delta-2}\right)\, dc}{\int_{Q_{v,s}} l(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\, dc + \int_{Q_{v,s}} q(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} O\left(|c|^{-\delta-2}\right)\, dc} \\
&\overset{(5.10)}{=} \lim_{\varepsilon\to 0} \frac{\int_{B\cap Q_{v,s}} l(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\, dc + \int_{B\cap Q_{v,s}} q(z)\, dz \cdot O\left(r^{-\delta-1}\right)}{\int_{Q_{v,s}} l(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\, dc + \alpha(v,s) \cdot O\left(r^{-\delta-1}\right)} \\
&= \lim_{\varepsilon\to 0} \frac{\int_{B\cap Q_{v,s}} l(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\, dc}{\int_{Q_{v,s}} l(z)\, dz \int_{-r-\varepsilon}^{-r+\varepsilon} c^{-2}\, dc} + O\left(\alpha(v,s) r^{-\delta-1}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\int_{B\cap Q_{v,s}} l(z)\,dz}{\int_{Q_{v,s}} l(z)\,dz} + O\left(\alpha(v,s)r^{-\delta-1}\right) \\
&= \frac{1}{\chi(v,s)} \int_{B\cap Q_{v,s}} l(z)\,dz + O\left(\alpha(v,s)r^{-\delta-1}\right).
\end{aligned}
$$

If (5.9) holds, we have for all Borel sets $B \subseteq R_{d-1}$ by the dominated convergence theorem (see for example Fristedt and Gray (1997, [28]), Section 8.2)

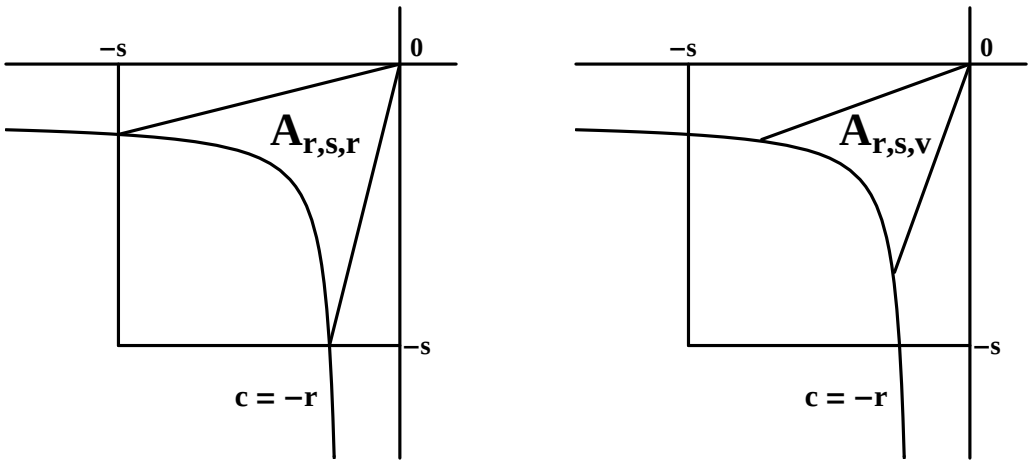
$$
\begin{aligned}
\lim_{r\to\infty, v\to\infty} P\left(Z \in B | C = -r, Z \in Q_{v,s}\right) &= \lim_{r\to\infty, v\to\infty} \frac{1}{\chi(v,s)} \int_{B\cap Q_{v,s}} l(z)\,dz + O\left(\alpha(v,s)r^{-\delta-1}\right) \\
&= \int_B \frac{l(z)}{d^*}\,dz.
\end{aligned}
$$

□

Again we want to investigate the shown asymptotic closer. For this we define for $v \le r$ the set

$$
A_{r,s,v} := \left\{ x \in K_s \,\middle|\, c = \frac{1}{x_1} + \ldots + \frac{1}{x_d} < -r, z = \left(\frac{\frac{1}{x_1}}{c}, \ldots, \frac{\frac{1}{x_{d-1}}}{c} \right) \in Q_{v,s} \right\}. \tag{5.11}
$$

In the bivariate case it can be illustrated as follows:



Due to Theorem 5.3.4 we are only interested in observations, which fall into the set $A_{r,s,v}$. Therefore, we make the following specification of Theorem 5.3.4.

Theorem 5.3.5
Let the random vector $X \in (-\infty, 0)^d$ follow a distribution function H, which lies in the δ–neighborhood of a generalized Pareto distribution W. Then

$$
\sup_{B\in\mathbb{B}_{d-1}\cap R_{d-1}} \left| P\left(Z \in B | X \in A_{r,s,v}\right) - \int_B \frac{l(z)}{d^*}\,dz \right| = O\left((d^* - \chi(v,s)) + \left(\alpha(v,s)r^{-\delta-1}\right) \right)
$$

holds for $v \le r$, where the $(d-1)$–dimensional Borel sets are again denoted by $\mathbb{B}_{d-1}$.

Proof:
We define, in analogy to the proof of Theorem 5.1.7, the conditional probability measure

$$
P_{r,s,v}(Z \in B) := P(Z \in B | X \in A_{r,s,v}),
$$

and by $\mathcal{C} := P_{r,s,v} * C$ we denote again the distribution of the random variable C, this time under the condition $X \in A_{r,s,v}$. Since $C < -r$ follows, we have

$$\int_{-\infty}^{-r} 1\, \mathcal{C}(dc) = \mathcal{C}((-\infty, -r)) = 1. \tag{5.12}$$

Analogously to the proof of Theorem 5.1.7 one shows that

$$P\left(Z \in B | X \in A_{r,s,v}\right) = \int_{-\infty}^{-r} P\left(Z \in B | C = c, Z \in Q_{v,s}\right) \mathcal{C}(dc).$$

With the help of Theorem 5.3.4 and analogously to the proof of Theorem 5.1.7, we obtain

$$\begin{aligned}
&\left| P\left(Z \in B | X \in A_{r,s,v}\right) - \int_B \frac{l(z)}{d^*}\, dz \right| \\
&= \left| \int_{-\infty}^{-r} P\left(Z \in B | C = c, Z \in Q_{v,s}\right) \mathcal{C}(dc) - \int_B \frac{l(z)}{d^*}\, dz \right| \\
&= \left| \int_{-\infty}^{-r} \frac{1}{\chi(v,s)} \int_{B \cap Q_{v,s}} l(z)\, dz + O\left(\alpha(v,s) c^{-\delta-1}\right) \mathcal{C}(dc) - \int_B \frac{l(z)}{d^*}\, dz \right| \\
&\leq \left| \int_{-\infty}^{-r} \frac{1}{\chi(v,s)} \int_{B \cap Q_{v,s}} l(z)\, dz\, \mathcal{C}(dc) - \int_B \frac{l(z)}{d^*}\, dz \right| + \left| \int_{-\infty}^{-r} O\left(\alpha(v,s) c^{-\delta-1}\right) \mathcal{C}(dc) \right| \\
&\overset{(5.12)}{\leq} \left| \frac{1}{\chi(v,s)} \int_{B \cap Q_{v,s}} l(z)\, dz - \int_B \frac{l(z)}{d^*}\, dz \right| + \int_{-\infty}^{-r} O\left(\alpha(v,s) r^{-\delta-1}\right) \mathcal{C}(dc) \\
&\leq \left| \frac{1}{\chi(v,s)} \int_{B \cap Q_{v,s}} l(z)\, dz - \frac{1}{\chi(v,s)} \int_B l(z)\, dz \right| \\
&\quad + \left| \frac{1}{\chi(v,s)} \int_B l(z)\, dz - \int_B \frac{l(z)}{d^*}\, dz \right| + O\left(\alpha(v,s) r^{-\delta-1}\right) \\
&= \int_{B \setminus Q_{v,s}} \frac{l(z)}{\chi(v,s)}\, dz + \int_B l(z)\, dz \left(\frac{1}{\chi(v,s)} - \frac{1}{d^*} \right) + O\left(\alpha(v,s) r^{-\delta-1}\right) \\
&\leq \int_{R_{d-1} \setminus Q_{v,s}} \frac{l(z)}{\chi(v,s)}\, dz + d^* \frac{d^* - \chi(v,s)}{d^* \chi(v,s)} + O\left(\alpha(v,s) r^{-\delta-1}\right) \\
&= \int_{R_{d-1}} \frac{l(z)}{\chi(v,s)}\, dz - \int_{Q_{v,s}} \frac{l(z)}{\chi(v,s)}\, dz + \frac{d^* - \chi(v,s)}{\chi(v,s)} + O\left(\alpha(v,s) r^{-\delta-1}\right) \\
&= \frac{d^*}{\chi(v,s)} - 1 + O\left(d^* - \chi(v,s)\right) + O\left(\alpha(v,s) r^{-\delta-1}\right) \\
&= O\left(\left(d^* - \chi(v,s)\right) + \alpha(v,s) r^{-\delta-1} \right).
\end{aligned}$$

The right hand side does not depend on B and, thus, uniform convergence holds

$$\sup_{B \in \mathbb{B}_{d-1} \cap R_{d-1}} \left| P\left(Z \in B | X \in A_{r,s,v}\right) - \int_B \frac{l(z)}{d^*}\, dz \right| = O\left(\left(d^* - \chi(v,s)\right) + \left(\alpha(v,s) r^{-\delta-1}\right) \right).$$

□

By Theorem 5.3.5 we know that the Pickands coordinate Z (with regard to Fréchet margins) has asymptotically (uniform) for $r \to \infty$ and $v \to \infty$ the conditional density $\frac{l(z)}{d^*}$ if the condition

$$r^{-\delta-1} \alpha(v,s) \to 0$$

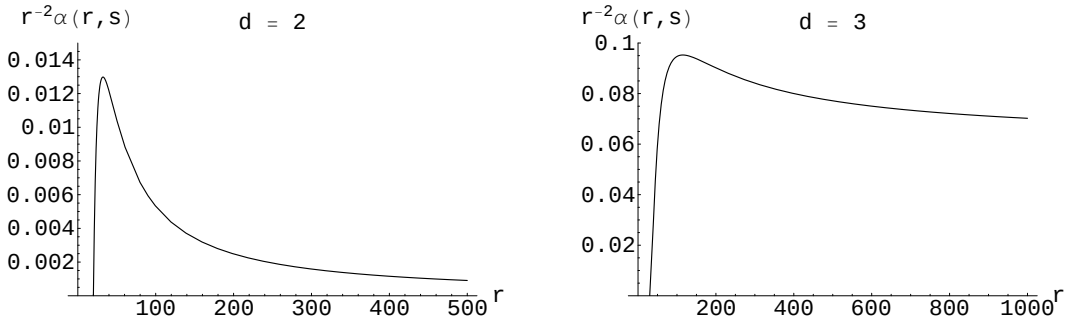
is fulfilled for $r \to \infty$ and $v \to \infty$ with $v \leq r$, as in Theorem 5.3.4.

Since, thus, the density of the Pickands coordinates with regard to Fréchet margins lies under certain regularity conditions close to the angular density for distributions in δ–neighborhoods, we have, as in Section 5.2, the possibility to estimate the angular density by means of (5.7). One has to bear in mind that here one can only use observations, for which $C^{(i)} < -r$ and additionally $Z^{(i)} \in Q_{v,s}$, i.e., $X^{(i)} \in A_{r,s,v}$, holds.

Example 5.3.6
In continuation of Example 5.3.3, we look at extreme value distributions. Before we estimate the angular density, we investigate condition (5.9) from Theorem 5.3.4 for the extreme value case and the special choice of

$$q(z) = \frac{1}{z_1^2 \cdot z_2^2 \cdot \ldots \cdot z_{d-1}^2 \left(1 - \sum_{i=1}^{d-1} z_i\right)^2},$$

see Example 5.3.3. We choose $r = v$ and plot numerically with `NIntegrate` the graphs of $r^{-2}\alpha(r, 0.1)$ for the cases $d = 2$ and $d = 3$.



In the case $d = 2$, we have clear convergence to 0, in the case $d = 3$, this convergence is already much slower. We note that in low dimensions the choice of $r = v$ is sufficient to imply the convergence (5.9), in higher dimensions one should choose $v = r^{\xi}$ with $\xi < 1$ to slow down the convergence of α to ∞. Further numerical investigations show that $\xi = \frac{1}{d-1}$ seems to be a good choice.

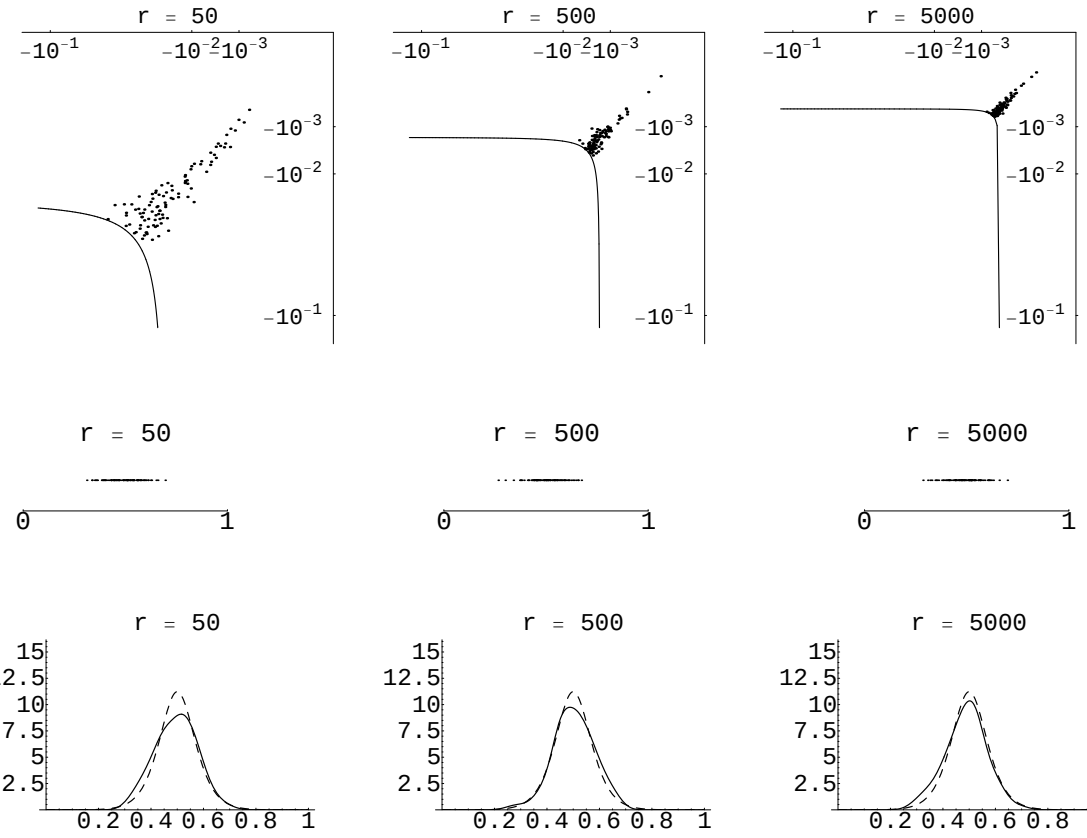
In the extreme value case, we evaluate the estimator $\hat{l}_{m,r}(z)$ with simulated data. Therefore, observations are generated, which are distributed according to an extreme value distribution G_λ of logistic type, i.e.,

$$G_\lambda(x_1, \ldots, x_d) = \exp\left(\left(\sum_{i=1}^{d} x_i\right) D_\lambda\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, \ldots, \frac{x_{d-1}}{\sum_{i=1}^{d} x_i}\right)\right) = \exp\left(-||x||_\lambda\right)$$

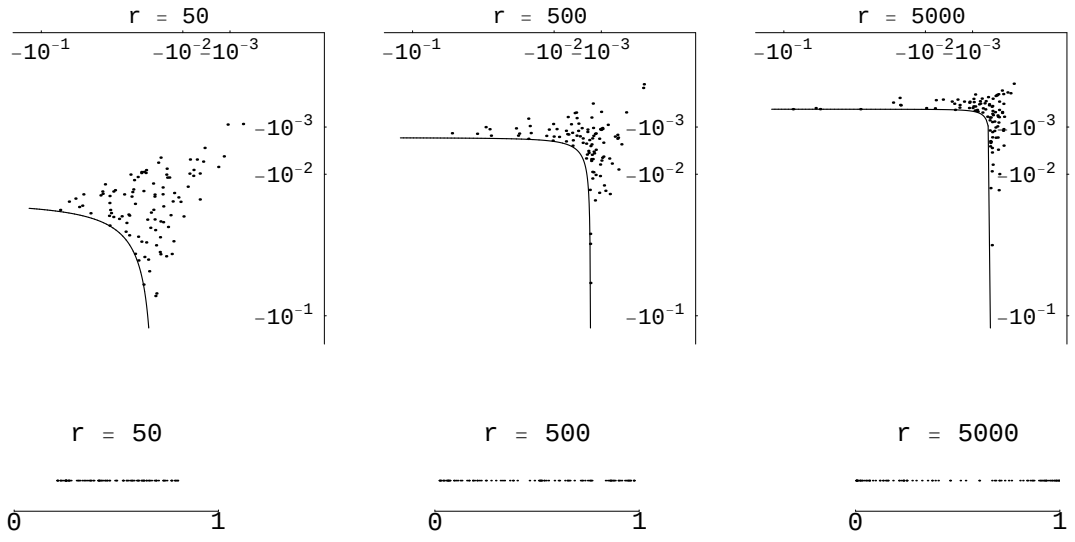
with D_λ from Definition 2.3.3. As shown in Example 5.3.3, this distribution function is in the 1–neighborhood of W_λ. An algorithm for the generation of these random vectors can be found in Stephenson (2003, [70]), Algorithm 1.1.

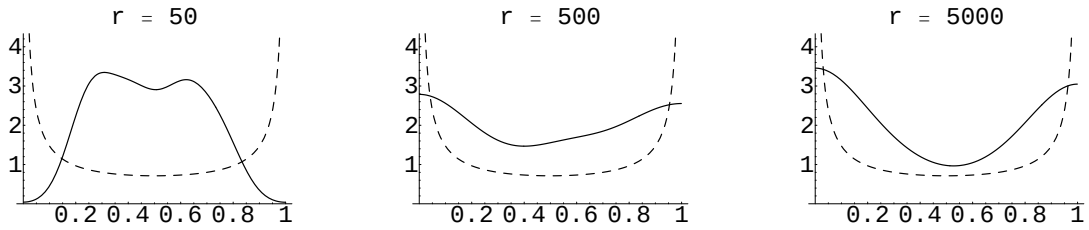
We choose $s = 0.1$, and in each case 100 data points are generated, for which $C < -r$ and $Z \in Q_{v,s}$ with $v, r = 50, 500, 5000$ hold respectively. Note that for these examples we always

chose $r = v$. For $\lambda = 6$ the following plots result (the original data are again transformed to a logarithmic scale):



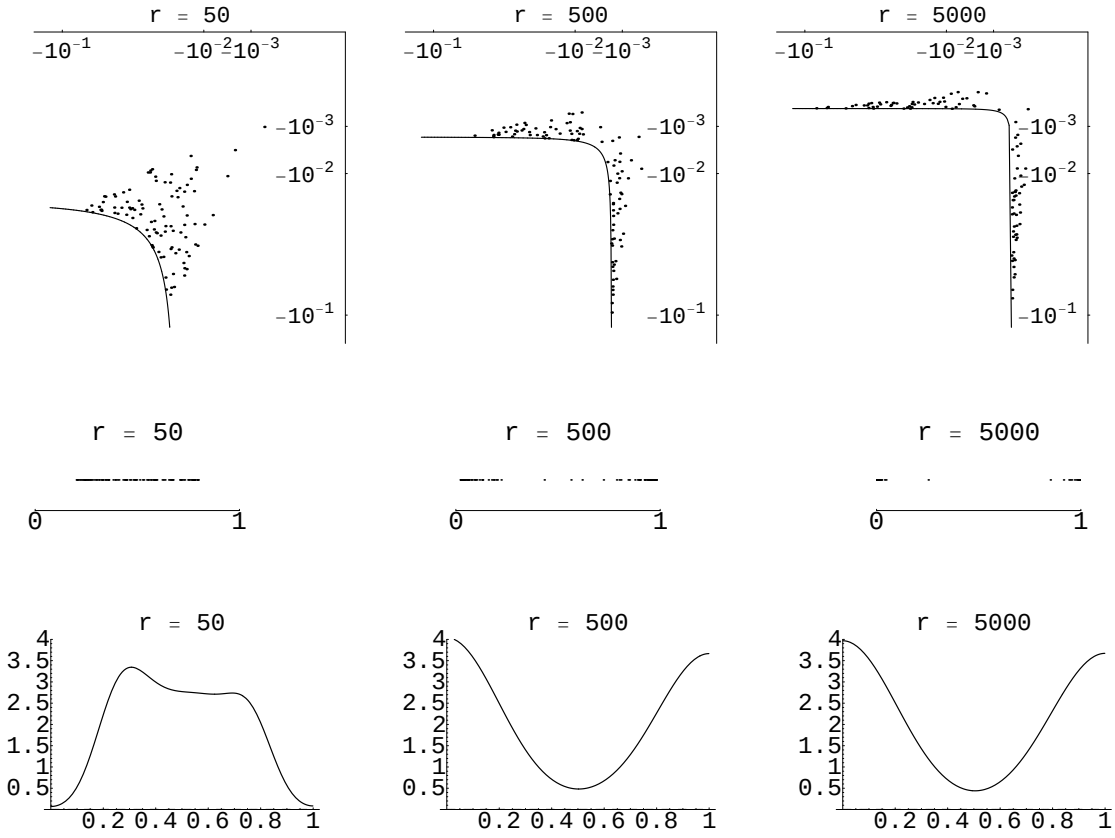
Once again we also investigate the case $\lambda = 1.2$.





As in Example 5.2.1 we see that cases with low dependence are more difficult to grasp.

Another interesting case, which is not covered by Theorems 5.3.4 and 5.3.5, is the case of independence, since we have $d^* = 0$ there. Nevertheless, we also investigate the estimator here for $d = 2$. The underlying angular measure is the point measure on $\{0, 1\}$. The angular density vanishes on the interior of R_{d-1}. But the angular components should accumulate at 0 and 1, and when one estimates the density of these, one should get a graph, which is close to 0 in the central area and converges at the boundary quickly to ∞.

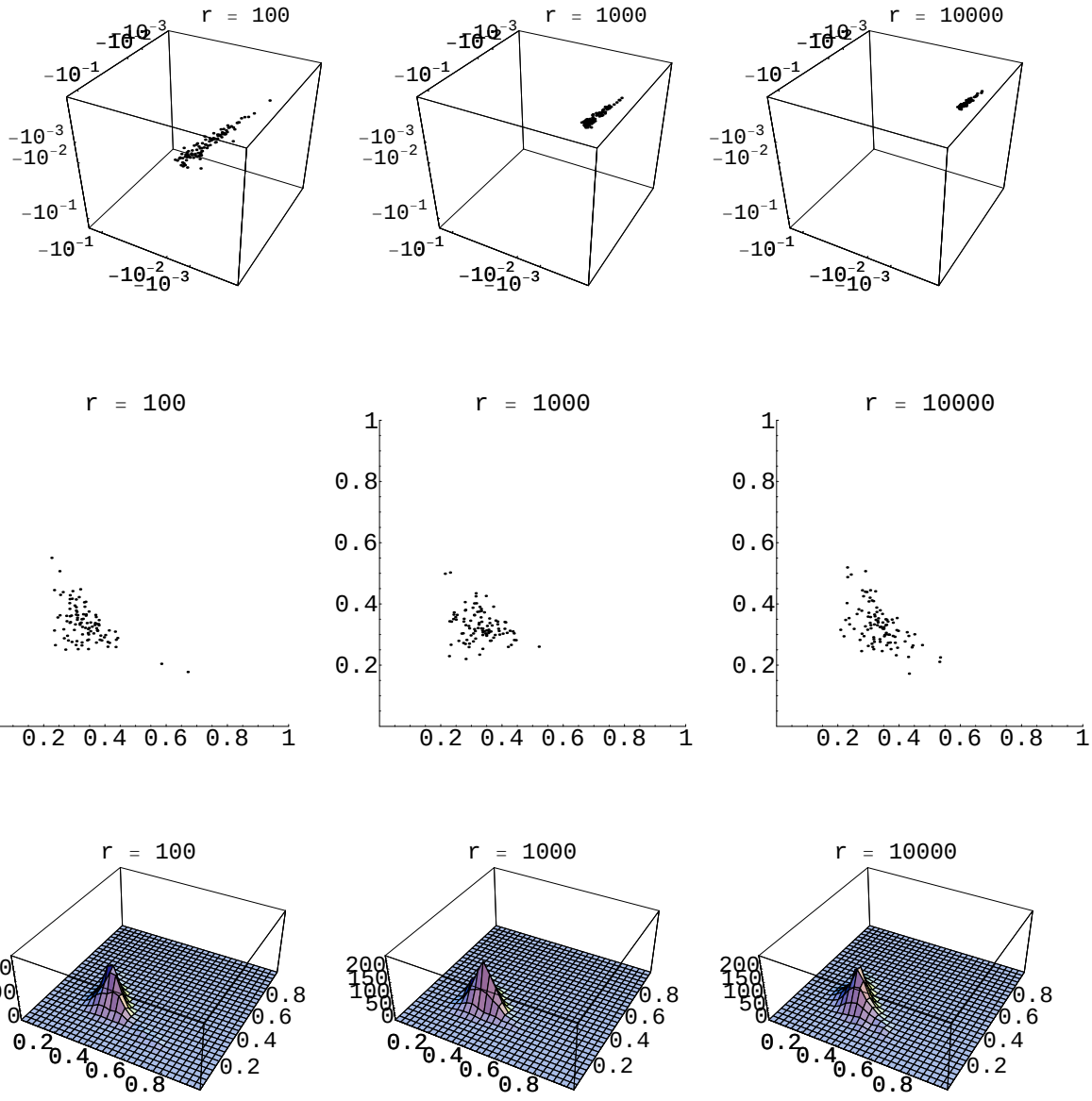


With the chosen bandwidth a better result of the kernel density estimator is not possible, since the angular components are almost only on the points 0 and 1 and are, thus, almost perfectly separated.

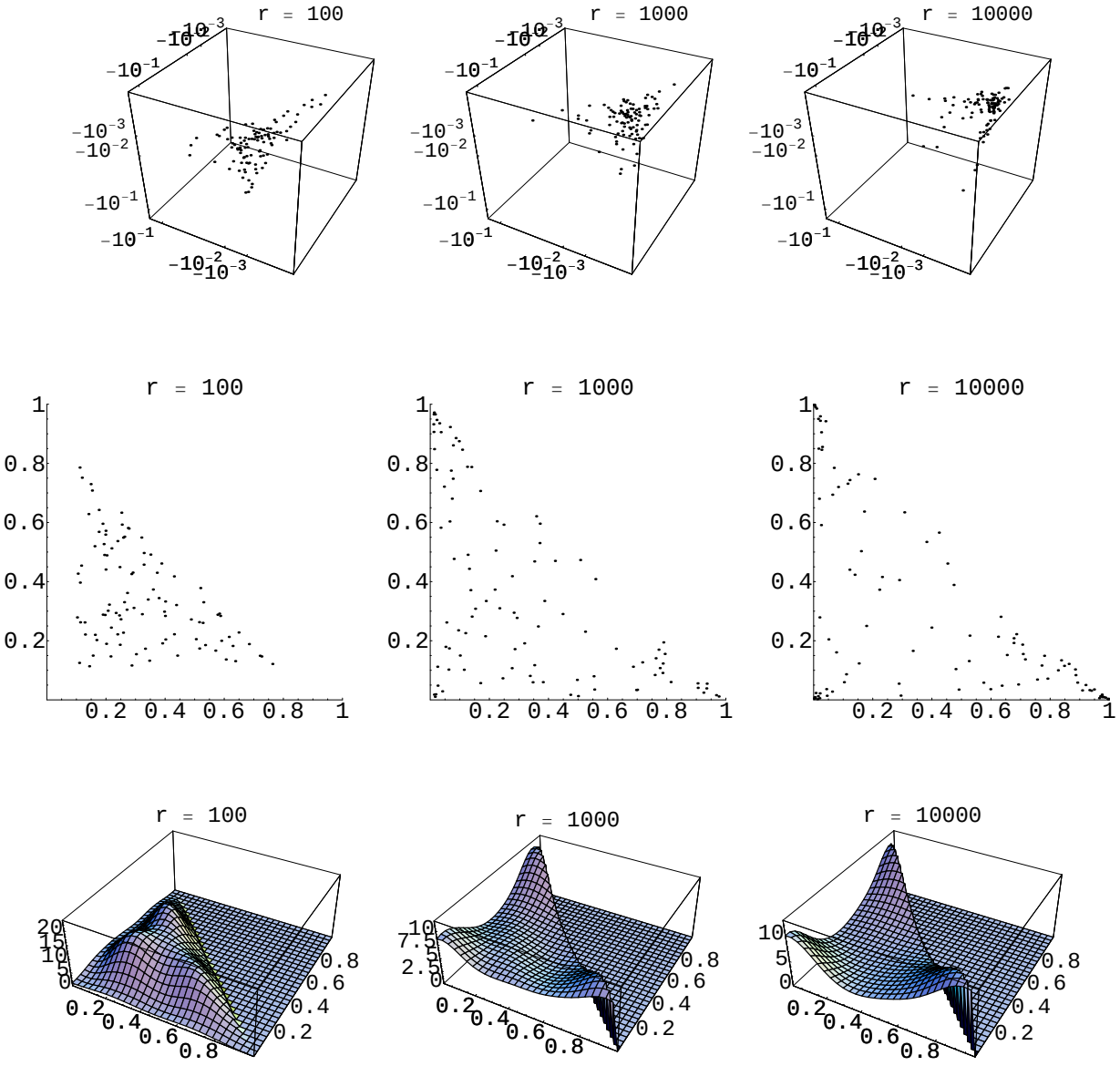
A further question would be whether the transformation to Pickands coordinates with regard to Fréchet margins in these δ-neighborhoods generates random variables, which have (asymptotically) the standardized angular distribution as distribution, although one knows that no

angular density exists or it vanishes in the interior of R_{d-1}. A mathematical affirmation of this is indicated in Joe et al. (1992, [43]) and Coles and Tawn (1994, [11]) for the extreme value case.

Again we investigate the case $d = 3$. For comparison the parameters are chosen exactly as in Example 5.2.1 with $r = v$. At first again $\lambda = 6$.



Then $\lambda = 1.2$.



The results resemble those in Example 5.2.1, see the corresponding remarks there.

◊

The estimation of the angular density in the extreme value case has been done in several papers, for example Coles and Tawn (1991, [10]), Coles and Tawn (1994, [11]), Coles et al. (1999, [9]) or Joe et al. (1992, [43]) with the help of the Pickands coordinates with regard to Fréchet margins. In contrast to our approach, an extreme value model was assumed which is valid not only close to the origin. Then one does not need to introduce the parameter v and is able to work with all angular components for which $C^{(i)} < -r$ holds. But also then one has problems and instability of the estimator close to the independence case. This can especially be seen in Section 4 of Coles et al. (1999, [9]), where only the estimated angular density in an interior set of R_{d-1} away from the boundary is considered reliable. This is also done here with the

introduction of v. In Section 5 of Joe et al. (1992, [43]), an example of this procedure with real data can be found and a practical way of choosing r is given.

The difference of our estimation procedure to the ones in the above cited references are the weaker assumptions of our estimator. We only assume an extreme value distribution close to the origin, for example in a cube around the origin, whereas the other procedures need that assumption for a larger area. With the concept of the δ-neighborhoods we are also able to transfer that method to other distribution functions close to a GPD, not only extreme value distributions.

5.4 Conditional Expectation and Conditional Variance of the Nonparametric Estimator

In the next three sections, we want to examine some mathematical properties of estimator (5.7), especially we will compute its asymptotic distribution. The asymptotic normality of both univariate and multivariate kernel density estimators under suitable regularity conditions is well known, see for example Chapters 2 and 3 of Prakasa Rao (1983, [55]), Chapter 4 of Wertz (1978, [74]) or Powell (2002, [54]). Since we base our estimator on observations, of which the distribution is only asymptotically known, the above references cannot directly be used as in Chapter 4. Instead, we will proceed analogously to Powell (2002, [54]) to show the asymptotic normality for our estimator under suitable regularity conditions.

For a mathematical simplification we look at the asymptotic properties without data sphering and reflection techniques, i.e., we consider an estimator of the shape

$$\hat{l}_{m,r}(z) = d \cdot \frac{1}{mh^{d-1}} \sum_{i=1}^{m} k\left(\frac{z - Z^{(i)}}{h}\right). \tag{5.13}$$

To our knowledge the computation of the asymptotic distribution of kernel density estimators with data sphering is, still, an open problem.

At first we assume that the m independent observations $X^{(1)}, \ldots, X^{(m)}$ with Pickands coordinates with regard to Fréchet margins $(Z^{(1)}, -r_1), \ldots, (Z^{(m)}, -r_m)$ and $||X^{(i)}||_\infty < s$ for $s > 0$ and $i = 1, \ldots, m$ are the basis of this estimator. Thereby the $X^{(i)}$ follow a generalized Pareto distribution W with angular density $l(z)$ and $r_i > r > 0$ holds for $i = 1, \ldots, m$. We begin by examining the expectation, variance and asymptotic distribution under the condition that the m observations fulfill $C^{(i)} < -r$ and that $C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m$ holds. Then we will remove these conditions in Section 5.6 and get to the unconditional asymptotic distribution of estimator (5.13).

By Theorem 5.1.6 we know that every $Z^{(i)}$ possesses the density

$$l_i(z) := \begin{cases} \frac{l(z)}{\chi(r_i,s)} & \text{for } z \in Q_{r_i,s}, \\ 0 & \text{else,} \end{cases} \tag{5.14}$$

under the above conditions. Thereby $Q_{r,s}$ and $\chi(r,s)$ are defined as in (5.2) and Definition 5.1.3 respectively. With the notation

$$\zeta_{im} := \frac{d}{h^{d-1}} k\left(\frac{z - Z^{(i)}}{h}\right),$$

we can write (5.13) as

$$\hat{l}_{m,r}(z) = \frac{1}{m}\sum_{i=1}^{m}\zeta_{im}.$$

Throughout the following we assume that $z \in Q_{r,s}$ and $l(z) \neq 0$ holds with z arbitrary but fixed. Furthermore, we assume that the underlying angular density $l(z)$ is smooth enough to have a Taylor series expansion of the form

$$l(z-hu) = l(z) - h\left(\nabla_z l(z)\cdot u\right) + \frac{h^2}{2}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}u_{j_1}u_{j_2}\right)_{1\le j_1,j_2\le d-1} + o(h^2) \tag{5.15}$$

for $h \to 0$ and $u \in \mathbb{R}^{d-1}$, see Theorem 168.2 in Heuser (1998, [38]). By $\nabla_z l$ we denote the gradient of l and with $(u\cdot v)$ the usual scalar product of two vectors u and v. Since this expansion implies that l is continuous, we have $d^* > 0$. The remainder term shall be bounded from above for every compact subset $U \subset \mathbb{R}^{d-1}$ by the third power of h and a constant M_U, depending only on U, i.e.,

$$\sup_{u\in U}\left|l(z-hu) - \left(l(z) - h\left(\nabla_z l(z)\cdot u\right) + \frac{h^2}{2}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}u_{j_1}u_{j_2}\right)_{1\le j_1,j_2\le d-1}\right)\right| \le M_U h^3, \tag{5.16}$$

thus being $o(h^2)$ for $u \in U$.

Also $l_i(z)$ has for $z \in Q_{r,s}$ and h small enough the same Taylor series expansion as (5.15) with the additional factor $\frac{1}{\chi(r_i,s)}$.

Let the kernel $k \ge 0$ have bounded support and for $j, j_1, j_2 \in \{1,\ldots,d-1\}$ fulfill

$$\int k(u)\,du = 1, \qquad \int u_j k(u)\,du = 0, \qquad \int u_{j_1}u_{j_2}k(u)\,du < \infty,$$

$$\int k^2(u)\,du < \infty \qquad \text{and} \qquad \int k^3(u)\,du < \infty. \tag{5.17}$$

The normal kernel for example fulfills conditions (5.17), but does not have bounded support, thus the asymptotic theory developed her is not valid for the normal kernel. The Epanechnikov kernel, however, has bounded support and fulfills conditions (5.17) and is, thus, to be preferred from an asymptotic viewpoint.

We begin with the asymptotic theory by examining the expectation.

Lemma 5.4.1
Under the above assumptions we have for the conditional expectation

$$\begin{aligned} &E\left(\zeta_{im}\,\middle|\,C^{(i)} = r_i, i = 1,\ldots,m\right) \\ &= \frac{d}{\chi(r_i,s)}l(z) + \frac{h^2}{2}\frac{d}{\chi(r_i,s)}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}\int u_{j_1}u_{j_2}k(u)\,du\right) + o\left(h^2\right), \end{aligned}$$

and the convergence of the remainder term is uniform in r.

Proof:
With a substitution, the Taylor series (5.15) and conditions (5.17) we get

$$\begin{aligned}
&E\left(\zeta_{im}\left|C^{(i)}=r_i, i=1,\dots,m\right.\right)\\
&= E\left(\frac{d}{h^{d-1}}k\left(\frac{z-Z^{(i)}}{h}\right)\left|C^{(i)}=r_i, i=1,\dots,m\right.\right)\\
&= \int \frac{d}{h^{d-1}}k\left(\frac{z-x}{h}\right)l_i(x)\,dx = \int dk(u)l_i(z-hu)\,du\\
&\overset{(5.15)}{=} \int dk(u)\frac{1}{\chi(r_i,s)}\left(l(z)-h\left(\nabla_z l(z)\cdot u\right)+\frac{h^2}{2}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}u_{j_1}u_{j_2}\right)+o\left(h^2\right)\right)du\\
&= \frac{d}{\chi(r_i,s)}l(z)\int k(u)\,du - h\frac{d}{\chi(r_i,s)}\left(\left(\nabla_z l(z)\right)\cdot\left(\int u_j k(u)\,du\right)_{1\le j\le d-1}\right)\\
&\quad +\frac{h^2}{2}\frac{d}{\chi(r_i,s)}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}\int u_{j_1}u_{j_2}k(u)\,du\right)+\int dk(u)\frac{1}{\chi(r_i,s)}o(h^2)\,du\\
&\overset{(5.17)}{=} \frac{d}{\chi(r_i,s)}l(z)+\frac{h^2}{2}\frac{d}{\chi(r_i,s)}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}\int u_{j_1}u_{j_2}k(u)\,du\right)+o\left(h^2\right).
\end{aligned}$$

We will show the above equality for the remainder term next. Since the kernel has bounded support $S := \overline{\{x\in\mathbb{R}^{d-1}|k(x)>0\}}$, and the remainder of the Taylor series is bounded from above by $M_S h^3$ for $u\in S$, see (5.16), the remainder term can be written as

$$\int dk(u)\frac{1}{\chi(r_i,s)}o(h^2)\,du \le \frac{dM_S}{\chi(r_i,s)}h^3.$$

Due to $\chi(r,s)\le\chi(r_i,s)\le d^*$ and $\chi(r,s)\to_{r\to\infty} d^*>0$ by Lemma 5.1.4, we have for the remainder term then

$$\int dk(u)\frac{1}{\chi(r_i,s)}o(h^2)\,du \le \frac{dM_S}{\chi(r_i,s)}h^3 \le \frac{dM_S}{\chi(r,s)}h^3 = o(h^2),$$

uniformly in r since the factor $\frac{dM_S}{\chi(r,s)}$ is bounded.

□

Theorem 5.4.2
Under the above assumptions we have

$$\begin{aligned}
&E\left(\hat{l}_{m,r}(z)\left|C^{(i)}=r_i, i=1,\dots,m\right.\right)\\
&= l(z)\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}+\frac{h^2}{2}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}\int u_{j_1}u_{j_2}k(u)\,du\right)\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}+o\left(h^2\right),
\end{aligned}$$

where the remainder term is uniform in m and r. Thereby

$$\frac{d}{d^*}\le\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}\le\frac{d}{\chi(r,s)}\to\frac{d}{d^*} \tag{5.18}$$

for $r\to\infty$.

Proof:
With Lemma 5.4.1,

$$\begin{aligned} E\left(\hat{l}_{m,r}(z)\,\middle|C^{(i)}=r_i, i=1,\ldots,m\right) &= \frac{1}{m}\sum_{i=1}^{m} E\left(\zeta_{im}\,\middle|C^{(i)}=r_i, i=1,\ldots,m\right) \\ &= \frac{1}{m}\sum_{i=1}^{m}\frac{d}{\chi(r_i,s)}l(z)+\frac{1}{m}\sum_{i=1}^{m}\frac{h^2}{2}\frac{d}{\chi(r_i,s)}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}\int u_{j_1}u_{j_2}k(u)\,du\right)+o\left(h^2\right) \\ &= l(z)\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}+\frac{h^2}{2}\mathrm{tr}\left(\frac{\partial^2 l(z)}{\partial z_{j_1}\partial z_{j_2}}\int u_{j_1}u_{j_2}k(u)\,du\right)\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}+o\left(h^2\right) \end{aligned}$$

holds. For the second equality sign we used $\frac{1}{m}\sum_{i=1}^{m} o(h^2) = o(h^2)$. We will show now that this is correct. From the proof of Lemma 5.4.1, we know that for $o(h^2)$ an upper bound can be given in terms of $o(h^2) \leq \frac{dM_S}{\chi(r,s)}h^3$ with M_S from the Taylor series expansion (5.15). We have now

$$\frac{1}{m}\sum_{i=1}^{m} o(h^2) \leq \frac{1}{m}\sum_{i=1}^{m}\frac{d}{\chi(r,s)}M_S h^3 = \frac{M_S d}{\chi(r,s)}h^3 = o(h^2)$$

uniformly in m and r, since $\chi(r,s) \to_{r\to\infty} d^* > 0$ monotonously increasing.

The assertion (5.18) follows from Lemma 5.1.4, since

$$\frac{d}{d^*} = \frac{d}{m}\sum_{i=1}^{m}\frac{1}{d^*} \leq \frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} \leq \frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r,s)} = \frac{d}{\chi(r,s)}$$

and $\chi(r,s) \to d^*$ for $r \to \infty$.

□

Theorem 5.4.2 states that the bias of the estimator (5.13) depends on the parameters h, r_i and d^*. We will derive conditions for the rate of convergence of h at the end of this, and the beginning of the next section. Since the rate of convergence of $\chi(r,s) = \int_{Q_{r,s}} l(z)\,dz$ is crucial for a suitable rate of convergence of the parameter r, the underlying and commonly unknown angular density l itself plays an important role there. We have seen in Example 5.2.1 that this rate can be very different. Which conditions must be fulfilled here, and how they link up to rate of convergence of h will be examined in Section 5.5. The bias coming from d^* cannot be controlled, but it consists of, as we will shortly see, only a constant factor, which equals 1 in the case $d^* = d$.

Next we turn to the investigation of the variance.

Lemma 5.4.3
Under the above assumptions, we have

$$\sigma_{im}^2 := \mathrm{Var}\left(\zeta_{im}\,\middle|C^{(i)}=r_i, i=1,\ldots,m\right) = \frac{d^2}{\chi(r_i,s)h^{d-1}}l(z)\int k^2(u)\,du + O\left(\frac{1}{h^{d-2}}\right)$$

with uniform convergence of the remainder term in r.

Proof:
Additional to Lemma 5.4.1 we use again a substitution, the Taylor series (5.15) and the conditions (5.17)

$$\sigma_{im}^2 = \mathrm{Var}(\zeta_{im}) = E(\zeta_{im}^2) - E(\zeta_{im})^2$$

$$
\begin{aligned}
&= E\left(\left(\frac{d}{h^{d-1}}k\left(\frac{z-Z^{(i)}}{h}\right)\right)^2\right) - E(\zeta_{im})^2 \\
&= \int \frac{d^2}{h^{2(d-1)}}k^2\left(\frac{z-x}{h}\right) l_i(x)\,dx - E(\zeta_{im})^2 \\
&= \frac{d^2}{h^{d-1}}\int k^2(u) l_i(z-hu)\,du - E(\zeta_{im})^2 \\
&= \frac{d^2}{h^{d-1}}\int k^2(u)\left(\frac{l(z)+O(h)}{\chi(r_i,s)}\right)du - E(\zeta_{im})^2 \\
&= \frac{d^2}{\chi(r_i,s)h^{d-1}}l(z)\int k^2(u)\,du + O\left(\frac{1}{h^{d-2}}\right) - E(\zeta_{im})^2 \\
&= \frac{d^2}{\chi(r_i,s)h^{d-1}}l(z)\int k^2(u)\,du + O\left(\frac{1}{h^{d-2}}\right),
\end{aligned}
$$

since $E(\zeta_{im})$ converges to $\frac{d}{d^*}\cdot l(z)$ for $h \to 0$, $r \to \infty$. The uniform convergence of the remainder term in r follows as in the proof of Lemma 5.4.1.

□

Theorem 5.4.4
Under the above assumptions we have

$$
\mathrm{Var}\left(\hat{l}_{m,r}(z)\,\middle|\,C^{(i)} = r_i, i = 1,\ldots,m\right) = l(z)\frac{1}{m^2}\left(\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}\right)\frac{d^2}{h^{d-1}}\int k^2(u)\,du + O\left(\frac{1}{mh^{d-2}}\right)
$$

with the convergence of the remainder term being uniformly in r.

Proof:
We have with Lemma 5.4.3 and the independence of the underlying observations

$$
\begin{aligned}
&\mathrm{Var}\left(\hat{l}_{m,r}(z)\,\middle|\,C^{(i)} = r_i, i = 1,\ldots,m\right) \\
&= \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m}\zeta_{im}\,\middle|\,C^{(i)} = r_i, i = 1,\ldots,m\right) = \frac{1}{m^2}\sum_{i=1}^{m}\sigma_{im}^2 \\
&= \frac{1}{m^2}\left(\frac{d^2}{h^{d-1}}l(z)\int k^2(u)\,du\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O\left(\frac{m}{h^{d-2}}\right)\right) \\
&= l(z)\frac{1}{m^2}\left(\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}\right)\frac{d^2}{h^{d-1}}\int k^2(u)\,du + O\left(\frac{1}{mh^{d-2}}\right).
\end{aligned}
$$

The uniform convergence of the remainder term in r and m follows as in Theorem 5.4.2.

□

Remark 5.4.5
The inequalities

$$
\frac{1}{d^*} \le \frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} \le \frac{1}{\chi(r,s)}, \tag{5.19}
$$

follow as in Theorem 5.4.2, thus an upper bound for the variance in Theorem 5.4.4 can be given by

$$\mathrm{Var}\left(\hat{l}_{m,r}(z)\,\middle|C^{(i)}=r_i, i=1,\ldots,m\right) \leq l(z)\frac{1}{m}\frac{1}{\chi(r,s)}\frac{d^2}{h^{d-1}}\int k^2(u)\,du + O\left(\frac{1}{mh^{d-2}}\right).$$

The term on the right side goes to 0 for $mh^{d-1} \to \infty$, where $m \to \infty$ and $h \to 0$. Also we have $l(z)\frac{d^2}{d^* mh^{d-1}}\int k^2(u)\,du + O\left(\frac{1}{mh^{d-2}}\right)$ for $r \to \infty$, which is by (5.19) a lower bound for the variance. This term converges for $mh^{d-1} \to \infty$ to 0 as well.

$\Diamond$

5.5 Conditional Asymptotic Distribution of the Nonparametric Estimator

In this section we consider the asymptotic distribution of the estimator (5.13). In this we follow the line of arguments suggested by Powell (2002, [54]). Before we begin though, we need some auxiliary results.

Lemma 5.5.1
Let X be a random variable. The a–norm is defined as $||X||_a := \left(E\left(|X|^a\right)\right)^{1/a}$ for $a \geq 1$ by the ath absolute moment. Then for all a, b with $1 \leq b \leq a$ the relation

$$||X||_b \leq ||X||_a \tag{5.20}$$

holds.

Proof:
The Hölder inequality (see Gänssler and Stute (1977, [30]), Theorem 1.13.2) states that for two random variables Y and Z and $p, q \geq 1$ with $\frac{1}{p}+\frac{1}{q}=1$, we have

$$||YZ||_1 = E(|YZ|) \leq \left(E\left(|Y|^p\right)\right)^{1/p}\left(E\left(|Z|^q\right)\right)^{1/q} = ||Y||_p||Z||_q. \tag{5.21}$$

By putting $Y = |X|^b$, $Z = 1$, $p = \frac{a}{b}$ and $q = \frac{a}{a-b}$ in (5.21), we arrive at

$$E\left(|X|^b\right) \leq \left(E\left(|X|^a\right)\right)^{b/a}\left(E\left(1^{a/(a-b)}\right)\right)^{(a-b)/a} = \left(E\left(|X|^a\right)\right)^{b/a}.$$

By taking the bth root on both sides (5.20) follows.

$\square$

Lemma 5.5.2
We have for $n \in \mathbb{N}$

$$||X - E(X)||_n \leq 2||X||_n.$$

Especially we have for the nth absolute central moment that it has the (suitably scaled) uncentered absolute moment as an upper bound, more precisely

$$E(|X - E(X)|^n) \leq 2^n E(|X|^n).$$

Proof:
With the triangular inequality, the binomial formula, Lemma 5.5.1 and the formula

$$2^n = (1+1)^n = \sum_{i=0}^{n}\binom{n}{i}1^i 1^{n-i} = \sum_{i=0}^{n}\binom{n}{i},$$

the relation

$$\begin{aligned} E(|X - E(X)|^n) &\leq E\left((|X| + E(|X|))^n\right) = E\left(\sum_{i=0}^{n} \binom{n}{i} |X|^i E(|X|)^{n-i}\right) \\ &= \sum_{i=0}^{n} \binom{n}{i} E(|X|^i) E(|X|)^{n-i} = \sum_{i=0}^{n} \binom{n}{i} ||X||_i^i ||X||_1^{n-i} \\ &\leq \sum_{i=0}^{n} \binom{n}{i} ||X||_n^i ||X||_n^{n-i} = ||X||_n^n \sum_{i=0}^{n} \binom{n}{i} = 2^n ||X||_n^n \end{aligned}$$

holds and, thus, the assertions follow.

□

To prove the asymptotic normality of estimator (5.13), we use the version of the Berry–Esséen Theorem from Section 1.9.5 of Serfling (1980, [63]). It assumes independent but not necessarily identically distributed random variables ζ_{im} with variance $\mathrm{Var}(\zeta_{im}) = \sigma_{im}^2$ and existing ath central moment $\rho_{im} := E(|\zeta_{im} - E(\zeta_{im})|^a) < \infty$ for an $a > 2$. If the Liapunov condition

$$\frac{\left(\sum_{i=1}^{m} \rho_{im}\right)^{1/a}}{\left(\sum_{i=1}^{m} \sigma_{im}^2\right)^{1/2}} \to 0 \tag{5.22}$$

holds for an $a > 2$, the sum

$$\bar{\zeta}_m = \frac{1}{m} \sum_{i=1}^{m} \zeta_{im}$$

is asymptotically normal, i.e.,

$$\frac{\bar{\zeta}_m - E(\bar{\zeta}_m)}{\sqrt{\mathrm{Var}(\bar{\zeta}_m)}} \longrightarrow_{\mathcal{D}} \mathcal{N}(0,1).$$

The Berry-Esséen Theorem from Section 1.9.5 of Serfling (1980, [63]) makes additional specifications on the rate of convergence, which we will use in the following lemma.

Lemma 5.5.3
Under the above assumptions, we have under the condition $C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m$ the convergence

$$\frac{\hat{l}_{m,r}(z) - E\left(\hat{l}_{m,r}(z)\right)}{\sqrt{\mathrm{Var}\left(\hat{l}_{m,r}(z)\right)}} \longrightarrow_{\mathcal{D}} \mathcal{N}(0,1), \tag{5.23}$$

if $mh^{d-1} \to \infty$ for $m \to \infty$ and $h \to 0$. More precisely

$$P\left(\frac{\hat{l}_{m,r}(z) - E\left(\hat{l}_{m,r}(z)\right)}{\sqrt{\mathrm{Var}\left(\hat{l}_{m,r}(z)\right)}} \leq x \,\middle|\, C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m\right) = \Phi(x) + O\left(\left(mh^{d-1}\right)^{-1/6}\right),$$

where Φ notes the distribution function of the standard normal distribution, and the remainder term converges uniformly in r.

Proof:
Our goal is to verify the Liapunov condition for $a = 3$. For that purpose we have with Lemma 5.5.2 and again the Taylor series expansion of l as in the lemmata of Section 5.4

$$\begin{aligned}\rho_{im} &= E(|\zeta_{im} - E(\zeta_{im})|^3) \le 8E(|\zeta_{im}|^3) \\ &= \frac{8d^3}{\chi(r_i,s)h^{2(d-1)}} l(z) \int k^3(u)\,du + o\left(\frac{1}{h^{2(d-1)}}\right)\end{aligned}$$

uniform in r. Together with Lemma 5.4.3 we get

$$\begin{aligned}\frac{\left(\sum_{i=1}^m \rho_{im}\right)^{1/3}}{\left(\sum_{i=1}^m \sigma_{im}^2\right)^{1/2}} &\le \frac{\left(\sum_{i=1}^m \frac{8d^3}{\chi(r_i,s)h^{2(d-1)}} l(z) \int k^3(u)\,du + o\left(\frac{1}{h^{2(d-1)}}\right)\right)^{1/3}}{\left(\sum_{i=1}^m \frac{d^2}{\chi(r_i,s)h^{d-1}} l(z) \int k^2(u)\,du + o\left(\frac{1}{h^{d-1}}\right)\right)^{1/2}} \\ &= \frac{\left(\frac{8d^3}{h^{2(d-1)}} l(z) \int k^3(u)\,du \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + o\left(\frac{m}{h^{2(d-1)}}\right)\right)^{1/3}}{\left(\frac{d^2}{h^{d-1}} l(z) \int k^2(u)\,du \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + o\left(\frac{m}{h^{d-1}}\right)\right)^{1/2}} \\ &\overset{(5.19)}{\le} \frac{\left(\frac{m}{\chi(r,s)} \frac{8d^3}{h^{2(d-1)}} l(z) \int k^3(u)\,du + o\left(\frac{m}{h^{2(d-1)}}\right)\right)^{1/3}}{\left(\frac{m}{d^*} \frac{d^2}{h^{d-1}} l(z) \int k^2(u)\,du + o\left(\frac{m}{h^{d-1}}\right)\right)^{1/2}} \\ &= O\left(\frac{m^{1/3}}{h^{\frac{2}{3}(d-1)}}\right) \cdot O\left(\frac{m^{-1/2}}{h^{-(d-1)/2}}\right) \\ &= O\left(\left(mh^{d-1}\right)^{-1/6}\right) \to 0,\end{aligned}$$

if $mh^{d-1} \to \infty$ for $m \to \infty$ and $h \to 0$, the same condition, which also guaranteed the convergence of the variance. The convergence is uniform in r, see the third to last line of the calculations above. The assertions now follow from the Theorem of Berry–Esséen, see Section 1.9.5 in Serfling (1980, [63]).

□

The targeted value of the estimator $\hat{l}_{m,r}(z)$ is actually $l(z)$ for $d^* = d$ or a constant multiple of $l(z)$ in case $d^* < d$, and not the expectation of $\hat{l}_{m,r}(z)$. By Theorem 5.4.2, the expectation is in general different from the targeted value. This bias of the estimation vanishes asymptotically though, which we want to show next.

Theorem 5.5.4
Under the above assumptions, we have, with the notation $\sigma := \sqrt{\frac{d^2}{d^*} l(z) \int k^2(u)\,du}$,

$$\begin{aligned}&P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - \frac{d}{d^*} l(z)\right) \le x \,\middle|\, C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m\right) \\ &= \Phi\left(\frac{x}{\sigma}\right) + O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}(d^* - \chi(r,s))\right)\end{aligned}$$

uniformly in h, m and r. I.e., under $m \to \infty$, $h \to 0$, $r \to \infty$, $mh^{d-1} \to \infty$, $mh^{d+3} \to 0$, $\sqrt{mh^{d-1}}(d^* - \chi(r,s)) \to 0$ the estimator $\hat{l}_{m,r}(z)$, conditional on $C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m$, is asymptotically unbiased and normally distributed.

Proof:
Before we can prove the main result, we have to show two auxiliary results, namely

$$O\left(\sqrt{\frac{d^2}{d^*}l(z)\int k^2(u)\,du} - \sqrt{d^2 l(z)\int k^2(u)\,du\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O(h)}\right) = O\left(h + (d^* - \chi(r,s))\right) \tag{5.24}$$

and

$$O\left(\sqrt{mh^{d-1}}\left(\frac{d}{d^*}l(z) - l(z)\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O\left(h^2\right)\right)\right) = O\left(\sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}(d^* - \chi(r,s))\right). \tag{5.25}$$

Because of Lemma 5.1.4 it follows for (5.24) that

$$\begin{aligned}
&O\left(\sqrt{\frac{d^2}{d^*}l(z)\int k^2(u)\,du} - \sqrt{d^2 l(z)\int k^2(u)\,du\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O(h)}\right)\\
&= O\left(\frac{\frac{d^2}{d^*}l(z)\int k^2(u)\,du - l(z)d^2\int k^2(u)\,du\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O(h)}{\sqrt{\frac{d^2}{d^*}l(z)\int k^2(u)\,du} + \sqrt{d^2 l(z)\int k^2(u)\,du\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O(h)}}\right)\\
&= O\left(\frac{d^2}{d^*}l(z)\int k^2(u)\,du - l(z)d^2\int k^2(u)\,du\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O(h)\right)\\
&= O\left(h + d^2 l(z)\int k^2(u)\,du\left(\frac{1}{d^*} - \frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}\right)\right)\\
&= O\left(h + \left(\frac{1}{d^*} - \frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}\right)\right)\\
&= O\left(h + \left(\frac{1}{d^*} - \frac{1}{m}\sum_{i=1}^{m}\frac{1}{\chi(r,s)}\right)\right)\\
&= O\left(h + \left(\frac{1}{d^*} - \frac{1}{\chi(r,s)}\right)\right)\\
&= O\left(h + \left(\frac{\chi(r,s) - d^*}{d^*\chi(r,s)}\right)\right)\\
&= O\left(h + (d^* - \chi(r,s))\right).
\end{aligned}$$

Again with Lemma 5.1.4, equation (5.25) also follows as above

$$\begin{aligned}
&O\left(\sqrt{mh^{d-1}}\left(\frac{d}{d^*}l(z) - l(z)\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)} + O\left(h^2\right)\right)\right)\\
&= O\left(\sqrt{mh^{d+3}} + l(z)\sqrt{mh^{d-1}}\left(\frac{d}{d^*} - \frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}\right)\right)\\
&= O\left(\sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}\left(d^* - \chi(r,s)\right)\right).
\end{aligned}$$

Throughout the following we will also need the Taylor series expansion

$$\Phi(x+h) - \Phi(x) = h\varphi(x) + o(h) = O(h) \tag{5.26}$$

of the distribution function of the standard normal distribution, see Theorem 168.1 in Heuser (1998, [38]). The convergence is independent of x, since φ and its derivative are bounded.

With the substitution

$$x = \frac{t}{\sqrt{d^2 l(z) \int k^2(u)\,du \frac{1}{m} \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + O(h)}},$$

we get from Lemma 5.5.3 and Theorem 5.4.4

$$\begin{aligned}
&P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - E\left(\hat{l}_{m,r}(z)\right)\right) \le t \,\middle|\, C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m\right) \\
&= \Phi\left(\frac{t}{\sqrt{d^2 l(z) \int k^2(u)\,du \frac{1}{m} \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + O(h)}}\right) + O\left(\left(mh^{d-1}\right)^{-1/6}\right) \\
&= \Phi\left(\frac{t}{\sigma}\right) + \Phi\left(\frac{t}{\sqrt{d^2 l(z) \int k^2(u)\,du \frac{1}{m} \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + O(h)}}\right) - \Phi\left(\frac{t}{\sigma}\right) + O\left(\left(mh^{d-1}\right)^{-1/6}\right) \\
&\overset{(5.26)}{=} \Phi\left(\frac{t}{\sigma}\right) + O\left(\frac{t}{\sqrt{d^2 l(z) \int k^2(u)\,du \frac{1}{m} \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + O(h)}} - \frac{t}{\sigma}\right) + O\left(\left(mh^{d-1}\right)^{-1/6}\right) \\
&= \Phi\left(\frac{t}{\sigma}\right) + O\left(\frac{t\sigma - t\sqrt{d^2 l(z) \int k^2(u)\,du \frac{1}{m} \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + O(h)}}{\sigma\sqrt{d^2 l(z) \int k^2(u)\,du \frac{1}{m} \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + O(h)}}\right) + O\left(\left(mh^{d-1}\right)^{-1/6}\right) \\
&= \Phi\left(\frac{t}{\sigma}\right) + O\left(\sigma - \sqrt{d^2 l(z) \int k^2(u)\,du \frac{1}{m} \sum_{i=1}^m \frac{1}{\chi(r_i,s)} + O(h)}\right) + O\left(\left(mh^{d-1}\right)^{-1/6}\right) \\
&\overset{(5.24)}{=} \Phi\left(\frac{t}{\sigma}\right) + O\left(h + (d^* - \chi(r,s)) + \left(mh^{d-1}\right)^{-1/6}\right).
\end{aligned}$$

Thus

$$\begin{aligned}
&\Phi\left(\frac{t}{\sigma}\right) + O\left(h + (d^* - \chi(r,s)) + \left(mh^{d-1}\right)^{-1/6}\right) \\
&= P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - E\left(\hat{l}_{m,r}(z)\right)\right) \le t \,\middle|\, C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m\right) \\
&= P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - \frac{d}{d^*} l(z)\right) + \sqrt{mh^{d-1}}\left(\frac{d}{d^*} l(z) - E\left(\hat{l}_{m,r}(z)\right)\right) \le t \,\middle|\right. \\
&\qquad \left. C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m\right) \\
&= P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - \frac{d}{d^*} l(z)\right) \le t - \sqrt{mh^{d-1}}\left(\frac{d}{d^*} l(z) - E\left(\hat{l}_{m,r}(z)\right)\right) \,\middle|\right. \\
&\qquad \left. C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m\right).
\end{aligned}$$

With the substitution

$$x = t - \sqrt{mh^{d-1}}\left(\frac{d}{d^*} l(z) - E\left(\hat{l}_{m,r}(z)\right)\right)$$

and Theorem 5.4.2, we finally arrive at

$$
\begin{aligned}
&P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z)-\frac{d}{d^*}l(z)\right)\le x\middle| C^{(1)}=-r_1,\dots,C^{(m)}=-r_m\right)\\
&= \Phi\left(\frac{x+\sqrt{mh^{d-1}}\left(\frac{d}{d^*}l(z)-E\left(\hat{l}_{m,r}(z)\right)\right)}{\sigma}\right)+O\left(h+(d^*-\chi(r,s))+\left(mh^{d-1}\right)^{-1/6}\right)\\
&= \Phi\left(\frac{x}{\sigma}\right)+\Phi\left(\frac{x+\sqrt{mh^{d-1}}\left(\frac{d}{d^*}l(z)-E\left(\hat{l}_{m,r}(z)\right)\right)}{\sigma}\right)-\Phi\left(\frac{x}{\sigma}\right)\\
&\quad +O\left(h+(d^*-\chi(r,s))+\left(mh^{d-1}\right)^{-1/6}\right)\\
&\overset{(5.26)}{=} \Phi\left(\frac{x}{\sigma}\right)+O\left(\frac{x+\sqrt{mh^{d-1}}\left(\frac{d}{d^*}l(z)-E\left(\hat{l}_{m,r}(z)\right)\right)-x}{\sigma}\right)\\
&\quad +O\left(h+(d^*-\chi(r,s))+\left(mh^{d-1}\right)^{-1/6}\right)\\
&= \Phi\left(\frac{x}{\sigma}\right)+O\left(\sqrt{mh^{d-1}}\left(\frac{d}{d^*}l(z)-l(z)\frac{d}{m}\sum_{i=1}^{m}\frac{1}{\chi(r_i,s)}+O\left(h^2\right)\right)\right)\\
&\quad +O\left(h+(d^*-\chi(r,s))+\left(mh^{d-1}\right)^{-1/6}\right)\\
&\overset{(5.25)}{=} \Phi\left(\frac{x}{\sigma}\right)+O\left(h+\left(mh^{d-1}\right)^{-1/6}+\sqrt{mh^{d+3}}+\sqrt{mh^{d-1}}(d^*-\chi(r,s))\right),
\end{aligned}
$$

which we had to show.

□

Note that for $d^*=d$, Theorem 5.5.4 states that we have an asymptotically unbiased estimator. The condition $\sqrt{mh^{d-1}}(d^*-\chi(r,s))\to 0$ can also be formulated as

$$\sqrt{mh^{d-1}}\left(\int_{R_{d-1}\setminus Q_{r,s}} l(z)\,dz\right)\to 0.$$

Thus the rate convergence at which r has to converge to ∞, does not only depend on m and h but also on the estimated angular density l itself. A sufficient condition for the above convergence, together with the other convergence conditions from Theorem 5.5.4, is, for example, that

$$\frac{d^*-\chi(r,s)}{h^2}$$

is bounded, since then

$$\sqrt{mh^{d-1}}(d^*-\chi(r,s))=\sqrt{mh^{d+3}}\cdot\frac{d^*-\chi(r,s)}{h^2}\to 0.$$

5.6 Asymptotic Normality of the Nonparametric Estimator

Let X be a random vector, which follows a generalized Pareto distribution by representation (2.5) with $U = K_s$. As in (5.5) let the set $A_{r,s}$ be given by

$$A_{r,s} = \left\{ x = (x_1, \ldots, x_d) \in K_s \,\middle|\, c = \frac{1}{x_1} + \ldots + \frac{1}{x_d} < -r \right\}.$$

Denote by $p_{r,s} := P(X \in A_{r,s})$ the probability that the random vector X falls into the set $A_{r,s}$. We have $p_{r,s} \to 0$ for $r \to \infty$. An illustration of the situation for the bivariate case can be found in Section 5.1.

Assume we have n independent copies $\tilde{X}^{(1)}, \ldots, \tilde{X}^{(n)}$ of the random vector X, and denote by $\tilde{Z}^{(i)}$ and $\tilde{C}^{(i)}$ the corresponding Pickands coordinates with regard to Fréchet margins, $i = 1, \ldots, n$. For the threshold $r > 0$ consider only those observations $\tilde{X}^{(i)}$ with $\tilde{X}^{(i)} \in A_{r,s}$. This is a random number $m = \tau_n$. We denote the resulting random vectors by $X^{(1)}, \ldots, X^{(m)}$, and $Z^{(i)}$ and $C^{(i)}$ designate the corresponding Pickands coordinates with regard to Fréchet margins, $i = 1, \ldots, m = \tau_n$. τ_n is a $B(n, p_{r,s})$ distributed random variable with expectation $np_{r,s}$ and is, by Theorem 1.4.1 in Reiss (1993, [56]), independent of $X_1, \ldots, X_m$. We assume $np_{r,s} \to \infty$ for $n \to \infty$, $r \to \infty$. Thus we also have $m \to \infty$ with probability 1, which we will show in the next lemma.

Lemma 5.6.1
Let τ_n be a $B(n, p_n)$ distributed random variable. For every $K \in \mathbb{N}_0$, we then have

$$P\left(\liminf_{n\to\infty} \tau_n > K \right) = 1$$

for $n \to \infty$, $p_n \to 0$ and $np_n \to \infty$ with the growth condition

$$\sum_{n\in\mathbb{N}} (np_n)^k \exp(-np_n) < \infty \tag{5.27}$$

for all $k \in \mathbb{N}$.

Proof:
We have

$$\begin{aligned} P\left(\liminf_{n\to\infty} \tau_n > K \right) &= 1 - P\left(\liminf_{n\to\infty} \tau_n \leq K \right) \\ &= 1 - P\left(\tau_n \leq K \text{ for infinitely many } n \right) \\ &= 1 - P\left(\limsup_{n\to\infty} A_n \right), \end{aligned}$$

with $A_n := \{\tau_n \leq K\}$. We know by the Borel-Cantelli-Lemma (see Lemma 3 in Section 6.2 of Fristedt and Grey (1997, [28])) that $P\left(\limsup_{n\to\infty} A_n\right) = 0$ if $\sum_{n=1}^{\infty} P(A_n) < \infty$. We will show that, and the assertion of the lemma follows.

$$\begin{aligned} \sum_{n=1}^{\infty} P(A_n) &= \sum_{n=1}^{\infty} B(n, p_n)\{0, , \ldots, K\} = \sum_{n=1}^{\infty} \sum_{k=0}^{K} B(n, p_n)\{k\} \\ &= \sum_{k=0}^{K} \sum_{n=1}^{\infty} B(n, p_n)\{k\}. \end{aligned}$$

The sums can be exchanged, since all summands are larger than 0. It suffices, thus, to show that $\sum_{n=1}^{\infty} B(n,p_n)\{k\} < \infty$ for every $0 \le k \le K$.

Our next aim is to show that

$$\left(1 - \frac{np_n}{n}\right)^n \le e^{-np_n} \tag{5.28}$$

for every $n \in \mathbb{N}$. By the general inequality $\log x \le x - 1$ for $x > 0$, we get

$$n \log(1 - p_n) \le n(1 - p_n - 1) = -np_n.$$

By applying the exponential function on both sides, we arrive at (5.28). Since $p_n \to_{n\to\infty} 0$, there exists a $p < 1$ and an $n_0 \in \mathbb{N}$, such that $p_n \le p$ for all $n \ge n_0$. Then we have

$$\begin{aligned}
&\sum_{n=1}^{\infty} B(n,p_n)\{k\} \\
&= \sum_{n=1}^{\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \sum_{n=1}^{\infty} \frac{n!}{k!(n-k)!} p_n^k \left(1 - \frac{np_n}{n}\right)^{n-k} \\
&\overset{(5.28)}{\le} \sum_{n=1}^{\infty} n \cdot \ldots \cdot (n-k+1) p_n^k e^{-np_n} (1-p_n)^{-k} \\
&\le \sum_{n=1}^{\infty} (np_n)^k e^{-np_n} (1-p_n)^{-k} \\
&\le \sum_{n=1}^{n_0-1} (np_n)^k e^{-np_n} (1-p_n)^{-k} + (1-p)^{-k} \sum_{n=n_0}^{\infty} (np_n)^k e^{-np_n} \\
&< \infty,
\end{aligned}$$

since the first sum is finite, and the second one exists by assumption (5.27).

□

Under the conditions $\tau_n = m$, $C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m$, we have shown in Section 5.5 that $\hat{l}_{\tau_n,r}(z)$ is asymptotically unbiased and normally distributed. In the course of this section, we will obtain unconditional results.

Lemma 5.6.2
Under $np_{r,s} \to \infty$ with growth condition (5.27), $mh^{d-1} \to \infty$, $mh^{d+3} \to 0$ and $\sqrt{mh^{d-1}}(d^* - \chi(r,s)) \to 0$ the relation

$$\begin{aligned}
&P\left(\sqrt{\tau_n h^{d-1}} \left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*} l(z)\right) \le x \,\middle|\, \tau_n = m\right) \\
&= \Phi\left(\frac{x}{\sigma}\right) + O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}\,(d^* - \chi(r,s))\right)
\end{aligned}$$

holds for $n \to \infty$, $h \to 0$ and $r \to \infty$, with σ defined as in Theorem 5.5.4.

Proof:
We note with $\mathcal{C}^{(1)} \times \ldots \times \mathcal{C}^{(m)}$ the distribution of the random variables $C^{(1)}, \ldots, C^{(m)}$ under the condition that $X^{(i)} \in A_{r,s}$, $i = 1, \ldots, m$. We then have

$$\int_{(-\infty,-r)^m} 1\left(\mathcal{C}^{(1)} \times \ldots \times \mathcal{C}^{(m)}\right)(d(r_1, \ldots, r_m)) = 1. \tag{5.29}$$

Therefore, and because of the independence of τ_n and $X_1, \ldots, X_m$ as well as the representation of a distribution by its conditional distribution (see for example Chapter V, §2 in Rényi (1966, [58])), Lemma 5.6.1 and Theorem 5.5.4, we get

$$\begin{aligned}
&P\left(\sqrt{\tau_n h^{d-1}}\left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*}l(z)\right) \leq x \middle| \tau_n = m\right) \\
&= \frac{P\left(\sqrt{\tau_n h^{d-1}}\left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*}l(z)\right) \leq x, \tau_n = m\right)}{P(\tau_n = m)} \\
&= \frac{P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - \frac{d}{d^*}l(z)\right) \leq x\right) P(\tau_n = m)}{P(\tau_n = m)} \\
&= P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - \frac{d}{d^*}l(z)\right) \leq x\right) \\
&= \int_{(-\infty,-r)^m} P\left(\sqrt{mh^{d-1}}\left(\hat{l}_{m,r}(z) - \frac{d}{d^*}l(z)\right) \leq x \middle| C^{(1)} = -r_1, \ldots, C^{(m)} = -r_m\right) \\
&\quad \left(\mathcal{C}^{(1)} \times \ldots \times \mathcal{C}^{(m)}\right)(d(r_1, \ldots, r_m)) \\
&= \int_{(-\infty,-r)^m} \Phi\left(\frac{x}{\sigma}\right) + O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}(d^* - \chi(r,s))\right) \\
&\quad \left(\mathcal{C}^{(1)} \times \ldots \times \mathcal{C}^{(m)}\right)(d(r_1, \ldots, r_m)) \\
&= \left(\Phi\left(\frac{x}{\sigma}\right) + O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}(d^* - \chi(r,s))\right)\right) \\
&\quad \cdot \int_{(-\infty,-r)^m} 1\left(\mathcal{C}^{(1)} \times \ldots \times \mathcal{C}^{(m)}\right)(d(r_1, \ldots, r_m)) \\
&\overset{(5.29)}{=} \Phi\left(\frac{x}{\sigma}\right) + O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}(d^* - \chi(r,s))\right).
\end{aligned}$$

□

In the final step we want to get rid of the condition $\tau_n = m$.

Theorem 5.6.3
Let τ_n be the number of observations with $\tilde{X}^{(i)} \in A_{r,s}$. Then we have under the growth condition (5.27) for $np_{r,s}$, $(np_{r,s} - \sqrt{n})\, h^{d-1} \to \infty$, $nh^{d+3} \to 0$, $\sqrt{nh^{d-1}}(d^* - \chi(r,s)) \to 0$ and $\Phi\left(-p_{r,s}^{-1/2}\right)\left(h^{d-1}\right)^{-1/6} \to 0$ the relation

$$\begin{aligned}
&P\left(\sqrt{\tau_n h^{d-1}}\left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*}l(z)\right) \leq x\right) \\
&= \Phi\left(\frac{x}{\sigma}\right) + O\left(h + \Phi\left(-p_{r,s}^{-1/2}\right)\left(h^{d-1}\right)^{-1/6} + \left(\left(np_{r,s} - \sqrt{n}\right)h^{d-1}\right)^{-1/6}\right. \\
&\quad \left. + \sqrt{nh^{d-1}}\,(d^* - \chi(r,s)) + \sqrt{nh^{d+3}}\right)
\end{aligned}$$

for $n \to \infty$, $h \to 0$ and $r \to \infty$ with the definition of σ from Theorem 5.5.4.

Proof:
First we note that with the convergence condition $(np_{r,s} - \sqrt{n}) \to \infty$ also $np_{r,s} \to \infty$ follows.

Before we begin with the actual proof, we need an auxiliary result on the distribution of τ_n.

Since τ_n is $B(n, p_{r,s})$ distributed, it follows from the Berry-Esséen Theorem (see Section 1.9.5 in Serfling (1980, [63])),

$$\begin{aligned}
P\left(\tau_n \le np_{r,s} - \sqrt{n}\right) &= P\left(\tau_n - np_{r,s} \le -\sqrt{n}\right) \\
&= P\left(\frac{\tau_n - np_{r,s}}{\sqrt{np_{r,s}(1-p_{r,s})}} \le -\frac{\sqrt{n}}{\sqrt{np_{r,s}(1-p_{r,s})}}\right) \\
&= P\left(\frac{\tau_n - np_{r,s}}{\sqrt{np_{r,s}(1-p_{r,s})}} \le -\frac{1}{\sqrt{p_{r,s}(1-p_{r,s})}}\right) \\
&= \Phi\left(-\frac{1}{\sqrt{p_{r,s}(1-p_{r,s})}}\right) + O\left(\frac{p_{r,s}(1-p_{r,s})\left((1-p_{r,s})^2 + p_{r,s}^2\right)}{\left(p_{r,s}(1-p_{r,s})\right)^{3/2}\sqrt{n}}\right) \\
&= \Phi\left(-\frac{1}{\sqrt{p_{r,s}(1-p_{r,s})}}\right) + O\left(\frac{1}{\sqrt{np_{r,s}}}\right) \\
&= O\left(\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right) + \frac{1}{\sqrt{np_{r,s}}}\right) \qquad (5.30)
\end{aligned}$$

for $n \to \infty$, $r \to \infty$.

In the special case $m = 0$, when no data fall over the threshold $-r$, we define $\hat{l}_{0,r}(z) = 0$. Then we have $P\left(\sqrt{\tau_n h^{d-1}}\left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*}l(z)\right) \le x | \tau_n = 0\right) = O(1)$. Thus we get with Lemma 5.6.2

$$\begin{aligned}
&P\left(\sqrt{\tau_n h^{d-1}}\left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*}l(z)\right) \le x\right) \\
&= P\left(\sqrt{\tau_n h^{d-1}}\left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*}l(z)\right) \le x \,\middle|\, \tau_n = 0\right) P(\tau_n = 0) \\
&\quad + \sum_{m=1}^{n} P\left(\sqrt{\tau_n h^{d-1}}\left(\hat{l}_{\tau_n,r}(z) - \frac{d}{d^*}l(z)\right) \le x \,\middle|\, \tau_n = m\right) P(\tau_n = m) \\
&= O(1)P(\tau_n = 0) + \sum_{m=1}^{n} \Phi\left(\frac{x}{\sigma}\right) P(\tau_n = m) \\
&\quad + \sum_{m=1}^{n} O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d+3}} + \sqrt{mh^{d-1}}\left(d^* - \chi(r,s)\right)\right) P(\tau_n = m) \\
&= \Phi\left(\frac{x}{\sigma}\right) + O(1)P(\tau_n = 0) \\
&\quad + \sum_{m=1}^{\lfloor np_{r,s} - \sqrt{n} \rfloor} O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d-1}}\left(d^* - \chi(r,s)\right) + \sqrt{mh^{d+3}}\right) P(\tau_n = m) \\
&\quad + \sum_{m=\lceil np_{r,s} - \sqrt{n} \rceil}^{n} O\left(h + \left(mh^{d-1}\right)^{-1/6} + \sqrt{mh^{d-1}}\left(d^* - \chi(r,s)\right) + \sqrt{mh^{d+3}}\right) P(\tau_n = m) \\
&= \Phi\left(\frac{x}{\sigma}\right) + O\left((1-p_{r,s})^n\right) \\
&\quad + \sum_{m=1}^{\lfloor np_{r,s} - \sqrt{n} \rfloor} O\left(h + \left(h^{d-1}\right)^{-1/6} + \sqrt{(np_{r,s} - \sqrt{n})h^{d-1}}\left(d^* - \chi(r,s)\right)\right. \\
&\quad \left. + \sqrt{(np_{r,s} - \sqrt{n})h^{d+3}}\right) \cdot P(\tau_n = m)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{m=\lceil np_{r,s}-\sqrt{n}\rceil}^{n} O\left(h+\left((np_{r,s}-\sqrt{n})h^{d-1}\right)^{-1/6}+\sqrt{nh^{d-1}}\,(d^*-\chi(r,s))+\sqrt{nh^{d+3}}\right)\\
&\quad\cdot P(\tau_n=m)\\
=\;&\Phi\left(\frac{x}{\sigma}\right)+O\left((1-p_{r,s})^n\right)\\
&+O\left(h+\left(h^{d-1}\right)^{-1/6}+\sqrt{(np_{r,s}-\sqrt{n})h^{d-1}}\,(d^*-\chi(r,s))+\sqrt{(np_{r,s}-\sqrt{n})h^{d+3}}\right)\\
&\quad\cdot P(\tau_n\le np_{r,s}-\sqrt{n})\\
&+O\left(h+\left((np_{r,s}-\sqrt{n})h^{d-1}\right)^{-1/6}+\sqrt{nh^{d-1}}\,(d^*-\chi(r,s))+\sqrt{nh^{d+3}}\right)\\
\overset{(5.30)}{=}\;&\Phi\left(\frac{x}{\sigma}\right)+O\left((1-p_{r,s})^n\right)\\
&+O\left(h+\left(h^{d-1}\right)^{-1/6}+\sqrt{(np_{r,s}-\sqrt{n})h^{d-1}}\,(d^*-\chi(r,s))+\sqrt{(np_{r,s}-\sqrt{n})h^{d+3}}\right)\\
&\quad\cdot O\left(\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)+\frac{1}{\sqrt{np_{r,s}}}\right)\\
&+O\left(h+\left((np_{r,s}-\sqrt{n})h^{d-1}\right)^{-1/6}+\sqrt{nh^{d-1}}\,(d^*-\chi(r,s))+\sqrt{nh^{d+3}}\right)\\
=\;&\Phi\left(\frac{x}{\sigma}\right)+O\left((1-p_{r,s})^n\right)\\
&+O\left(h\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)+\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)\left(h^{d-1}\right)^{-1/6}\right.\\
&\qquad+\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)\sqrt{(np_{r,s}-\sqrt{n})h^{d-1}}\,(d^*-\chi(r,s))\\
&\qquad\left.+\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)\sqrt{(np_{r,s}-\sqrt{n})h^{d+3}}\right)\\
&+O\left(\frac{h}{\sqrt{np_{r,s}}}+\left((np_{r,s})^3h^{d-1}\right)^{-1/6}+\sqrt{\left(1-\frac{1}{\sqrt{n}p_{r,s}}\right)h^{d-1}}\,(d^*-\chi(r,s))\right.\\
&\qquad\left.+\sqrt{\left(1-\frac{1}{\sqrt{n}p_{r,s}}\right)h^{d+3}}\right)\\
&+O\left(h+\left((np_{r,s}-\sqrt{n})h^{d-1}\right)^{-1/6}+\sqrt{nh^{d-1}}\,(d^*-\chi(r,s))+\sqrt{nh^{d+3}}\right)\\
=\;&\Phi\left(\frac{x}{\sigma}\right)+O\left((1-p_{r,s})^n+h\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)+\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)\left(h^{d-1}\right)^{-1/6}\right.\\
&\quad+\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)\sqrt{(np_{r,s}-\sqrt{n})h^{d-1}}\,(d^*-\chi(r,s))+\Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)\sqrt{(np_{r,s}-\sqrt{n})h^{d+3}}\\
&\quad+\frac{h}{\sqrt{np_{r,s}}}+\left((np_{r,s})^3h^{d-1}\right)^{-1/6}+\sqrt{\left(1-\frac{1}{\sqrt{n}p_{r,s}}\right)h^{d-1}}\,(d^*-\chi(r,s))\\
&\quad+\sqrt{\left(1-\frac{1}{\sqrt{n}p_{r,s}}\right)h^{d+3}}+h+\left((np_{r,s}-\sqrt{n})h^{d-1}\right)^{-1/6}\\
&\quad\left.+\sqrt{nh^{d-1}}\,(d^*-\chi(r,s))+\sqrt{nh^{d+3}}\right)
\end{aligned}
$$

$$= \Phi\left(\frac{x}{\sigma}\right) + O\left(h + \Phi\left(-\frac{1}{\sqrt{p_{r,s}}}\right)\left(h^{d-1}\right)^{-1/6} + \left((np_{r,s} - \sqrt{n})h^{d-1}\right)^{-1/6} + \sqrt{nh^{d-1}}\,(d^* - \chi(r,s)) + \sqrt{nh^{d+3}}\right)$$

for $n \to \infty$, $r \to \infty$, $h \to 0$. Note that by (5.28) we have $(1-p_{r,s})^n = \left(1 - \frac{np_{r,s}}{n}\right)^n \leq e^{-np_{r,s}} \to 0$ for $np_{r,s} \to \infty$, and since this behaves, thus, asymptotically as $e^{-np_{r,s}}$, it converges faster to 0 than $\left((np_{r,s} - \sqrt{n})h^{d-1}\right)^{-1/6}$.

□

Remark that in the proof of Theorem 5.6.3 one can replace the rate of convergence $np_{r,s} - \sqrt{n}$ by $np_{r,s} - n^a$ with $0 < a < \frac{1}{2}$ and in this manner get further relations between the rates of convergence of n, h and r, which might be useful in certain situations.

The next goal is to find some other criteria, which are easier to handle and also imply the asymptotic normality of the estimator $\hat{l}_{m,r}(z)$ at least in some cases. For this we need an auxiliary result on $p_{r,s}$.

Lemma 5.6.4
We have

$$\frac{\chi(r,s)}{r} \leq p_{r,s} = \int_{-\infty}^{-r} c^{-2}\chi(-c,s)\,dc \leq \frac{d^*}{r}.$$

Proof:
The assertion follows from

$$\begin{aligned} p_{r,s} &= P(X \in A_{r,s}) = \int_{-\infty}^{-r}\int_{Q_{-c,s}} c^{-2} l(z)\,dz\,dc = \int_{-\infty}^{-r} c^{-2}\chi(-c,s)\,dc \geq \int_{-\infty}^{-r} c^{-2}\chi(r,s)\,dc \\ &= \chi(r,s)\int_{-\infty}^{-r} c^{-2}\,dc = \chi(r,s)\left[-\frac{1}{c}\right]_{-\infty}^{-r} = \frac{\chi(r,s)}{r} \end{aligned}$$

and

$$p_{r,s} = \int_{-\infty}^{-r} c^{-2}\chi(-c,s)\,dc \leq \int_{-\infty}^{-r} c^{-2} d^*\,dc = d^*\int_{-\infty}^{-r} c^{-2}\,dc = d^*\left[-\frac{1}{c}\right]_{-\infty}^{-r} = \frac{d^*}{r}.$$

□

Corollary 5.6.5
Let τ_n be defined as in Theorem 5.6.3 and

$$\left(\frac{n}{r} - \sqrt{n}\right)h^{d-1} \to \infty, \tag{5.31}$$

$$nh^{d+3} \to 0, \tag{5.32}$$

$$\sqrt{nh^{d-1}}(d^* - \chi(r,s)) \to 0, \tag{5.33}$$

$$\Phi\left(-\sqrt{r}\right)\left(h^{d-1}\right)^{-\frac{1}{6}} \to 0, \tag{5.34}$$

$$\sum_{n=1}^{\infty}\left(\frac{n}{r}\right)^k \exp\left(-\frac{n}{r}\right) < \infty, \quad \text{where } k \in \mathbb{N}, \tag{5.35}$$

for $n \to \infty$, $r = r_n \to \infty$ and $h = h_n \to 0$. Then we have

$$\sqrt{\tau_n h^{d-1}} \left(\hat{l}_{\tau_n, r}(z) - \frac{d}{d^*} l(z) \right) \longrightarrow_{\mathcal{D}} \mathcal{N} \left(0, \frac{d^2}{d^*} l(z) \int k^2(u)\, du \right).$$

Proof:
It suffices to show that the conditions in Theorem 5.6.3 are fulfilled. We have with Lemma 5.6.4, for r large enough,

$$\left(n p_{r,s} - \sqrt{n} \right) h^{d-1} \geq \left(n \frac{\chi(r,s)}{r} - \sqrt{n} \right) h^{d-1} \geq \left(\frac{n}{r} - \sqrt{n} \right) h^{d-1} \to \infty,$$

and

$$\Phi\left(-p_{r,s}^{-1/2}\right) \left(h^{d-1}\right)^{-1/6} \leq \Phi\left(-\left(\frac{d^*}{r}\right)^{-\frac{1}{2}} \right) \left(h^{d-1}\right)^{-1/6} = \Phi\left(-\sqrt{\frac{r}{d^*}} \right) \left(h^{d-1}\right)^{-1/6} \to 0,$$

since the constant factor $\frac{1}{d^*}$ does not matter for $r \to \infty$. To show growth condition (5.27) we let $r = r_n$ depend explicitly on n. Since $r_n \to \infty$, there exists n_0 such that $\chi(r_n, s) \geq \frac{d^*}{2}$ for all $n > n_0$. Thus we get with Lemma 5.6.4

$$\begin{aligned}
\sum_{n=1}^{\infty} (n p_{r_n,s})^k e^{-n p_{r_n,s}} &\leq \sum_{n=1}^{\infty} \left(\frac{n d^*}{r_n} \right)^k e^{-\frac{n\chi(r_n,s)}{r_n}} \\
&\leq \sum_{n=1}^{n_0} \left(\frac{n d^*}{r_n} \right)^k e^{-\frac{n\chi(r_n,s)}{r_n}} + 2^k \sum_{n=n_0+1}^{\infty} \left(\frac{d^*}{2} \cdot \frac{n}{r_n} \right)^k e^{-\frac{d^*}{2} \cdot \frac{n}{r_n}} \\
&< \infty,
\end{aligned}$$

because the first sum is finite and the second one exists by (5.35), since the factor $\frac{d^*}{2}$ does not influence the convergence.

□

Remark that if any of the convergence conditions (5.31) to (5.34) is modified in such a way that only convergence to a fixed real number greater than 0 is demanded, we know by Theorem 5.6.3 that our estimator will be biased, i.e., the asymptotic normal distribution will not have mean 0. However, the effects of that bias are uncontrolled, since we do not have an explicit representation for it.

We now want to check the assumptions of Corollary 5.6.5 for the case of a bounded angular density. In the logistic case, the angular density is bounded if $\lambda \geq d$ holds, i.e., if we are closer to the case of complete dependence. This is also the situation, where our estimator worked well in Example 5.2.1.

Lemma 5.6.6
Let the angular density l of a generalized Pareto distribution be bounded, i.e., there exists an $M > 0$ with $l(z) \leq M$ for all $z \in R_{d-1}$. Then we have

$$d^* - \chi(r,s) = \int_{R_{d-1} \setminus Q_{r,s}} l(z)\, dz \leq \frac{dM}{(d-2)!} \frac{1}{rs} + o\left(\frac{1}{r}\right).$$

Proof:
By Lemma 2.1.2, we have for the volume of R_{d-1} the relation

$$\mathrm{vol}(R_{d-1}) = \int_{R_{d-1}} 1\,dz = \frac{1}{(d-1)!}.$$

Since by Lemma 5.1.2 the set $Q_{r,s}$ is a shifted simplex of edge length $1 - \frac{d}{rs}$, we also have by Lemma 2.1.2

$$\mathrm{vol}(Q_{r,s}) = \int_{Q_{r,s}} 1\,dz = \frac{\left(1-\frac{d}{rs}\right)^{d-1}}{(d-1)!}.$$

Thus, with the usage of the binomial formula,

$$\begin{aligned}
\int_{R_{d-1}\setminus Q_{r,s}} l(z)\,dz &\leq M\left(\int_{R_{d-1}} 1\,dz - \int_{Q_{r,s}} 1\,dz\right) = M\left(\frac{1}{(d-1)!} - \frac{\left(1-\frac{d}{rs}\right)^{d-1}}{(d-1)!}\right) \\
&= \frac{M}{(d-1)!}\left(1 - \left(1-\frac{d}{rs}\right)^{d-1}\right) \\
&= \frac{M}{(d-1)!}\left(1 - \sum_{i=0}^{d-1}\binom{d-1}{i}\left(-\frac{d}{rs}\right)^{d-1-i}\right) \\
&= \frac{M}{(d-1)!}\left(1 - 1 + \binom{d-1}{d-2}\frac{d}{rs} + o\left(\frac{1}{r}\right)\right) \\
&= \frac{M}{(d-1)!}\frac{(d-1)!}{(d-2)!}\frac{d}{rs} + o\left(\frac{1}{r}\right) = \frac{dM}{(d-2)!}\frac{1}{rs} + o\left(\frac{1}{r}\right).
\end{aligned}$$

□

For the case of a bounded angular density one can very easily give suitable rates of convergence.

Corollary 5.6.7
By choosing $r = n^{\frac{2}{d+3}}$ and $h = n^{-\frac{1}{d+2}}$, the convergence conditions of Corollary 5.6.5 are fulfilled for a bounded angular density.

Proof:
First we obviously have $r \to \infty$ and $h \to 0$ for $n \to \infty$. In addition,

$$\begin{aligned}
\left(\frac{n}{r} - \sqrt{n}\right)h^{d-1} &= \left(\frac{n}{n^{\frac{2}{d+3}}} - \sqrt{n}\right)n^{-\frac{d-1}{d+2}} = \left(n^{1-\frac{2}{d+3}} - n^{\frac{1}{2}}\right)n^{-\frac{d-1}{d+2}} = \left(n^{\frac{d+1}{d+3}} - n^{\frac{1}{2}}\right)n^{-\frac{d-1}{d+2}} \\
&= n^{\frac{d+1}{d+3}-\frac{d-1}{d+2}} - n^{\frac{1}{2}-\frac{d-1}{d+2}} \to \infty,
\end{aligned}$$

since

$$\frac{d+1}{d+3} > \frac{d-1}{d+2} \Leftrightarrow (d+2)(d+1) > (d-1)(d+3) \Leftrightarrow d^2+3d+2 > d^2+2d-3 \Leftrightarrow d > -5$$

and

$$\frac{d+1}{d+3} - \frac{d-1}{d+2} > \frac{1}{2} - \frac{d-1}{d+2} \Leftrightarrow d+1 > \frac{d}{2} + \frac{3}{2} \Leftrightarrow \frac{1}{2}d > \frac{1}{2} \Leftrightarrow d > 1.$$

Furthermore,

$$nh^{d+3} = n^{1-\frac{d+3}{d+2}} = n^{-\frac{1}{d+2}} \to 0,$$

and with Lemma 5.6.6,

$$\sqrt{nh^{d-1}}(d^* - \chi(r,s)) = n^{\frac{1}{2}\left(1-\frac{d-1}{d+2}\right)} \cdot O\left(\frac{1}{n^{\frac{2}{d+3}}}\right) = n^{\frac{3}{2(d+3)}} \cdot O\left(n^{-\frac{4}{2(d+3)}}\right) = O\left(n^{-\frac{1}{2(d+3)}}\right) \to 0.$$

The condition

$$\Phi\left(-\sqrt{r}\right)\left(h^{d-1}\right)^{-\frac{1}{6}} \quad \to \quad 0$$

is clear, since $\Phi(-\sqrt{r})$ converges exponentially to 0, whereas h^{d-1} converges only polynomially to 0 and, thus, $\left(h^{d-1}\right)^{-\frac{1}{6}}$ converges only polynomially to ∞.
The growth condition (5.35) is in our case

$$\sum_{n=1}^{\infty}\left(n^{\frac{d+1}{d+3}}\right)^k \exp\left(-n^{\frac{d+1}{d+3}}\right) < \infty.$$

This can be checked by the existence of the integrals $\int_0^\infty x^k e^{-x} = \Gamma(k+1)$ and the comparison of sums criterion, see Theorem 88.1 in Heuser (1998, [37]). By the substitution $y = x^{\frac{d+1}{d+3}}$ it is easily shown that

$$\int_0^\infty \left(x^{\frac{d+1}{d+3}}\right)^k \exp\left(-x^{\frac{d+1}{d+3}}\right) dx = \frac{d+3}{d+1} \cdot \Gamma\left(k + \frac{d+3}{d+1}\right) < \infty.$$

□

In the case of an unbounded angular density, the situation cannot be as easily handled. The verification of suitable convergence conditions there is, still, an open problem. One might have to use variants of Theorem 5.6.3 with $np_{r,s} - \sqrt{n}$ replaced by $np_{r,s} - n^a$, see above.

Remark 5.6.8
In the unbiased case of $d = d^*$, approximate confidence intervals of level α for $l(z)$ of the type

$$\left[\hat{l}_{m,r}(z) - \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \cdot \frac{dl(z)\int k^2(u)\,du}{\sqrt{mh^{d-1}}}, \hat{l}_{m,r}(z) + \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \cdot \frac{dl(z)\int k^2(u)\,du}{\sqrt{mh^{d-1}}}\right]$$

can be derived from Theorem 5.6.3 by using the corresponding quantiles from the standard normal distribution. Replacing the unknown angular density $l(z)$ again by its estimator, one arrives at estimated 95% confidence intervals of the type

$$\left[\hat{l}_{m,r}(z) - \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \cdot \frac{d\hat{l}_{m,r}(z)\int k^2(u)\,du}{\sqrt{mh^{d-1}}}, \hat{l}_{m,r}(z) + \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \cdot \frac{d\hat{l}_{m,r}(z)\int k^2(u)\,du}{\sqrt{mh^{d-1}}}\right]$$

for $l(z)$. But these confidence intervals have to be viewed with much care, see Powell (2002, [54]).
◊

5.7 A Supplement on the Integral of the Pickands Density

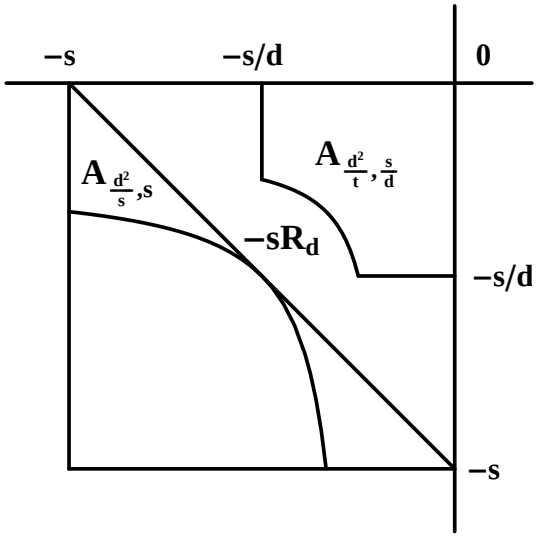
In this section we want to show the limit (3.11), used in Section 3.2.3 for the computation of runtimes, since now we have the necessary tools at our disposal. Once again we can see here the importance of the Pickands coordinates with regard to Fréchet margins, without which the proof below would not be possible. We begin with a geometric lemma.

Lemma 5.7.1
Let $d \in \mathbb{N}$, $d \geq 2$ and $s > t > 0$. Then we have

$$A_{\frac{d^2}{t},\frac{s}{d}} \subseteq -sR_d \subseteq A_{\frac{d^2}{s},s},$$

with $A_{r,s}$ defined as in (5.5).

Proof:
To illustrate the assertion of the lemma we look at the situation for the bivariate case.



We first consider the left inclusion. Let $x = (x_1, \ldots, x_d) \in A_{\frac{d^2}{t},\frac{s}{d}}$. Then $0 > x_i > -\frac{s}{d}$ for all $i = 1, \ldots, d$. Thus $0 > x_1 + \ldots + x_d > d \cdot \left(-\frac{s}{d}\right) = -s$, so $x \in -sR_d$.

For the right inclusion let $x = (x_1, \ldots, x_d) \in -sR_d$, i.e., $0 > x_i > -s$ for $i = 1, \ldots, d$ and $0 > \sum_{i=1}^{d} x_i > -s$. Then we have $x \in K_s$. To conclude that $x \in A_{\frac{d^2}{s},s}$, we still have to show $\sum_{i=1}^{d} \frac{1}{x_i} < -\frac{d^2}{s}$.

We have $x_d > -s - \sum_{i=1}^{d-1} x_i$ and, thus,

$$\sum_{i=1}^{d} \frac{1}{x_i} < \sum_{i=1}^{d-1} \frac{1}{x_i} + \frac{1}{-s - \sum_{i=1}^{d-1} x_i} =: f(x_1, \ldots, x_{d-1}).$$

We determine the maximum of the function $f : -sR_{d-1} \to \mathbb{R}$ via partial derivatives. Since the function f converges on the boundary of $-sR_{d-1}$ to $-\infty$, a unique local maximum must then be a global maximum. We have for $i = 1, \ldots, d-1$

$$\frac{\partial}{\partial x_i} f(x_1, \ldots, x_{d-1}) \quad = \quad -\frac{1}{x_i^2} + \frac{1}{(s + x_1 + \ldots + x_{d-1})^2} = 0$$

$$
\begin{aligned}
&\iff -(s + x_1 + \ldots + x_{d-1})^2 + x_i^2 = 0 \\
&\iff -\left(s + \sum_{j=1, j\neq i}^{d-1} x_j\right)^2 - 2\left(s + \sum_{j=1, j\neq i}^{d-1} x_j\right) x_i - x_i^2 + x_i^2 = 0 \\
&\iff 2\left(s + \sum_{j=1, j\neq i}^{d-1} x_j\right) x_i = -\left(s + \sum_{j=1, j\neq i}^{d-1} x_j\right)^2 \\
&\iff 2x_i = -\left(s + \sum_{j=1, j\neq i}^{d-1} x_j\right) \\
&\iff 2x_i + \sum_{j=1, j\neq i}^{d-1} x_j = -s.
\end{aligned}
$$

This leads to the system of equations

$$
Ax := \begin{pmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ 1 & 1 & 2 & & \vdots \\ \vdots & \vdots & & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{d-1} \end{pmatrix} = \begin{pmatrix} -s \\ -s \\ -s \\ \vdots \\ -s \end{pmatrix}.
$$

The matrix A is a $(d-1) \times (d-1)$ matrix with full rank, thus the system of equations is uniquely solvable. This matrix also occurred in the proof of Theorem 2.3.7. By choosing $x_1 = \ldots = x_{d-1} = -\frac{s}{d}$ we see that this is one and, thus, the only solution of the system of equations.

Therefore, $\left(-\frac{s}{d}, \ldots, -\frac{s}{d}\right)$ is the maximum of the function f. We have

$$
\begin{aligned}
f\left(-\frac{s}{d}, \ldots, -\frac{s}{d}\right) &= -\sum_{i=1}^{d-1} \frac{d}{s} + \frac{1}{-s + \sum_{i=1}^{d-1} \frac{s}{d}} = -(d-1)\frac{d}{s} + \frac{1}{-s + (d-1)\frac{s}{d}} \\
&= -\frac{d^2}{s} + \frac{d}{s} - \frac{d}{s} = -\frac{d^2}{s},
\end{aligned}
$$

thus

$$
\sum_{i=1}^{d} \frac{1}{x_i} < f(x_1, \ldots, x_{d-1}) \leq -\frac{d^2}{s},
$$

which had to be shown.

□

Theorem 5.7.2
Let W be a d–dimensional generalized Pareto distribution, T_P the transformation to Pickands coordinates, and

$$
\phi(z) = |c|^{d-1} \left(\frac{\partial^d}{\partial x_1 \cdots \partial x_d} W\right) \left(T_P^{-1}(z, c)\right)
$$

the corresponding Pickands density. Then we have for the number μ from Theorem 2.2.5

$$
\mu = \int_{R_{d-1}} \phi(z)\, dz \leq \frac{d^*}{d^2} \leq \frac{1}{d}. \tag{5.36}
$$

Suppose that W_λ, $\lambda \in \Lambda \subseteq \mathbb{R}^k$ is a parametric family of generalized Pareto distributions. We assume the existence of a $\lambda_0 \in \overline{\Lambda}$ (the closure is to be taken with regard to $\overline{\mathbb{R}^k}$ such that $\lambda_0 = \infty$ for example in the one-parametric case is possible) with

$$d^*_{\lambda_0} := \lim_{\lambda \to \lambda_0} \chi_\lambda(r,s) = \lim_{\lambda \to \lambda_0} d^*_\lambda, \tag{5.37}$$

for s small enough and any $r > \frac{d}{s}$ fixed. Then we have

$$\lim_{\lambda \to \lambda_0} \mu_\lambda = \frac{d^*_{\lambda_0}}{d^2}. \tag{5.38}$$

If $\lambda_1 \in \overline{\Lambda}$ exists with

$$\lim_{\lambda \to \lambda_1} \chi_\lambda(r,s) = 0 \tag{5.39}$$

for s small enough and any $r > \frac{d}{s}$ fixed, then we have

$$\lim_{\lambda \to \lambda_1} \mu_\lambda = 0.$$

Proof:
Let X be a d–dimensional random vector, which follows a GPD, i.e., there exists $s_0 > 0$, such that X possesses on K_{s_0} the distribution function W. Let $0 < s < s_0$, then we have with the Pickands coordinates and by Lemma 5.4.1 from Falk et al. (2004, [21]), as well as by Fubini's Theorem (see for example Fristedt and Gray (1997, [28]), Section 9.2),

$$\begin{aligned} P(X \in -sR_d) &= \int_{-sR_d} \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W(x_1, \ldots, x_d)\, d(x_1, \ldots, x_d) \\ &= \int_{T_P(-sR_d)} |c|^{d-1} \frac{\partial^d}{\partial x_1 \cdots \partial x_d} W(T_P^{-1}(z,c))\, d(z,c) \\ &= \int_{R_{d-1} \times (-s,0)} \phi(z)\, d(z,c) \\ &= \int_{-s}^{0} \int_{R_{d-1}} \phi(z)\, dz\, dc = \int_{-s}^{0} \mu\, dc = s\mu. \end{aligned} \tag{5.40}$$

Furthermore, we have by Lemma 5.6.4

$$P\left(X \in A_{\frac{d^2}{s},s}\right) \leq \frac{d^*}{\frac{d^2}{s}} = \frac{sd^*}{d^2}.$$

Since $-sR_d \subseteq A_{\frac{d^2}{s},s}$ holds by Lemma 5.7.1, $\mu \leq \frac{d^*}{d^2} \leq \frac{1}{d}$ immediately follows.

Using the lower bound from Lemma 5.6.4 for the parametric case, we get for every $t < s$

$$P\left(X \in A_{\frac{d^2}{t},\frac{s}{d}}\right) \geq \frac{\chi_\lambda\left(\frac{d^2}{t}, \frac{s}{d}\right)}{\frac{d^2}{t}}.$$

Since $t < s$, we conclude $\frac{d^2}{t} > \frac{d}{\frac{s}{d}}$ and, thus, the assumptions for the passage to the limit are fulfilled for s small enough. Together with the first inclusion in Lemma 5.7.1, we arrive at

$$\lim_{\lambda \to \lambda_0} s\mu_\lambda \geq \lim_{\lambda \to \lambda_0} \frac{\chi_\lambda\left(\frac{d^2}{t}, \frac{s}{d}\right)}{\frac{d^2}{t}} \overset{(5.37)}{=} \frac{d^*_{\lambda_0}}{\frac{d^2}{t}} = \frac{t d^*_{\lambda_0}}{d^2}.$$

From the assertion (5.36) we get

$$\lim_{\lambda\to\lambda_0} s\mu_\lambda \leq \lim_{\lambda\to\lambda_0} s\frac{d^*_\lambda}{d^2} \overset{(5.37)}{=} s\frac{d^*_{\lambda_0}}{d^2}.$$

If now t converges to s, we conclude

$$\frac{d^*_{\lambda_0}}{d^2} \geq \lim_{\lambda\to\lambda_0} \mu_\lambda \geq \frac{d^*_{\lambda_0}}{d^2}$$

and, thus, the assertion (5.38) holds. Finally, we have again with Lemmata 5.7.1 and 5.6.4

$$\begin{aligned}\lim_{\lambda\to\lambda_1} \mu_\lambda &= \lim_{\lambda\to\lambda_1} \frac{s\mu_\lambda}{s} \overset{(5.40)}{=} \lim_{\lambda\to\lambda_1} \frac{P(X \in -sR_d)}{s} \leq \lim_{\lambda\to\lambda_1} \frac{P\left(X \in A_{\frac{d^2}{s},s}\right)}{s} \\ &\overset{5.6.4}{=} \frac{1}{s} \lim_{\lambda\to\lambda_1} \int_{-\infty}^{-r} c^{-2}\chi_\lambda(c,s)\,dc = \frac{1}{s}\int_{-\infty}^{-r} c^{-2} \lim_{\lambda\to\lambda_1} \chi_\lambda(c,s)\,dc \overset{(5.39)}{=} 0,\end{aligned}$$

where the integral and the limit can be exchanged due to the dominated convergence theorem (see for example Fristedt and Gray (1997, [28]), Section 8.2), since $\chi_\lambda(c,s)$ is bounded by $d^*_\lambda \leq d$, and c^{-2} is integrable over $(-\infty,-r)$.

□

Remark 5.7.3
In the logistic case, the conditions of Theorem 5.7.2 are fulfilled with $\lambda_0 = \infty$ and $\lambda_1 = 1$, the cases of complete dependence and independence and $d^* = d$ for $\lambda > 1$, thus $\lim_{\lambda\to\infty} \mu_\lambda = \frac{1}{d}$ follows and the limit (3.11) holds.

◊

Chapter 6

Parametric Estimation in Generalized Pareto Models

In Chapters 4 and 5 we used nonparametric methods to estimate the angular density. In this chapter we want to estimate it, when we assume a parametric model underlying the data. The estimation of the angular density then reduces to the estimation of the model parameters. In Section 6.1, we present two maximum likelihood (ML) methods based on the angular density to achieve this goal. In Section 6.2, we will also introduce a maximum likelihood method but use the Pickands density instead of the angular density. Relative frequencies will be used in Section 6.3 for another estimation procedure. For all procedures, the asymptotic behavior is studied. Some combined estimation procedures will be given in Section 6.4, and all procedures will be compared extensively with the help of simulated data for the logistic case.

In the literature, the maximum likelihood method is quite popular for parametric estimation in extreme value models, see for example Section 8.2 of Coles (2001, [8]), Coles and Tawn (1991, [10]), Coles and Tawn (1994, [11]) or Coles et al. (1999, [9]). In Section 9.3 of Reiss and Thomas (2001, [57]), two other methods are mentioned, one is the method of moments, the other is a Pickands estimator in a Marshall-Olkin model. In Paper B, Section 8, of Tajvidi (1996, [71]), the ML method is used to estimate in bivariate GPD models, which is by our knowledge the only source of parametric estimation for GPDs. The estimation via relative frequencies in Section 6.3 follows an idea of Falk (1998, [20]).

Possible parametric models are given in Section 2.3. Further models, described for the extreme value case, can be found in Section 3.5 of Kotz and Nadarajah (2000, [47]).

6.1 Maximum Likelihood Estimation with the Angular Density

For general information on the maximum likelihood method we refer to Section 2.6 of Coles (2001, [8]) and Section 4.2 of Serfling (1980, [63]).

In this chapter, we assume that we have n independent copies $\tilde{X}^{(1)}, \ldots, \tilde{X}^{(n)}$ of a random vector X, which follows a generalized Pareto distribution $W_{\lambda_1,\ldots,\lambda_k}$ from a k–parametric family with $U = K_s$ by representation (2.5). Let $W_{\lambda_1,\ldots,\lambda_k}$ have the angular density $l_{\lambda_1,\ldots,\lambda_k}$, and suppose that $d^*_{\lambda_1,\ldots,\lambda_k} > 0$. An example is the logistic distribution with $k = 1$. To keep the notation as simple as possible we set $\lambda := (\lambda_1, \ldots, \lambda_k)$, where $\lambda \in \Lambda \subseteq \mathbb{R}^k$.

We assume, furthermore, that for the copies $||\tilde{X}^{(i)}||_\infty < s$ holds, and denote again by $\tilde{Z}^{(i)}$ and $\tilde{C}^{(i)}$ the corresponding Pickands coordinates with regard to Fréchet margins, $i = 1, \ldots, n$. Again we choose the threshold $r > 0$ and consider only those observations $\tilde{X}^{(i)}$ with $\tilde{C}^{(i)} < -r$. We denote these by $X^{(1)}, \ldots, X^{(m)}$. They are independent from each other and from the random number $m = \tau_n$ (see Sections 4.1 and 5.2).

By Theorem 5.1.7 they have a density, which is not exactly known but close to $\frac{l_\lambda(z)}{d^*_\lambda}$. This is a suitable approach for an (asymptotic) maximum likelihood estimation of $\lambda = (\lambda_1, \ldots, \lambda_k)$ by choosing $\hat{\lambda}_{m,r}$ such that the expression

$$\prod_{i=1}^{m} \frac{l_\lambda\left(Z^{(i)}\right)}{d^*_\lambda}$$

is maximized in λ. This is equivalent to maximizing the term

$$\Upsilon(\lambda) := \log\left(\prod_{i=1}^{m} \frac{l_\lambda\left(Z^{(i)}\right)}{d^*_\lambda}\right) = \sum_{i=1}^{m} \log\left(\frac{l_\lambda\left(Z^{(i)}\right)}{d^*_\lambda}\right)$$

in $\lambda = (\lambda_1, \ldots, \lambda_k)$. Theoretically one must only compute the gradient $\nabla_\lambda \Upsilon(\lambda)$ of this term, set it to 0 and solve this equation in $\lambda_1, \ldots, \lambda_k$. Since, however, in most cases the expression $\frac{l_\lambda}{d^*_\lambda}$ is quite complicated in λ, one can hardly proceed in this manner. Instead, one must try to maximize this expression numerically. Therefore, one needs a method of maximizing a function without using the derivatives. The simplest method is the algorithm by Nelder-Mead, see Section 8.1 in Kelley (1999, [46]). Most common software packages like MATHEMATICA or MATLAB have implemented algorithms based on Nelder-Mead for the numerical computation of a maximum, which will be used in this manuscript. They all have in common that they cannot guarantee the convergence to a global maximum but only convergence to a local maximum.

These procedures have to evaluate the function, which is to be maximized, at a lot of points. Therefore, this evaluation should be as efficient as possible. An analytical expression of d^*_λ is known in most common parametric models like the logistic or asymmetric logistic model (see Section 2.3). If this is not the case, one has to rely on numerical methods for the evaluation of d^*_λ, since this is the integral of the angular density. These numerical integration methods would make the computational effort much higher and, thus, should be avoided if possible.

The asymptotic consistency, normality and efficiency of maximum likelihood estimators are well known, see for example Section 4.2 in Serfling (1980, [63]), Section 33.2 in Cramer (1963, [12]), Section 2.5 in Witting and Nölle (1970, [75]) or Section 7.5 of Pfanzagl (1994, [52]). However, we did not choose the exact procedure of the maximum likelihood estimation method, since we do not insert the observations into their density, but into a function, which is only close to their density (see Theorem 5.1.7). Therefore, we refer to this method as the asymptotic maximum likelihood (Asymptotic ML) method, and the asymptotic behavior of the above defined Asymptotic ML estimator $\hat{\lambda}_{m,r}$, which maximizes $\Upsilon(\lambda)$, must be examined more closely. Thereby we proceed analogously to the references above. First we begin with a lemma.

Lemma 6.1.1
Let f_n, $n \in \mathbb{N}$ be a sequence of k–dimensional random vectors with the property

$$\sqrt{n} f_n \longrightarrow_{\mathcal{D}} \mathcal{N}(0, \Pi)$$

with a $k \times k$ matrix Π. Let, in addition, A_n, $n \in \mathbb{N}$, be a sequence of $k \times k$ matrices with $A_n \to_{n\to\infty} A$ componentwise to a $k \times k$ matrix A. Then we have

$$\sqrt{n} A_n f_n \longrightarrow_{\mathcal{D}} \mathcal{N}(0, A\Pi A^T).$$

Proof:
We have (see for example Falk et al. (2002, [23]), Theorem 3.3.7)

$$\sqrt{n}Af_n \longrightarrow_{\mathcal{D}} \mathcal{N}(0, A\Pi A^T).$$

Thus

$$\sqrt{n}A_nf_n = \sqrt{n}A_nf_n - \sqrt{n}Af_n + \sqrt{n}Af_n = \underbrace{\sqrt{n}(A_n - A)f_n}_{\to 0 \text{ for } n\to\infty} + \sqrt{n}Af_n \longrightarrow_{\mathcal{D}} \mathcal{N}(0, A\Pi A^T),$$

as asserted.

□

Now we come to a first main result of this section.

Theorem 6.1.2
Let $s > 0$ be fixed and $r > 0$ such that $Q_{r,s} \neq \emptyset$. Let $X_1, \ldots, X_m$ be independent random variables, distributed according to a generalized Pareto distribution W_λ with $d_\lambda^* > 0$, and with $C^{(i)} < -r$, $X_i \in K_s$. Let the function

$$\varrho(\lambda, z) := \frac{l_\lambda(z)}{d_\lambda^*}$$

be three times continuously differentiable in λ, where $l_\lambda(z)$ is the angular density. In addition, we assume that for every $\tilde{\theta} \in \Lambda$ there exist functions $g(z)$, $h(z)$ and $H(z) > 0$, possibly depending on $\tilde{\theta}$, such that the relations

$$\left|\frac{\partial}{\partial\theta_j}\varrho(\theta, z)\right| \leq g(z), \quad \left|\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}}\varrho(\theta, z)\right| \leq h(z), \quad \left|\frac{\partial^3}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}}\log\left(\varrho(\theta, z)\right)\right| \leq H(z), \tag{6.1}$$

hold for θ from a neighborhood $N(\tilde{\theta})$ of $\tilde{\theta}$, for $j, j_1, j_2, j_3 = 1, \ldots, k$, and for all $z \in R_{d-1}$. Next we demand that

$$\begin{aligned}
&\int_{R_{d-1}} g(z)\,dz < \infty, && \int_{R_{d-1}} h(z)\,dz < \infty,\\
&\int_{R_{d-1}} H(z)\,dz < \infty, && \int_{R_{d-1}} H(z)\varrho(\theta, z)\,dz < \infty \text{ for } \theta \in N(\tilde{\theta}),\\
&\int_{R_{d-1}} H^2(z)\,dz < \infty, && \int_{R_{d-1}} H^2(z)\varrho(\theta, z)\,dz < \infty \text{ for } \theta \in N(\tilde{\theta}).
\end{aligned} \tag{6.2}$$

Furthermore, we assume that for $\lambda \in \Lambda$ and $j, j_1, j_2 = 1, \ldots, k$, the integrals

$$0 < v_{\lambda,j_1,j_2} := \int_{R_{d-1}} \frac{1}{\varrho(\lambda, z)}\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda, z)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda, z)\right) dz < \infty, \tag{6.3}$$

$$\int_{R_{d-1}} \frac{1}{\varrho(\lambda, z)}\left|\frac{\partial}{\partial\lambda_j}\varrho(\lambda, z)\right| dz < \infty, \tag{6.4}$$

$$\begin{aligned}
&\int_{R_{d-1}} \frac{1}{\varrho(\lambda, z)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho(\lambda, z)\right)\\
&\qquad - \left(\frac{1}{\varrho(\lambda, z)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda, z)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda, z)\right) dz < \infty,
\end{aligned} \tag{6.5}$$

$$u_{\lambda,j_1,j_2} := \int_{R_{d-1}} \left(\frac{1}{\varrho(\lambda,z)}\right)^2 \left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda,z)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda,z)\right) dz < \infty, \quad (6.6)$$

$$\int_{R_{d-1}} \left(\frac{1}{\varrho(\lambda,z)}\right)\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho(\lambda,z)\right)^2 + \left(\frac{1}{\varrho(\lambda,z)}\right)^3 \left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda,z)\right)^2\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda,z)\right)^2 - 2\left(\frac{1}{\varrho(\lambda,z)}\right)^2 \left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda,z)\right)\cdot\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda,z)\right)\cdot\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho(\lambda,z)\right) dz < \infty, \quad (6.7)$$

$$\int_{R_{d-1}} \left(\frac{1}{\varrho(\lambda,z)}\right)^2\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho(\lambda,z)\right)^2 + \left(\frac{1}{\varrho(\lambda,z)}\right)^4 \left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda,z)\right)^2\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda,z)\right)^2 - 2\left(\frac{1}{\varrho(\lambda,z)}\right)^3 \cdot\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho(\lambda,z)\right)\cdot\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda,z)\right)\cdot\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda,z)\right) dz < \infty, \quad (6.8)$$

$$\int_{R_{d-1}} \left(\frac{1}{\varrho(\lambda,z)}\right)^2 \cdot\left|\frac{\partial}{\partial\lambda_j}\varrho(\lambda,z)\right|^3 dz < \infty, \quad (6.9)$$

$$\int_{R_{d-1}} \left(\frac{1}{\varrho(\lambda,z)}\right)^3 \cdot\left|\frac{\partial}{\partial\lambda_j}\varrho(\lambda,z)\right|^3 dz < \infty, \quad (6.10)$$

exist. These regularity conditions (6.1) to (6.10) make sure that certain moments of diverse random variables in the proof exist and that certain limits and integrals can be exchanged.

Let the matrix

$$V_\lambda := \left(v_{\lambda,j_1,j_2}\right)_{j_1,j_2=1,\ldots,k}$$

be invertible. Then there exists a sequence of the maximum likelihood estimators $\hat{\lambda}_{m,r}$, which solve $\nabla_\theta \Upsilon(\theta) = 0$, and for whom the convergence

$$\sqrt{m}\left(\hat{\lambda}_{m,r} - \lambda\right) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, V_\lambda^{-1}\right)$$

holds for $m \to \infty$ and $r \to \infty$ if

$$\sqrt{m}\,\left(d_\lambda^* - \chi_\lambda(r,s)\right) \to_{m\to\infty, r\to\infty} 0. \quad (6.11)$$

The matrix V_λ is the Fisher information matrix. Thus, the estimation is asymptotically efficient in the sense that it possesses for $m \to \infty$, $r \to \infty$ the inverse of the Fisher information matrix from the information inequality (see Section 4.1.3 in Serfling (1980, [63])) as covariance matrix.

Proof:
Because of the differentiability of $\varrho(\theta, z)$ with regard to θ, there exists a Taylor series expansion for the gradient ∇_θ of the function $\log(\varrho(\theta, z))$ in a neighborhood $N(\lambda)$ of λ, the underlying parameter. We denote it as

$$\nabla_\theta \log(\varrho(\theta,z)) = \nabla_\theta \log(\varrho(\theta,z))|_{\theta=\lambda} + \left(\mathrm{He}_\theta \log(\varrho(\theta,z))|_{\theta=\lambda} + R(z,\theta,\lambda)\right)(\theta-\lambda),$$

where He_θ is the Hessian matrix, $R(z, \theta, \lambda)$ is a matrix with

$$||R(z,\theta,\lambda)||_\infty \le H(z)||\theta-\lambda||_\infty,$$

and H is the function from condition (6.1), see Theorems 168.4 and 168.5 in Heuser (1998, [38]). We put

$$\begin{aligned} A_m &:= \frac{1}{m}\sum_{i=1}^{m} \nabla_\theta \log\left(\varrho\left(\theta, Z^{(i)}\right)\right)\Big|_{\theta=\lambda}, \\ B_m &:= \frac{1}{m}\sum_{i=1}^{m} \mathrm{He}_\theta \log\left(\varrho\left(\theta, Z^{(i)}\right)\right)\Big|_{\theta=\lambda}, \\ C_m &:= \frac{1}{m}\sum_{i=1}^{m} H\left(Z^{(i)}\right). \end{aligned}$$

Then we have

$$\frac{1}{m}\nabla_\theta \Upsilon(\theta) = A_m + \left(B_m + \frac{1}{m}\sum_{i=1}^{m} R\left(Z^{(i)}, \theta, \lambda\right)\right)(\theta - \lambda), \tag{6.12}$$

where

$$\left|\left|\frac{1}{m}\sum_{i=1}^{m} R\left(Z^{(i)}, \theta, \lambda\right)\right|\right|_\infty \leq |C_m| \cdot ||\theta - \lambda||_\infty. \tag{6.13}$$

Due to conditions (6.1) and (6.2), differentiation and integration can be interchanged, and we get for $j = 1, \ldots, k$

$$\int_{R_{d-1}} \frac{\partial}{\partial\theta_j}\varrho(\theta, z)\, dz = \frac{\partial}{\partial\theta_j}\int_{R_{d-1}} \varrho(\theta, z)\, dz = \frac{\partial}{\partial\theta_j}\int_{R_{d-1}} \frac{l_\theta(z)}{d_\theta^*}\, dz \overset{(2.11)}{=} \frac{\partial}{\partial\theta_j}(1) = 0, \tag{6.14}$$

and equally for $j_1, j_2 = 1, \ldots, k$

$$\int_{R_{d-1}} \frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}}\varrho(\theta, z)\, dz = 0. \tag{6.15}$$

Theorem 5.1.7 implies that the $Z^{(i)}$ possess a density, which can be expanded to

$$\frac{l_\lambda(z)}{d_\lambda^*} + O\left(d_\lambda^* - \chi_\lambda(r,s)\right) = \varrho(\lambda, z) + O\left(d_\lambda^* - \chi_\lambda(r,s)\right).$$

With that we have for $j = 1, \ldots, k$ and $Z = Z^{(i)}$

$$\begin{aligned} E_\lambda\left(\frac{\partial \log\left(\varrho\left(\lambda, Z\right)\right)}{\partial\lambda_j}\right) &= \int_{R_{d-1}} \frac{1}{\varrho\left(\lambda, z\right)}\left(\frac{\partial}{\partial\lambda_j}\varrho\left(\lambda, z\right)\right)\left(\varrho(\lambda, z) + O\left(d_\lambda^* - \chi_\lambda(r,s)\right)\right) dz \\ &= \int_{R_{d-1}} \frac{1}{\varrho\left(\lambda, z\right)}\left(\frac{\partial}{\partial\lambda_j}\varrho\left(\lambda, z\right)\right)\varrho(\lambda, z)\, dz \\ &\quad + \int_{R_{d-1}} \frac{1}{\varrho\left(\lambda, z\right)}\left(\frac{\partial}{\partial\lambda_j}\varrho\left(\lambda, z\right)\right) O\left(d_\lambda^* - \chi_\lambda(r,s)\right) dz \\ &\overset{(6.14)}{=} \left(\int_{R_{d-1}} \frac{1}{\varrho\left(\lambda, z\right)}\left(\frac{\partial}{\partial\lambda_j}\varrho\left(\lambda, z\right)\right) dz\right) O\left(d_\lambda^* - \chi_\lambda(r,s)\right) \\ &=: t_j(\lambda, r). \end{aligned}$$

Thereby $\lim_{r\to\infty} t_j(\lambda, r) = 0$ holds due to (6.4). Furthermore, we have for $j_1, j_2 = 1, \ldots, k$

$$
\begin{aligned}
&E_\lambda\left(\frac{\partial^2 \log\left(\varrho\left(\lambda, Z\right)\right)}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\right) = \\
&= \int_{R_{d-1}} \left(\frac{1}{\varrho\left(\lambda,z\right)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right) - \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\right) \\
&\quad \left(\varrho(\lambda,z) + O\left(d_\lambda^* - \chi_\lambda(r,s)\right)\right) dz \\
&= \int_{R_{d-1}} \frac{1}{\varrho\left(\lambda,z\right)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\varrho(\lambda,z)\, dz \\
&\quad + \int_{R_{d-1}} \frac{1}{\varrho\left(\lambda,z\right)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right) O\left(d_\lambda^* - \chi_\lambda(r,s)\right) dz \\
&\quad - \int_{R_{d-1}} \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\left(\varrho(\lambda,z) + O\left(d_\lambda^* - \chi_\lambda(r,s)\right)\right) dz \\
&\overset{(6.15)}{=} O\left(d_\lambda^* - \chi_\lambda(r,s)\right)\int_{R_{d-1}} \frac{1}{\varrho\left(\lambda,z\right)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\lambda_{j_2}}\varrho\left(\lambda,z\right)\right) \\
&\quad - \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right) dz \\
&\quad - \int_{R_{d-1}} \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\varrho(\lambda,z)\, dz \\
&\overset{(6.3)}{=:} \tilde{t}_{j_1,j_2}(\lambda,r) - v_{\lambda,j_1,j_2}.
\end{aligned}
$$

Due to (6.5), we have $\lim_{r\to\infty} \tilde{t}_{j_1,j_2}(\lambda, r) = 0$. Analogously,

$$
\begin{aligned}
&E_\lambda\left(\frac{\partial \log\left(\varrho\left(\lambda, Z\right)\right)}{\partial\lambda_{j_1}} \cdot \frac{\partial \log\left(\varrho\left(\lambda, Z\right)\right)}{\partial\lambda_{j_2}}\right) = \\
&= \int_{R_{d-1}} \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\left(\varrho(\lambda,z) + O\left(d_\lambda^* - \chi_\lambda(r,s)\right)\right) dz \\
&= \int_{R_{d-1}} \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right) dz \\
&\quad + O\left(d_\lambda^* - \chi_\lambda(r,s)\right)\int_{R_{d-1}} \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right) dz \\
&\overset{(6.3),(6.6)}{=} v_{\lambda,j_1,j_2} - O\left(d_\lambda^* - \chi_\lambda(r,s)\right) u_{\lambda,j_1,j_2}. \qquad (6.16)
\end{aligned}
$$

Next we consider

$$
\begin{aligned}
&E_\lambda\left(\left(\frac{\partial^2 \log\left(\varrho\left(\lambda, Z\right)\right)}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\right)^2\right) = \\
&= \int_{R_{d-1}} \left(\frac{1}{\varrho\left(\lambda,z\right)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right) - \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\right)^2 \\
&\quad \left(\varrho(\lambda,z) + O\left(d_\lambda^* - \chi_\lambda(r,s)\right)\right) dz \\
&= \int_{R_{d-1}} \frac{1}{\varrho\left(\lambda,z\right)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)^2 + \left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^3\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)^2\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&-2\left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\,dz\\
&+O\left(d_\lambda^*-\chi_\lambda(r,s)\right)\int_{R_{d-1}}\left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)^2\\
&+\left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^4\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)^2\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)^2\\
&-2\left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^3\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_1}}\varrho\left(\lambda,z\right)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho\left(\lambda,z\right)\right)\,dz\\
=:\;& w_{\lambda,j_1,j_2}+q'_{j_1,j_2}(\lambda,r).
\end{aligned}
$$

The corresponding integrals exist by (6.7) and (6.8) and, thus, $\lim_{r\to\infty} q'_{j_1,j_2}(\lambda,r)=0$. Finally we have

$$
\begin{aligned}
&E_\lambda\left(\left|\frac{\partial\log\left(\varrho\left(\lambda,Z\right)\right)}{\partial\lambda_j}\right|^3\right)=\\
=\;&\int_{R_{d-1}}\left(\frac{1}{\varrho\left(\lambda,z\right)}\left|\frac{\partial}{\partial\lambda_j}\varrho\left(\lambda,z\right)\right|\right)^3\left(\varrho(\lambda,z)+O\left(d_\lambda^*-\chi_\lambda(r,s)\right)\right)\,dz\\
=\;&\int_{R_{d-1}}\left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^2\left|\frac{\partial}{\partial\lambda_j}\varrho\left(\lambda,z\right)\right|^3\,dz+O\left(d_\lambda^*-\chi_\lambda(r,s)\right)\int_{R_{d-1}}\left(\frac{1}{\varrho\left(\lambda,z\right)}\right)^3\left|\frac{\partial}{\partial\lambda_j}\varrho\left(\lambda,z\right)\right|^3\,dz\\
=:\;& c_{\lambda,j}+\tilde{q}_j(\lambda,r). \qquad (6.17)
\end{aligned}
$$

The corresponding integrals exist by assumptions (6.9) and (6.10) and, thus, $\lim_{r\to\infty}\tilde{q}_j(\lambda,r)=0$.

From the above considerations we conclude that

(i) A_m is an arithmetic mean of independent, identically distributed random vectors with expectation $t(\lambda,r):=(t_1(\lambda,r),\ldots,t_k(\lambda,r))$ and covariance matrix

$$
\begin{aligned}
&\left(v_{\lambda,j_1,j_2}+O(d_\lambda^*-\chi_\lambda(r,s))u_{\lambda,j_1,j_2}-t_{j_1}(\lambda,r)t_{j_2}(\lambda,r)\right)_{j_1,j_2=1,\ldots,k}\\
=:\;& V_\lambda+O(d_\lambda^*-\chi_\lambda(r,s))U_\lambda+t(\lambda,r)t^T(\lambda,r).
\end{aligned}
$$

(ii) B_m is an arithmetic mean of independent, identically distributed random vectors (noted as a matrix) with expectation matrix $\left(-v_{\lambda,j_1,j_2}+\tilde{t}_{j_1,j_2}(\lambda,r)\right)_{j_1,j_2=1,\ldots,k}=:-V_\lambda+\tilde{t}(\lambda,r)$. The components of the random vectors underlying B_m have the variance

$$
w_{\lambda,j_1,j_2}+q'_{j_1,j_2}(\lambda,r)-\left(\tilde{t}_{j_1,j_2}(\lambda,r)-v_{\lambda,j_1,j_2}\right)^2\to_{r\to\infty} w_{\lambda,j_1,j_2}-v^2_{\lambda,j_1,j_2}.
$$

(iii) C_m is an arithmetic mean of independent, identically distributed random variables with expectation

$$
E_\lambda(H(Z))=\underbrace{\int_{R_{d-1}}H(z)\varrho(\lambda,z)\,dz}_{=:F_\lambda}+\underbrace{O(d_\lambda^*-\chi_\lambda(r,s))\int_{R_{d-1}}H(z)\,dz}_{=:t'(\lambda,r)}<\infty
$$

due to (6.2) and $\lim_{r\to\infty}t'(\lambda,r)=0$. The variance is

$$
\begin{aligned}
&E_\lambda(H^2(Z))-(E_\lambda(H(Z)))^2\\
=\;&\underbrace{\int_{R_{d-1}}H^2(z)\varrho(\lambda,z)\,dz}_{=:L_\lambda}+\underbrace{O(d_\lambda^*-\chi_\lambda(r,s))\int_{R_{d-1}}H^2(z)\,dz}_{=:q(\lambda,r)}-\left(F_\lambda+t'(\lambda,r)\right)^2.
\end{aligned}
$$

Because of (6.2), the variance exists and converges to $L_\lambda - F_\lambda^2$ for $r \to \infty$.

From the strong law of large numbers (see, for example, Serfling (1980, [63]), Theorem 1.8B), we conclude with probability 1

$$A_m \to_{m\to\infty} t(\lambda, r), \qquad B_m \to_{m\to\infty} -V_\lambda + \tilde{t}(\lambda, r), \qquad C_m \to_{m\to\infty} F_\lambda + t'(\lambda, r). \tag{6.18}$$

Furthermore, we have the convergences

$$A_m \to_{m\to\infty, r\to\infty} 0, \qquad B_m \to_{m\to\infty, r\to\infty} -V_\lambda, \qquad C_m \to_{m\to\infty, r\to\infty} F_\lambda \tag{6.19}$$

in probability, as we will show in the following. For $\varepsilon > 0$ we get with the help of the inequality by Tchebychev, see Corollary 1.18.3 in Gänssler and Stute (1977, [30]), where $A_{m,j}$ denotes the jth component of A_m,

$$\begin{aligned}
P\left(||A_m||_\infty \geq \varepsilon\right) &= P\left(||A_m - t(\lambda,r) + t(\lambda,r)||_\infty \geq \varepsilon\right) \\
&\leq P\left(||A_m - t(\lambda,r)||_\infty + ||t(\lambda,r)||_\infty \geq \varepsilon\right) \\
&= P\left(||A_m - t(\lambda,r)||_\infty \geq \varepsilon - ||t(\lambda,r)||_\infty\right) \\
&= P\left(\bigcup_{j=1}^{k} \left\{|A_{m,j} - t_j(\lambda,r)| \geq \varepsilon - ||t(\lambda,r)||_\infty\right\}\right) \\
&\leq \sum_{j=1}^{k} P\left(|A_{m,j} - t_j(\lambda,r)| \geq \varepsilon - ||t(\lambda,r)||_\infty\right) \\
&\leq \sum_{j=1}^{k} \frac{\mathrm{Var}(A_{m,j})}{\left(\varepsilon - ||t(\lambda,r)||_\infty\right)^2} \\
&\overset{(i)}{=} \frac{1}{m} \sum_{j=1}^{k} \frac{v_{\lambda,j,j} + O(d_\lambda^* - \chi_\lambda(r,s))u_{\lambda,j,j} - t_j^2(\lambda,r)}{\left(\varepsilon - ||t(\lambda,r)||_\infty\right)^2} \\
&\to_{m\to\infty, r\to\infty} 0.
\end{aligned}$$

Since the denominator does not vanish for r large enough due to $t(\lambda, r) \to_{r\to\infty} 0$, the sum converges for $r \to \infty$ to a fixed value, and with the factor $\frac{1}{m}$ we get the convergence to 0. Thus, we have shown the first convergence of (6.19).

Analogously we show the assertion for B_m, where B_{m,j_1,j_2} denotes the (j_1, j_2)th component of the random matrix B_m.

$$\begin{aligned}
P\left(||B_m + V_\lambda||_\infty \geq \varepsilon\right) &= P\left(||B_m + V_\lambda - \tilde{t}(\lambda,r) + \tilde{t}(\lambda,r)||_\infty \geq \varepsilon\right) \\
&\leq P\left(||B_m + V_\lambda - \tilde{t}(\lambda,r)||_\infty + ||\tilde{t}(\lambda,r)||_\infty \geq \varepsilon\right) \\
&= P\left(||B_m + V_\lambda - \tilde{t}(\lambda,r)||_\infty \geq \varepsilon - ||\tilde{t}(\lambda,r)||_\infty\right) \\
&= P\left(\bigcup_{j_1,j_2=1}^{k} \left\{|B_{m,j_1,j_2} + v_{\lambda,j_1,j_2} - \tilde{t}_{j_1,j_2}(\lambda,r)| \geq \varepsilon - ||\tilde{t}(\lambda,r)||_\infty\right\}\right) \\
&\leq \sum_{j_1,j_2=1}^{k} P\left(|B_{m,j_1,j_2} + v_{\lambda,j_1,j_2} - \tilde{t}_{j_1,j_2}(\lambda,r)| \geq \varepsilon - ||\tilde{t}(\lambda,r)||_\infty\right) \\
&\leq \sum_{j_1,j_2=1}^{k} \frac{\mathrm{Var}(B_{m,j_1,j_2})}{\left(\varepsilon - ||\tilde{t}(\lambda,r)||_\infty\right)^2}
\end{aligned}$$

$$\begin{aligned} &\overset{(ii)}{=} && \frac{1}{m}\sum_{j_1,j_2=1}^{k} \frac{w_{\lambda,j_1,j_2} + q'_{j_1,j_2}(\lambda,r) - \left(\tilde{t}_{j_1,j_2}(\lambda,r) - v_{\lambda,j_1,j_2}\right)^2}{\left(\varepsilon - ||\tilde{t}(\lambda,r)||_\infty\right)^2} \\ &\to_{m\to\infty, r\to\infty} && 0. \end{aligned}$$

Likewise we conclude for C_m

$$\begin{aligned} P\left(|C_m - F_\lambda| \geq \varepsilon\right) &= P\left(|C_m - F_\lambda - t'(\lambda,r) + t'(\lambda,r)| \geq \varepsilon\right) \\ &\leq P\left(|C_m - F_\lambda - t'(\lambda,r)| + |t'(\lambda,r)| \geq \varepsilon\right) \\ &= P\left(|C_m - F_\lambda - t'(\lambda,r)| \geq \varepsilon - |t'(\lambda,r)|\right) \\ &\leq \frac{\mathrm{Var}(C_m)}{\left(\varepsilon - |t'(\lambda,r)|\right)^2} \\ &\overset{(iii)}{=} \frac{1}{m} \cdot \frac{L_\lambda + q(\lambda,r) - \left(F_\lambda + t'(\lambda,r)\right)^2}{\left(\varepsilon - |t'(\lambda,r)|\right)^2} \\ &\to_{m\to\infty, r\to\infty} 0. \end{aligned}$$

The aim is now to show that

$$\sqrt{m}A_m \longrightarrow_{\mathcal{D}} \mathcal{N}(0, V_\lambda) \tag{6.20}$$

holds. For this purpose we decompose

$$\sqrt{m}A_m = \sqrt{m}(A_m - t(\lambda,r)) + \sqrt{m}t(\lambda,r).$$

The second term converges by condition (6.11) for $m \to \infty$, $r \to \infty$ to 0. It remains to show that

$$\sqrt{m}(A_m - t(\lambda,r)) \longrightarrow_{\mathcal{D}} \mathcal{N}(0, V_\lambda).$$

For that we use a version of the central limit theorem by Lindeberg for the multivariate case, as described in Corollary 18.3 of Bhattacharya and Rao (1976, [4]). For reasons of clarity, we let $r = r_m$ depend on m. We put

$$X_{i,m} := \nabla_\lambda \log\left(\varrho\left(\lambda, Z^{(i)}\right)\right) - t(\lambda, r_m).$$

Then we have

$$\sqrt{m}(A_m - t(\lambda, r_m)) = \frac{1}{\sqrt{m}}\sum_{i=1}^{m} X_{i,m}.$$

We set

$$T_m := \left(\frac{1}{m}\sum_{i=1}^{m} \mathrm{Cov}(X_{i,m})\right)^{-1/2}.$$

From the multivariate central limit theorem by Lindeberg, we conclude the asymptotic normality if

$$\frac{1}{m^{3/2}}\sum_{i=1}^{m} E\left(||T_m X_{i,m}||_\infty^3\right) \to_{m\to\infty} 0. \tag{6.21}$$

We will now show the condition (6.21). Then with Lemma 6.1.1 the assertion on the limit covariance matrix follows, since Corollary 18.3 of Bhattacharya and Rao (1976, [4]) is formulated for the standard normal distribution.

First we get by (i)

$$\begin{aligned} T_m &= \left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{Cov}(X_{i,m})\right)^{-1/2} \\ &= \left(\frac{1}{m}\sum_{i=1}^{m}V_\lambda + O(d_\lambda^* - \chi_\lambda(r_m,s))U_\lambda + t(\lambda,r_m)t^T(\lambda,r_m)\right)^{-1/2} \\ &= \left(V_\lambda + O(d_\lambda^* - \chi_\lambda(r_m,s))U_\lambda + t(\lambda,r_m)t^T(\lambda,r_m)\right)^{-1/2} \\ &\to_{m\to\infty} V_\lambda^{-1/2}, \end{aligned}$$

and then, where $X_{i,m}^{(j)}$ denotes the jth individual component of $X_{i,m}$, with Lemma 5.5.2

$$\begin{aligned} &\frac{1}{m^{3/2}}\sum_{i=1}^{m}E\left(||T_mX_{i,m}||_\infty^3\right) \\ &\quad\le \frac{1}{m^{3/2}}||T_m||_\infty^3\sum_{i=1}^{m}E\left(||X_{i,m}||_\infty^3\right) \\ &\quad\le \frac{1}{m^{3/2}}||T_m||_\infty^3\sum_{i=1}^{m}E\left(\sum_{j=1}^{k}\left|X_{i,m}^{(j)}\right|^3\right) \\ &\quad= \frac{1}{m^{3/2}}||T_m||_\infty^3\sum_{i=1}^{m}\sum_{j=1}^{k}E\left(\left|X_{i,m}^{(j)}\right|^3\right) \\ &\quad\le \frac{1}{m^{3/2}}||T_m||_\infty^3\sum_{i=1}^{m}\sum_{j=1}^{k}8E\left(\left|\frac{\partial}{\partial\lambda_j}\log\left(\varrho\left(\lambda,Z^{(i)}\right)\right)\right|^3\right) \\ &\quad\overset{(6.17)}{=} \frac{8}{m^{3/2}}||T_m||_\infty^3\sum_{i=1}^{m}\sum_{j=1}^{k}(c_{\lambda,j}+\tilde{q}_j(\lambda,r_m)) \\ &\quad= \frac{8}{m^{1/2}}||T_m||_\infty^3\sum_{j=1}^{k}(c_{\lambda,j}+\tilde{q}_j(\lambda,r_m)) \\ &\quad\to_{m\to\infty} 0. \end{aligned}$$

Thus (6.20) is shown.

Let $\varepsilon > 0$. We will now show that under certain conditions a solution of the maximum likelihood equation exists in the ball $K_\varepsilon^\infty(\lambda)$ with radius ε around λ with regard to the ∞–norm.

For this purpose we put

$$\eta := \min\left\{\varepsilon, \frac{\varepsilon}{||V_\lambda||_\infty}, \frac{1}{3|F_\lambda|\cdot||V_\lambda^{-1}||_\infty}\right\}. \tag{6.22}$$

Define

$$\Delta := \left\{V_\lambda\varphi\,\middle|\,\varphi\in\mathbb{R}^k, ||\varphi||_\infty\le\eta\right\}.$$

Then Δ is compact and convex, as well as $\Delta\subseteq K_\varepsilon^\infty(0)$, since by (6.22)

$$||V_\lambda\varphi||_\infty \le ||V_\lambda||_\infty||\varphi||_\infty \le ||V_\lambda||_\infty\eta \le \varepsilon.$$

We define now the function $f : \Delta \to \mathbb{R}^k$ by

$$f(x) := \frac{1}{m}\nabla_\theta \Upsilon(\theta)\Big|_{\theta=\lambda+V_\lambda^{-1}x} + x.$$

Due to the assumptions, f is continuous. We show next that $f(x) \in \Delta$ for $x \in \Delta$. For that we check $||V_\lambda^{-1} f(x)||_\infty \leq \eta$. From this we can conclude $f(x) = V_\lambda V_\lambda^{-1} f(x) \in \Delta$. With the representation $x = V_\lambda \varphi$ for a $\varphi \in K_\eta^\infty(0)$ we get

$$\begin{aligned}
||V_\lambda^{-1} f(x)||_\infty &\leq ||V_\lambda^{-1}||_\infty ||f(x)||_\infty = ||V_\lambda^{-1}||_\infty ||f(V_\lambda\varphi)||_\infty \\
&= ||V_\lambda^{-1}||_\infty \Bigg|\Bigg| A_m + B_m\left(\lambda + V_\lambda^{-1}V_\lambda\varphi - \lambda\right) \\
&\quad + \left(\frac{1}{m}\sum_{i=1}^m R(Z^{(i)}, \lambda + V_\lambda^{-1}V_\lambda\varphi, \lambda)\right)\left(\lambda + V_\lambda^{-1}V_\lambda\varphi - \lambda\right) + V_\lambda\varphi \Bigg|\Bigg|_\infty \\
&= ||V_\lambda^{-1}||_\infty \left|\left| A_m + (B_m + V_\lambda)\varphi + \left(\frac{1}{m}\sum_{i=1}^m R(Z^{(i)}, \lambda+\varphi, \lambda)\right)\varphi \right|\right|_\infty \\
&\overset{(6.13)}{\leq} ||V_\lambda^{-1}||_\infty ||A_m||_\infty + ||V_\lambda^{-1}||_\infty ||(B_m + V_\lambda)||_\infty ||\varphi||_\infty + ||V_\lambda^{-1}||_\infty |C_m| \cdot ||\lambda + \varphi - \lambda||_\infty \cdot ||\varphi||_\infty \\
&\overset{(6.22)}{\leq} ||V_\lambda^{-1}||_\infty ||A_m||_\infty + ||V_\lambda^{-1}||_\infty ||(B_m + V_\lambda)||_\infty \eta + \frac{||V_\lambda^{-1}||_\infty \cdot |C_m| \cdot \eta}{3|F_\lambda| \cdot ||V_\lambda^{-1}||_\infty} \\
&\leq ||V_\lambda^{-1}||_\infty ||A_m||_\infty + ||V_\lambda^{-1}||_\infty \eta ||(B_m + V_\lambda)||_\infty + \frac{\eta}{3}\frac{|C_m|}{|F_\lambda|}.
\end{aligned}$$

When the conditions

$$||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \quad ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2} \tag{6.23}$$

are fulfilled, the inequality

$$||V_\lambda^{-1} f(x)||_\infty \leq \frac{\eta}{4} + \frac{\eta}{4} + \frac{\eta}{3}\cdot\frac{3}{2} = \eta$$

follows.

We know that under (6.23) f is a continuous function with $f : \Delta \to \Delta$. By the Brower fixed point theorem (see Theorem 229.2 in Heuser (1998, [38])), a fixed point $x \in \Delta$ with $f(x) = x$ therefore exists. For this fixed point x, we have

$$\frac{1}{m}\nabla_\theta \Upsilon(\theta)\Big|_{\theta=\lambda+V_\lambda^{-1}x} + x = x \Longleftrightarrow \frac{1}{m}\nabla_\theta \Upsilon(\theta)\Big|_{\theta=\lambda+V_\lambda^{-1}x} = 0.$$

Thus $\lambda + V_\lambda^{-1}x$ is a solution of the maximum likelihood equation. If one writes x by $x = V_\lambda\varphi$ with $||\varphi||_\infty \leq \eta$, one arrives at $\lambda + V_\lambda^{-1}V_\lambda\varphi = \lambda + \varphi$. Thus under condition (6.23) a solution of the maximum likelihood equation lies in $K_\eta^\infty(\lambda) \subseteq K_\varepsilon^\infty(\lambda)$, what had to be shown.

Especially $K_\varepsilon^\infty(\lambda)$ contains, due to the continuity of $\nabla_\theta\Upsilon(\theta)$, the special, recursively defined solution

$$\hat\lambda_{m,r,\varepsilon} := \begin{pmatrix} \hat\lambda_{m,r,\varepsilon}^{(1)} \\ \hat\lambda_{m,r,\varepsilon}^{(2)} \\ \vdots \\ \hat\lambda_{m,r,\varepsilon}^{(k)} \end{pmatrix} = \begin{pmatrix} \inf\{\theta_1 | \theta \in K_\varepsilon^\infty(\lambda), \nabla_\theta\Upsilon(\theta) = 0\} \\ \inf\left\{\theta_2 | \theta \in K_\varepsilon^\infty(\lambda), \nabla_\theta\Upsilon(\theta) = 0, \theta_1 = \hat\lambda_{m,r,\varepsilon}^{(1)}\right\} \\ \vdots \\ \inf\left\{\theta_k | \theta \in K_\varepsilon^\infty(\lambda), \nabla_\theta\Upsilon(\theta) = 0, \theta_1 = \hat\lambda_{m,r,\varepsilon}^{(1)}, \ldots, \theta_{k-1} = \hat\lambda_{m,r,\varepsilon}^{(k-1)}\right\} \end{pmatrix}$$

with $\theta = (\theta_1, \ldots, \theta_k)^T$. Because of the construction of $\hat{\lambda}_{m,r,\varepsilon}$ with infima, $\hat{\lambda}_{m,r,\varepsilon}$ is a random variable, see the proof of the theorem in Section 4.2.2 of Serfling (1980, [63]). In case that the condition (6.23) is not fulfilled, we set $\hat{\lambda}_{m,r,\varepsilon} := 0$.

We now construct a random variable $\hat{\lambda}_{m,r}$, which does not depend on the choice of ε. We put, therefore, $\hat{\lambda}_{m,r_m} := \hat{\lambda}_{m,r_m,\frac{1}{m}}$. For reasons of clarity, we again let $r = r_m$ depend on m. We show that this random variable converges in probability to λ. Let $\varepsilon > 0$ and η be defined as in (6.22).

$$
\begin{aligned}
& P\left(||\hat{\lambda}_{m,r_m} - \lambda||_\infty \geq \varepsilon\right) \\
= \;& P\left(||\hat{\lambda}_{m,r_m} - \lambda||_\infty \geq \varepsilon, ||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \quad ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right) \\
& + P\left(||\hat{\lambda}_{m,r_m} - \lambda||_\infty \geq \varepsilon, \left\{||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \right.\right. \\
& \qquad \left.\left. ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right\}^c\right) \\
\leq \;& P\left(||\hat{\lambda}_{m,r_m} - \lambda||_\infty \geq \varepsilon, ||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \quad ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right) \\
& + P\left(\left\{||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \quad ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right\}^c\right) \\
= \;& P\left(||\hat{\lambda}_{m,r_m} - \lambda||_\infty \geq \varepsilon, ||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \quad ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right) \\
& + P\left(\left(\left\{||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}\right\} \cap \left\{||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}\right\} \cap \left\{\frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right\}\right)^c\right) \\
= \;& P\left(||\hat{\lambda}_{m,r_m} - \lambda||_\infty \geq \varepsilon, ||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \quad ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right) \\
& + P\left(\left\{||A_m||_\infty > \frac{\eta}{4||V_\lambda^{-1}||_\infty}\right\} \cup \left\{||B_m + V_\lambda||_\infty > \frac{1}{4||V_\lambda^{-1}||_\infty}\right\} \cup \left\{\frac{|C_m|}{|F_\lambda|} > \frac{3}{2}\right\}\right) \\
\leq \;& P\left(||\hat{\lambda}_{m,r_m} - \lambda||_\infty \geq \varepsilon, ||A_m||_\infty \leq \frac{\eta}{4||V_\lambda^{-1}||_\infty}, \quad ||B_m + V_\lambda||_\infty \leq \frac{1}{4||V_\lambda^{-1}||_\infty}, \quad \frac{|C_m|}{|F_\lambda|} \leq \frac{3}{2}\right) \\
& + P\left(||A_m||_\infty > \frac{\eta}{4||V_\lambda^{-1}||_\infty}\right) + P\left(||B_m + V_\lambda||_\infty > \frac{1}{4||V_\lambda^{-1}||_\infty}\right) + P\left(\frac{|C_m|}{|F_\lambda|} > \frac{3}{2}\right) \\
& \to_{m\to\infty} 0.
\end{aligned}
$$

The convergence of the last three terms follows from (6.19). The estimator $\hat{\lambda}_{m,r_m}$ is, when fulfilling the secondary conditions (6.23), by definition in a ball around λ, with a radius smaller than ϵ from this m onward. Thus, the first term converges to 0, since $\frac{1}{m} < \varepsilon$ for m large enough.

We have now

$$
0 = \frac{1}{m}\, \nabla_\theta \Upsilon(\theta)|_{\theta=\hat{\lambda}_{m,r}} = A_m + \underbrace{\left(B_m + \frac{1}{m}\sum_{i=1}^{m} R\left(Z^{(i)}, \hat{\lambda}_{m,r}, \lambda\right)\right)}_{=:G_{m,r}} \left(\hat{\lambda}_{m,r} - \lambda\right),
$$

with a probability converging to 1 for all m and r large enough. For $m \to \infty$, $r \to \infty$, we have $G_{m,r} \to -V_\lambda$ due to (6.13), and $\hat{\lambda}_{m,r} \to \lambda$ in probability. Since V_λ is invertible by assumption, $G_{m,r}$ is also invertible for m and r large enough. For these m and r, we have

$$\sqrt{m}\left(\hat{\lambda}_{m,r} - \lambda\right) = -G_{m,r}^{-1}\sqrt{m}A_m.$$

By the asymptotic normality (6.20) of A_m, Lemma 6.1.1 and the symmetry of V_λ, we finally conclude

$$\sqrt{m}\left(\hat{\lambda}_{m,r} - \lambda\right) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, \left(V_\lambda^{-1}\right)^T V_\lambda V_\lambda^{-1}\right) = \mathcal{N}\left(0, V_\lambda^{-1}\right)$$

for $m \to \infty$, $r \to \infty$.

Because of (6.16), we have for the Fisher information I_λ of the angular component Z of a random vector X distributed by W_λ, for whom $C < -r$ and $X \in K_s$ holds,

$$I_\lambda = V_\lambda + O\left(d_\lambda^* - \chi_\lambda(r,s)\right) U_\lambda \to_{r\to\infty} V_\lambda.$$

I.e., the estimation procedure has asymptotically the inverse of the Fisher information as covariance matrix and is, thus, asymptotically efficient.

□

Example 6.1.3
We choose again the logistic distribution W_λ as an example. With the angular density given in Example 2.3.10, we calculate as function, which is to be maximized,

$$\begin{aligned}\Upsilon(\lambda) &= \sum_{i=1}^{m}\left(\sum_{j=1}^{d-1}\log(j\lambda - 1)\right) + (-\lambda - 1)\sum_{j=1}^{d-1}\log\left(Z_j^{(i)}\right) + (-\lambda - 1)\log\left(1 - \sum_{j=1}^{d-1} Z_j^{(i)}\right) \\ &\quad + \left(\frac{1}{\lambda} - d\right)\log\left(\sum_{j=1}^{d-1}\left(Z_j^{(i)}\right)^{-\lambda} + \left(1 - \sum_{j=1}^{d-1} Z_j^{(i)}\right)^{-\lambda}\right) - m\log d\end{aligned}$$

with $Z^{(i)} = \left(Z_1^{(i)}, \ldots, Z_{d-1}^{(i)}\right)$. Remark that in the logistic case $d_\lambda^* = d$ holds for $\lambda > 1$.

The subtraction of the constant $m \log d$ does not influence the maximization, thus one can maximize the term

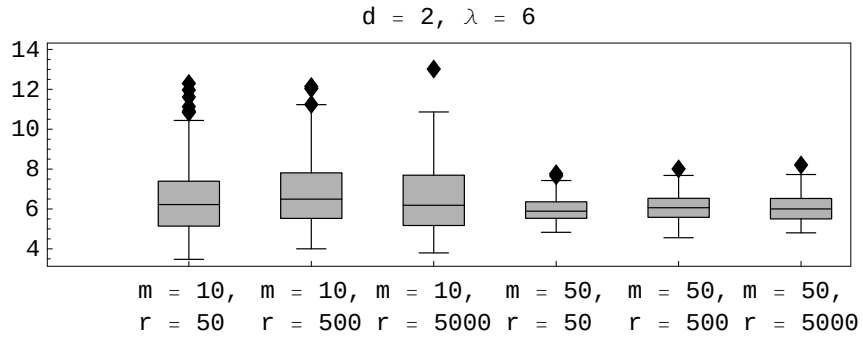
$$\sum_{i=1}^{m}\log\left(l_\lambda\left(Z^{(i)}\right)\right)$$

in λ.

The only realistic way to find maximal values of Υ in practice is, as stated above, by numerical methods. In this example, we use the function `NMaximize` of the software package MATHEMATICA, version 5.2.

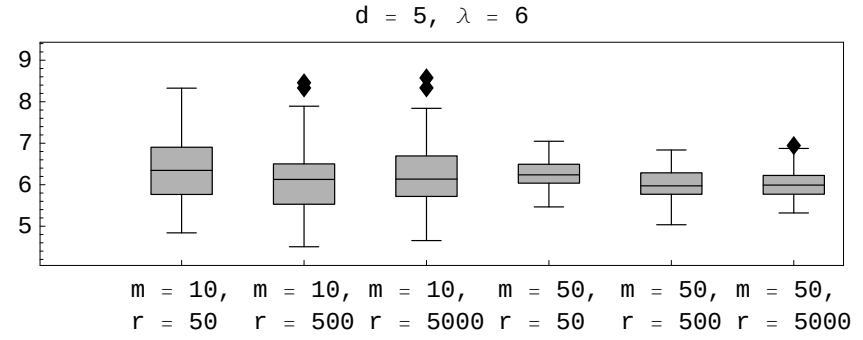
With Algorithm 3.1.11, we generated the necessary data for simulations of $\hat{\lambda}_{m,r}$. Thereby, for each combination of the parameter values $m = 10, 50$ and $r = 50, 500, 5000$, a data set was created with the fixed parameter values $d = 2$ and $\lambda = 6$. For each of the six data sets λ was

estimated. This was done 100 times for each setting of parameters, and in the following graphics, the 100 estimations $\hat{\lambda}_{m,r}$ for λ are represented by a boxplot.



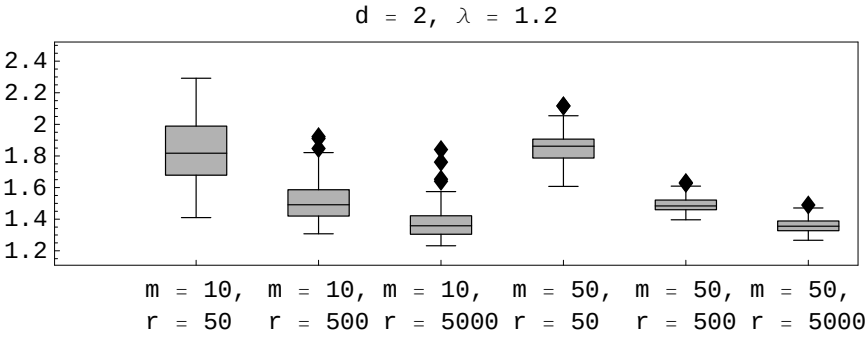
As was to be expected, the variance of the estimations reduces, when they are based on more data. The boxplots become more and more symmetric with increasing parameters. Thus, the estimations seem to work well and estimate the right parameter on average for these parameter settings, independent of r.

For the next plot we set $d = 5$, all other parameters remain unchanged.



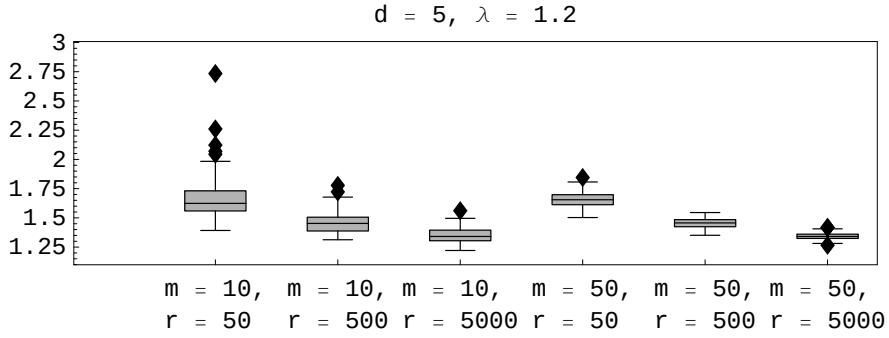
Once again the results are satisfactory.

In the next step d, is set to 2 again and λ is set to 1.2.



Here we see a clear dependence of the results on the choice of r. The value 1.2, which is to be estimated, lies for all boxplots outside the whiskers. Thus, we have a bias, which vanishes asymptotically by Theorem 6.1.2. The overestimation of λ reduces, as one should expect, only

with increasing r, but even for the value of $r = 5000$ it can still be seen in the above graphic. The same effect can be observed for $d = 5$.



So we have an estimation procedure, which is asymptotically efficient and works fine for cases of high dependence, but is biased for fixed sample sizes when approaching the independence case and, thus, not reliable there. This is not an uncommon phenomenon in extreme value analysis. For example the ML estimators presented in Tawn (1988, [72]) are shown to behave badly when close to the independence case. In Section 6.3, we will introduce an estimation procedure, which has reversed properties, i.e, is reliable close to the independence case and has high variability close to the dependence case.

$\Diamond$

The approximation of the density of the $Z^{(i)}$ by l_λ / d^*_λ seems to be very crude for fixed r, when one is close to the independence case. If we look at the density of the $Z^{(i)}$ under the additional condition $C^{(i)} = -r_i$ and $Z^{(i)} \in Q_{r_i,s}$, we know by Theorem 5.1.6 that the $Z^{(i)}$ have the density

$$\tilde{l}_{\lambda,r_i}(z) = \begin{cases} \frac{l_\lambda(z)}{\chi_\lambda(r_i,s)} & \text{for } z \in Q_{r_i,s}, \\ 0 & \text{else.} \end{cases}$$

This can also be used for a conditional approach of a maximum likelihood estimation of $\lambda = (\lambda_1, \ldots, \lambda_k)$ by choosing $\hat{\lambda}_{m,r}$ such that the expression

$$\prod_{i=1}^{m} \tilde{l}_{\lambda,r_i}\left(Z^{(i)}\right)$$

is maximized in λ. This is equivalent to maximizing the term

$$\begin{aligned} \tilde{\Upsilon}(\lambda) &:= \log\left(\prod_{i=1}^{m} \tilde{l}_{\lambda,r_i}\left(Z^{(i)}\right)\right) \\ &= \sum_{i=1}^{m} \log\left(\tilde{l}_{\lambda,r_i}\left(Z^{(i)}\right)\right) = \sum_{i=1}^{m} \log\left(\frac{l_\lambda\left(Z^{(i)}\right)}{\chi_\lambda(r_i,s)}\right) \\ &= \sum_{i=1}^{m} \log\left(l_\lambda\left(Z^{(i)}\right)\right) - \sum_{i=1}^{m} \log\left(\chi_\lambda(r_i,s)\right) \end{aligned}$$

in $\lambda = (\lambda_1, \ldots, \lambda_k)$. Since we are using conditional densities we refer to this method as the conditional maximum likelihood (Conditional ML) method.

In contrast to the Asymptotic ML method via $\Upsilon(\lambda)$, we have the additional complication that $\chi_\lambda(r_i, s)$ is an integral and must be evaluated numerically, since analytical expressions are not even known for the most common parametric models like the logistic or asymmetric logistic model. Here m different integrals have to be evaluated for $\tilde{\Upsilon}(\lambda)$, whereas for a computation of $\Upsilon(\lambda)$ at most one numerical integral has to be evaluated if no analytical expression of d_λ^* is known. This makes an evaluation of $\tilde{\Upsilon}(\lambda)$ much more costly and, thus, prolongs the runtime for the maximization of $\tilde{\Upsilon}(\lambda)$ by a high factor.

We show next that the Conditional ML estimation with $\tilde{\Upsilon}(\lambda)$ also leads to an estimator, which is asymptotically normal under suitable regularity conditions.

Theorem 6.1.4
Let $s > 0$ be fixed and $r > 0$ such that $Q_{r,s} \neq \emptyset$. Let $X_1, \ldots, X_m$ be independent random vectors, which follow a generalized Pareto distribution W_λ with $d_\lambda^* > 0$, with $C^{(i)} = -r_i$ and $r_i > r$ for all $i = 1, \ldots, m$, as well as $X_i \in K_s$. Let the function

$$\varrho(\lambda, z, r) := \frac{l_\lambda(z)}{\chi_\lambda(r, s)}$$

be three times continuously differentiable in λ for every r large enough. $l_\lambda(z)$ is the angular density, and χ is defined as in Definition 5.1.3. In addition, we assume that for every $\tilde{\theta} \in \Lambda$ there exist functions $g_r(z)$, $h_r(z)$ and $H_r(z) > 0$, possibly depending on $\tilde{\theta}$ and r, such that the relations

$$\left|\frac{\partial}{\partial\theta_j}\varrho(\theta, z, r)\right| \leq g_r(z), \qquad \left|\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}}\varrho(\theta, z, r)\right| \leq h_r(z), \tag{6.24}$$
$$\left|\frac{\partial^3}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}}\log\left(\varrho(\theta, z, r)\right)\right| \leq H_r(z),$$

hold for r large enough, for θ in a neighborhood $N(\tilde{\theta})$ of $\tilde{\theta}$, for $j, j_1, j_2, j_3 = 1, \ldots, k$, and for all $z \in R_{d-1}$. Next we demand that

$$\int_{R_{d-1}} g_r(z)\, dz < \infty, \qquad \int_{R_{d-1}} h_r(z)\, dz < \infty,$$
$$t_{\theta,r} := \int_{Q_{r,s}} H_r(z)\varrho(\theta, z, r)\, dz < \infty, \qquad \tilde{t}_{\theta,r} := \int_{Q_{r,s}} H_r^2(z)\varrho(\theta, z, r)\, dz < \infty \tag{6.25}$$

for $\theta \in N(\tilde{\theta})$ and all r large enough. Furthermore, we assume that for $\lambda \in \Lambda$, r large enough and $j, j_1, j_2 = 1, \ldots, k$, the integrals

$$0 < v_{\lambda,j_1,j_2,r} := \int_{Q_{r,s}} \frac{1}{\varrho(\lambda, z, r)} \left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda, z, r)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda, z, r)\right) dz < \infty, \tag{6.26}$$

$$w_{\lambda,j_1,j_2,r} := \int_{Q_{r,s}} \left(\frac{1}{\varrho(\lambda, z, r)}\left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}}\varrho(\lambda, z, r)\right)\right.$$
$$\left. - \left(\frac{1}{\varrho(\lambda, z, r)}\right)^2 \frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda, z, r) \cdot \frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda, z, r)\right)^2 \varrho(\lambda, z, r)\, dz < \infty, \tag{6.27}$$

$$u_{\lambda,j,r} := \int_{Q_{r,s}} \left(\frac{1}{\varrho(\lambda, z, r)}\right)^2 \cdot \left|\frac{\partial}{\partial\lambda_j}\varrho(\lambda, z, r)\right|^3 dz < \infty, \tag{6.28}$$

exist. These regularity conditions (6.24) to (6.28) make sure that certain moments of diverse random variables in the proof exist and that certain limits and integrals can be exchanged. In addition, we require that the behavior of these integrals with regard to r is regular, more precisely there have to exist real numbers F_λ, β_λ, γ_{λ,j_1,j_2}, δ_{λ,j_1,j_2} and $\alpha_{\lambda,j}$ such that

$$\begin{aligned} t_{\theta,r} &\to F_\lambda, & \tilde{t}_{\theta,r} &\to \beta_\lambda, \\ v_{\lambda,j_1,j_2,r} &\to \gamma_{\lambda,j_1,j_2}, & w_{\lambda,j_1,j_2,r} &\to \delta_{\lambda,j_1,j_2}, \\ & & \text{and} \quad u_{\lambda,j,r} &\to \alpha_{\lambda,j}, \end{aligned} \tag{6.29}$$

monotone increasing for $r \to \infty$.

Let the matrix

$$\Gamma_\lambda := \left(\gamma_{\lambda,j_1,j_2}\right)_{j_1,j_2=1,\ldots,k}$$

be invertible. Then there exists a sequence of maximum likelihood estimators $\hat{\lambda}_{m,r}$, which solve $\nabla_\theta \tilde{\Upsilon}(\theta) = 0$ and for whom the convergence

$$\sqrt{m}\left(\hat{\lambda}_{m,r} - \lambda\right) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, \Gamma_\lambda^{-1}\right)$$

holds for $m \to \infty$ and $r \to \infty$.

Proof:
We will proceed analogously as in the proof of Theorem 6.1.2.

Because of the differentiability of $\varrho(\theta, z, r)$ with regard to θ, there exists a Taylor series expansion for the gradient ∇_θ of the function $\log(\varrho(\theta, z, r))$ in a neighborhood $N(\lambda)$ of λ, the underlying parameter. We write it as

$$\nabla_\theta \log(\varrho(\theta, z, r)) = \nabla_\theta \log(\varrho(\theta, z, r))|_{\theta=\lambda} + \left(\mathrm{He}_\theta \log(\varrho(\theta, z, r))|_{\theta=\lambda} + R(z, \theta, \lambda, r)\right)(\theta - \lambda),$$

where He_θ again denotes the Hessian matrix, where $R(z, \theta, \lambda, r)$ is a matrix with

$$||R(z, \theta, \lambda, r)||_\infty \leq H_r(z)||\theta - \lambda||_\infty$$

and H_r is the function from condition (6.24), see Theorems 168.4 and 168.5 in Heuser (1998, [38]). We put as before

$$\begin{aligned} A_m &:= \frac{1}{m}\sum_{i=1}^m \nabla_\theta \log\left(\varrho\left(\theta, Z^{(i)}, r_i\right)\right)\Big|_{\theta=\lambda}, \\ B_m &:= \frac{1}{m}\sum_{i=1}^m \mathrm{He}_\theta \log\left(\varrho\left(\theta, Z^{(i)}, r_i\right)\right)\Big|_{\theta=\lambda}, \\ C_m &:= \frac{1}{m}\sum_{i=1}^m H_{r_i}\left(Z^{(i)}\right). \end{aligned}$$

Then we have

$$\frac{1}{m}\nabla_\theta \tilde{\Upsilon}(\theta) = A_m + \left(B_m + \frac{1}{m}\sum_{i=1}^m R\left(Z^{(i)}, \theta, \lambda, r_i\right)\right)(\theta - \lambda), \tag{6.30}$$

where

$$\left|\left|\frac{1}{m}\sum_{i=1}^m R\left(Z^{(i)}, \theta, \lambda, r_i\right)\right|\right|_\infty \leq |C_m| \cdot ||\theta - \lambda||_\infty. \tag{6.31}$$

Due to conditions (6.24) and (6.25) differentiation and integration can be interchanged, and we get for $j = 1, \ldots, k$ and r large enough

$$\int_{Q_{r,s}} \frac{\partial}{\partial \theta_j} \varrho(\theta, z, r)\, dz = \frac{\partial}{\partial \theta_j} \int_{Q_{r,s}} \varrho(\theta, z, r)\, dz = \frac{\partial}{\partial \theta_j} \int_{Q_{r,s}} \frac{l_\theta(z)}{\chi_\theta(r,s)}\, dz \overset{(5.4)}{=} \frac{\partial}{\partial \theta_j}(1) = 0, \quad (6.32)$$

and equally for $j_1, j_2 = 1, \ldots, k$

$$\int_{Q_{r,s}} \frac{\partial^2}{\partial \theta_{j_1} \partial \theta_{j_2}} \varrho(\theta, z, r)\, dz = 0. \quad (6.33)$$

Theorem 5.1.6 states that the $Z^{(i)}$ possess a density, which is $\varrho(\theta, z, r_i)$ for $z \in Q_{r,s}$ and 0 else. Note that since we condition on $C^{(i)} = r_i$ the $Z^{(i)}$ are, in contrast to the proof of Theorem 6.1.2, not identically distributed. So we have for $j = 1, \ldots, k$ and $i = 1, \ldots, m$

$$E_\lambda \left(\frac{\partial \log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right)}{\partial \lambda_j} \right) = \int_{Q_{r_i,s}} \frac{1}{\varrho\left(\lambda, z, r_i\right)} \left(\frac{\partial}{\partial \lambda_j} \varrho\left(\lambda, z, r_i\right) \right) \varrho(\lambda, z, r_i)\, dz \overset{(6.32)}{=} 0.$$

Furthermore, we have for $j_1, j_2 = 1, \ldots, k$ and $i = 1, \ldots, m$

$$\begin{aligned}
&E_\lambda \left(\frac{\partial^2 \log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right)}{\partial \lambda_{j_1} \partial \lambda_{j_2}} \right) = \\
&= \int_{Q_{r_i,s}} \left(\frac{1}{\varrho\left(\lambda, z, r_i\right)} \left(\frac{\partial^2}{\partial \lambda_{j_1} \partial \lambda_{j_2}} \varrho\left(\lambda, z, r_i\right) \right) \right. \\
&\quad \left. - \left(\frac{1}{\varrho\left(\lambda, z, r_i\right)} \right)^2 \left(\frac{\partial}{\partial \lambda_{j_1}} \varrho\left(\lambda, z, r_i\right) \right) \left(\frac{\partial}{\partial \lambda_{j_2}} \varrho\left(\lambda, z, r_i\right) \right) \right) \varrho(\lambda, z, r_i)\, dz \\
&\overset{(6.33)}{=} -\int_{Q_{r_i,s}} \frac{1}{\varrho(\lambda, z, r_i)} \left(\frac{\partial}{\partial \lambda_{j_1}} \varrho(\lambda, z, r_i) \right) \left(\frac{\partial}{\partial \lambda_{j_2}} \varrho(\lambda, z, r_i) \right) dz \\
&\overset{(6.26)}{=} -v_{\lambda, j_1, j_2, r_i}.
\end{aligned}$$

Analogously

$$\begin{aligned}
&E_\lambda \left(\frac{\partial \log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right)}{\partial \lambda_{j_1}} \cdot \frac{\partial \log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right)}{\partial \lambda_{j_2}} \right) = \\
&= \int_{Q_{r_i,s}} \left(\frac{1}{\varrho\left(\lambda, z, r_i\right)} \right)^2 \left(\frac{\partial}{\partial \lambda_{j_1}} \varrho\left(\lambda, z, r_i\right) \right) \left(\frac{\partial}{\partial \lambda_{j_2}} \varrho\left(\lambda, z, r_i\right) \right) \varrho(\lambda, z, r_i)\, dz \\
&= \int_{Q_{r_i,s}} \frac{1}{\varrho\left(\lambda, z, r_i\right)} \left(\frac{\partial}{\partial \lambda_{j_1}} \varrho\left(\lambda, z, r_i\right) \right) \left(\frac{\partial}{\partial \lambda_{j_2}} \varrho\left(\lambda, z, r_i\right) \right) dz \\
&\overset{(6.26)}{=} v_{\lambda, j_1, j_2, r_i}.
\end{aligned}$$

Next we consider

$$E_\lambda \left(\left(\frac{\partial^2 \log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right)}{\partial \lambda_{j_1} \partial \lambda_{j_2}} \right)^2 \right) =$$

$$
\begin{aligned}
&= \int_{Q_{r_i},s} \left(\frac{1}{\varrho\left(\lambda, z, r_i\right)} \left(\frac{\partial^2}{\partial\lambda_{j_1}\partial\lambda_{j_2}} \varrho\left(\lambda, z, r_i\right) \right) \right. \\
&\quad \left. - \left(\frac{1}{\varrho\left(\lambda, z, r_i\right)} \right)^2 \left(\frac{\partial}{\partial\lambda_{j_1}} \varrho\left(\lambda, z, r_i\right) \right) \left(\frac{\partial}{\partial\lambda_{j_2}} \varrho\left(\lambda, z, r_i\right) \right) \right)^2 \varrho(\lambda, z, r_i)\, dz \\
&\overset{(6.27)}{=} w_{\lambda,j_1,j_2,r_i}.
\end{aligned}
$$

Finally we have

$$
\begin{aligned}
&E_\lambda \left(\left| \frac{\partial \log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right)}{\partial\lambda_j} \right|^3 \right) = \\
&= \int_{Q_{r_i},s} \left(\frac{1}{\varrho\left(\lambda, z, r_i\right)} \left| \frac{\partial}{\partial\lambda_j} \varrho\left(\lambda, z, r_i\right) \right| \right)^3 \varrho(\lambda, z, r_i)\, dz \\
&= \int_{Q_{r_i},s} \left(\frac{1}{\varrho\left(\lambda, z, r_i\right)} \right)^2 \left| \frac{\partial}{\partial\lambda_j} \varrho\left(\lambda, z, r_i\right) \right|^3 dz \\
&\overset{(6.28)}{=} u_{\lambda,j,r_i}. \qquad (6.34)
\end{aligned}
$$

From the above considerations we conclude that

(i) A_m is an arithmetic mean of independent but not identically distributed random vectors with expectation 0 and covariance matrix $V_{\lambda,r_i} = \left(v_{\lambda,j_1,j_2,r_i}\right)_{j_1,j_2=1,\ldots,k}$.

(ii) B_m is an arithmetic mean of independent but not identically distributed random vectors (noted as a matrix) with expectation matrix $-V_{\lambda,r_i}$. The components of the random vectors underlying B_m have the variance $w_{\lambda,j_1,j_2,r_i} - v^2_{\lambda,j_1,j_2,r_i}$.

(iii) C_m is an arithmetic mean of independent, but not identically distributed random variables with expectation t_{λ,r_i} and with variance $\tilde{t}_{\lambda,r_i} - t^2_{\lambda,r_i}$ due to (6.25).

Next, we will show the convergences

$$
A_m \to_{m\to\infty} 0, \qquad B_m \to_{m\to\infty,r\to\infty} -\Gamma_\lambda, \qquad C_m \to_{m\to\infty,r\to\infty} F_\lambda \qquad (6.35)
$$

in probability. We denote by $A_{m,j}$ the jth component of A_m, and for $\varepsilon > 0$ we get with the help of the inequality by Tchebychev, see Corollary 1.18.3 in Gänssler and Stute (1977, [30])

$$
\begin{aligned}
P\left(||A_m||_\infty \geq \varepsilon\right) &= P\left(\bigcup_{j=1}^k \left\{ |A_{m,j}| \geq \varepsilon \right\} \right) \leq \sum_{j=1}^k P\left(|A_{m,j}| \geq \varepsilon \right) \leq \sum_{j=1}^k \frac{\mathrm{Var}(A_{m,j})}{\varepsilon^2} \\
&\overset{(i)}{=} \frac{1}{m^2\varepsilon^2} \sum_{j=1}^k \sum_{i=1}^m v_{\lambda,j,j,r_i} \leq \frac{1}{m\varepsilon^2} \sum_{j=1}^k \gamma_{\lambda,j,j} \\
&\to_{m\to\infty} 0.
\end{aligned}
$$

Thus, we have shown the first convergence of (6.35).

Analogously we show the assertion for B_m, where B_{m,j_1,j_2} denotes the (j_1,j_2)th component of the random matrix B_m.

$$
\begin{aligned}
&P\left(||B_m+\Gamma_\lambda||_\infty \geq \varepsilon\right)\\
&\quad = \quad P\left(\left|\left|B_m+\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}+\Gamma_\lambda-\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}\right|\right|_\infty \geq \varepsilon\right)\\
&\quad \leq \quad P\left(\left|\left|B_m+\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}\right|\right|_\infty \geq \varepsilon-\left|\left|\Gamma_\lambda-\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}\right|\right|_\infty\right)\\
&\quad = \quad P\left(\bigcup_{j_1,j_2=1}^k\left\{\left|B_{m,j_1,j_2}+\frac{1}{m}\sum_{i=1}^m v_{\lambda,j_1,j_2,r_i}\right| \geq \varepsilon-\left|\left|\Gamma_\lambda-\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}\right|\right|_\infty\right\}\right)\\
&\quad \leq \quad \sum_{j_1,j_2=1}^k P\left(\left\{\left|B_{m,j_1,j_2}+\frac{1}{m}\sum_{i=1}^m v_{\lambda,j_1,j_2,r_i}\right| \geq \varepsilon-\left|\left|\Gamma_\lambda-\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}\right|\right|_\infty\right\}\right)\\
&\quad \overset{(ii)}{\leq} \quad \sum_{j_1,j_2=1}^k \frac{\frac{1}{m^2}\sum_{i=1}^m w_{\lambda,j_1,j_2,r_i}-v^2_{\lambda,j_1,j_2,r_i}}{\left(\varepsilon-||\Gamma_\lambda-\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}||_\infty\right)^2}\\
&\quad \overset{(6.29)}{\leq} \quad \sum_{j_1,j_2=1}^k \frac{\frac{1}{m^2}\sum_{i=1}^m \delta_{\lambda,j_1,j_2}-v^2_{\lambda,j_1,j_2,r}}{\left(\varepsilon-||\Gamma_\lambda-\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}||_\infty\right)^2}\\
&\quad = \quad \frac{1}{m}\sum_{j_1,j_2=1}^k \frac{\delta_{\lambda,j_1,j_2}-v^2_{\lambda,j_1,j_2,r}}{\left(\varepsilon-||\Gamma_\lambda-\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i}||_\infty\right)^2}\\
&\quad \to_{m\to\infty,r\to\infty} \quad 0,
\end{aligned}
$$

since the numerator is bounded and the denominator is bounded away from 0 for r large enough, due to $\frac{1}{m}\sum_{i=1}^m V_{\lambda,r_i} \to_{r\to\infty} \Gamma_\lambda$ by (6.29).

Likewise we conclude for C_m

$$
\begin{aligned}
P\left(|C_m-F_\lambda| \geq \varepsilon\right) \quad &= \quad P\left(\left|C_m-\frac{1}{m}\sum_{i=1}^m t_{\lambda,r_i}-\left(F_\lambda-\frac{1}{m}\sum_{i=1}^m t_{\lambda,r_i}\right)\right| \geq \varepsilon\right)\\
&\leq \quad P\left(\left|C_m-\frac{1}{m}\sum_{i=1}^m t_{\lambda,r_i}\right| \geq \varepsilon-\left|F_\lambda-\frac{1}{m}\sum_{i=1}^m t_{\lambda,r_i}\right|\right)\\
&\overset{(iii)}{\leq} \quad \frac{\frac{1}{m^2}\sum_{i=1}^m \tilde{t}_{\lambda,r_i}-t^2_{\lambda,r_i}}{\left(\varepsilon-\left|F_\lambda-\frac{1}{m}\sum_{i=1}^m t_{\lambda,r_i}\right|\right)^2}\\
&\leq \quad \frac{1}{m}\frac{\beta_\lambda-t^2_{\lambda,r}}{\left(\varepsilon-\left|F_\lambda-\frac{1}{m}\sum_{i=1}^m t_{\lambda,r_i}\right|\right)^2}\\
&\to_{m\to\infty,r\to\infty} \quad 0,
\end{aligned}
$$

since the numerator is bounded and the denominator is bounded away from 0 for r large enough, due to $\frac{1}{m}\sum_{i=1}^m t_{\lambda,r_i} \to_{r\to\infty} F_\lambda$ by (6.29).

The aim is now to show that

$$\sqrt{m}A_m \longrightarrow_{\mathcal{D}} \mathcal{N}(0,\Gamma_\lambda) \tag{6.36}$$

holds.

For that we use again the version of the central limit theorem by Lindeberg for the multivariate case as described in Corollary 18.3 of Bhattacharya and Rao (1976, [4]) and already used in the proof of Theorem 6.1.2. We put

$$X_{i,m} := \nabla_\lambda \log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right).$$

Then we have

$$\sqrt{m}A_m = \frac{1}{\sqrt{m}}\sum_{i=1}^{m} X_{i,m}.$$

We set

$$T_m := \left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{Cov}(X_{i,m})\right)^{-1/2}.$$

From the multivariate central limit theorem by Lindeberg, we conclude the asymptotic normality if

$$\frac{1}{m^{3/2}}\sum_{i=1}^{m} E\left(||T_m X_{i,m}||_\infty^3\right) \to_{m\to\infty, r\to\infty} 0. \tag{6.37}$$

We will now show the condition (6.37) as before. Then with Lemma 6.1.1, the assertion on the limit covariance matrix follows.

First we get by (i)

$$T_m = \left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{Cov}(X_{i,m})\right)^{-1/2} = \left(\frac{1}{m}\sum_{i=1}^{m} V_{\lambda,r_i}\right)^{-1/2} \to_{m\to\infty, r\to\infty} \Gamma_\lambda^{-1/2},$$

since $\frac{1}{m}\sum_{i=1}^{m} v_{\lambda,j_1,j_2,r_i} \to \gamma_{\lambda,j_1,j_2}$ by (6.29).

Then, we denote with $X_{i,m}^{(j)}$ the jth individual component of $X_{i,m}$ and obtain

$$\begin{aligned}
&\frac{1}{m^{3/2}}\sum_{i=1}^{m} E\left(||T_m X_{i,m}||_\infty^3\right) \\
&\quad\leq \quad \frac{1}{m^{3/2}}||T_m||_\infty^3 \sum_{i=1}^{m} E\left(||X_{i,m}||_\infty^3\right) \\
&\quad\leq \quad \frac{1}{m^{3/2}}||T_m||_\infty^3 \sum_{i=1}^{m} E\left(\sum_{j=1}^{k}\left|X_{i,m}^{(j)}\right|^3\right) \\
&\quad= \quad \frac{1}{m^{3/2}}||T_m||_\infty^3 \sum_{i=1}^{m}\sum_{j=1}^{k} E\left(\left|\frac{\partial}{\partial\lambda_j}\log\left(\varrho\left(\lambda, Z^{(i)}, r_i\right)\right)\right|^3\right) \\
&\quad\overset{(6.34)}{=} \quad \frac{1}{m^{3/2}}||T_m||_\infty^3 \sum_{i=1}^{m}\sum_{j=1}^{k} u_{\lambda,j,r_i} \\
&\quad\overset{(6.29)}{\leq} \quad \frac{1}{m^{3/2}}||T_m||_\infty^3 \sum_{i=1}^{m}\sum_{j=1}^{k} \alpha_{\lambda,j} = \frac{1}{m^{1/2}}||T_m||_\infty^3 \sum_{j=1}^{k}\alpha_{\lambda,j} \\
&\to_{m\to\infty, r\to\infty} \quad 0.
\end{aligned}$$

Thus (6.36) is shown.

Now we are exactly in the same situation as in the proof of Theorem 6.1.2. The existence of a solution of the ML equation and its convergence in probability to λ follow in exactly the same manner. Consequently,

$$0 = \frac{1}{m}\,\nabla_\theta \tilde{\Upsilon}(\theta)\Big|_{\theta=\hat{\lambda}_{m,r}} = A_m + \underbrace{\left(B_m + \frac{1}{m}\sum_{i=1}^{m} R\left(Z^{(i)}, \hat{\lambda}_{m,r}, \lambda, r_i\right)\right)}_{=:G_{m,r,r_1,\ldots,r_m}} \left(\hat{\lambda}_{m,r} - \lambda\right),$$

with a probability which converges to 1 for all m and r large enough. We have

$$\begin{aligned} ||G_{m,r,r_1,\ldots,r_m} + \Gamma_\lambda||_\infty &= \left|\left|B_m + \Gamma_\lambda + \frac{1}{m}\sum_{i=1}^{m} R\left(Z^{(i)}, \hat{\lambda}_{m,r}, \lambda, r_i\right)\right|\right|_\infty \\ &\leq ||B_m + \Gamma_\lambda||_\infty + \left|\left|\frac{1}{m}\sum_{i=1}^{m} R\left(Z^{(i)}, \hat{\lambda}_{m,r}, \lambda, r_i\right)\right|\right|_\infty \\ &\leq ||B_m + \Gamma_\lambda||_\infty + |C_m| \cdot \left|\left|\hat{\lambda}_{m,r} - \lambda\right|\right|_\infty, \end{aligned}$$

thus $G_{m,r,r_1,\ldots,r_m} \to_{m\to\infty, r\to\infty} -\Gamma_\lambda$ in probability. The asymptotic normality of $\sqrt{m}\left(\hat{\lambda}_{m,r} - \lambda\right)$ now follows as in the proof of Theorem 6.1.2.

□

Remark that although we conditioned on $C^{(i)} = r_i$, all convergences in the above proof where shown to be only depending on r.

Corollary 6.1.5
Let $\tilde{X}^{(1)}, \ldots, \tilde{X}^{(n)}$ follow a GPD with the assumptions from Theorem 6.1.2. Then the Conditional ML estimator $\hat{\lambda}_{m,r}$, based on the random number $m = \tau_n$ of observations with $\tilde{C}^{(i)} < -r$, $||X^{(i)}||_\infty < s$, is asymptotically normal, i.e.,

$$\sqrt{m}\left(\hat{\lambda}_{m,r} - \lambda\right) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, \Gamma_\lambda^{-1}\right),$$

if $np_{r,s} \to \infty$, where $p_{r,s} = P\left(\tilde{X}^{(i)} \in A_{r,s}\right)$, and $n \to \infty$, $r \to \infty$ with growth condition (5.27) fulfilled.

Proof:
Due to $np_{r,s} \to \infty$ and (5.27), we know by Lemma 5.6.1 that $\tau_n = m \to \infty$ with probability 1 and, thus, with probability 1 a triangular array of radial components

$$r_n < r_{1,n}, \ldots, r_{m,n}$$

with $r_n \to \infty$, $m = \tau_n \to \infty$ is generated by the underlying random variables. Moving along this triangular array the assertion follows with Theorem 6.1.4, since all convergences there only depend on r.

□

Remark 6.1.6
Note that the quantity A_m in the proof of Theorem 6.1.4 has expectation 0 independent of r and r_i, whereas in the proof of Theorem 6.1.2 the quantity A_m had expectation $t(\lambda, r)$, which

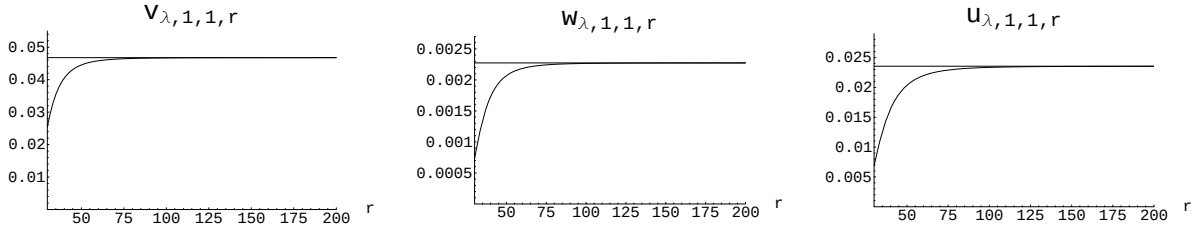
converged to 0 for $r \to \infty$ but did not vanish for fixed r. This resulted in a bias, which could be seen in the simulations of Example 6.1.3. We, thus, expect not to see a bias in simulations for the Conditional ML method, which are carried out in Example 6.1.7.

In addition to the conditions for the asymptotic normality of the Asymptotic ML estimator in Theorem 6.1.2, we also needed here that the conditions (6.29), concerning the behavior of diverse integrals with regard to r, are fulfilled. We take a short look at the condition concerning $v_{\lambda,j_1,j_2,r}$. The assumption is that a $\gamma_{\lambda,j_1,j_2} < \infty$ exists, such that $v_{\lambda,j_1,j_2,r} \to_{r\to\infty} \gamma_{\lambda,j_1,j_2}$ monotonously increasing. If we could exchange the corresponding limits with the integration and differentiation, we would get

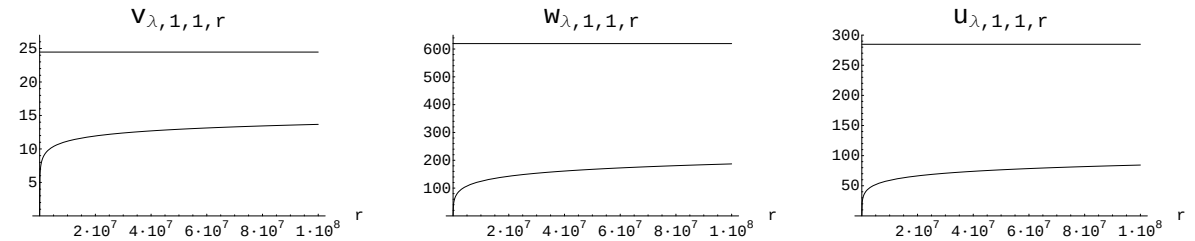
$$\begin{aligned}
\gamma_{\lambda,j_1,j_2} &= \lim_{r\to\infty} v_{\lambda,j_1,j_2,r} = \lim_{r\to\infty} \int_{Q_{r,s}} \frac{1}{\varrho(\lambda,z,r)} \left(\frac{\partial}{\partial\lambda_{j_1}}\varrho(\lambda,z,r)\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\varrho(\lambda,z,r)\right) dz \\
&= \lim_{r\to\infty} \int_{Q_{r,s}} \frac{\chi_\lambda(r,s)}{l_\lambda(z)} \left(\frac{\partial}{\partial\lambda_{j_1}}\frac{l_\lambda(z)}{\chi_\lambda(r,s)}\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\frac{l_\lambda(z)}{\chi_\lambda(r,s)}\right) dz \\
&= \int_{R_{d-1}} \frac{d^*_\lambda}{l_\lambda(z)} \left(\frac{\partial}{\partial\lambda_{j_1}}\frac{l_\lambda(z)}{d^*_\lambda}\right)\left(\frac{\partial}{\partial\lambda_{j_2}}\frac{l_\lambda(z)}{d^*_\lambda}\right) dz \\
&= v_{\lambda,j_1,j_2}
\end{aligned}$$

with v_{λ,j_1,j_2} being defined as in Theorem 6.1.2. Thus condition (6.29) is fulfilled if the limit $r \to \infty$ can be interchanged with integration and differentiation, the limit integral exists and the convergence is monotonous in r. Considerations similar to those above can also be made for $w_{\lambda,j_1,j_2,r}$ and $u_{\lambda,j,r}$. Thus the requirements (6.29) are not unreasonable ones.

We want to illustrate conditions (6.29) for the logistic case by plotting the quantities $v_{\lambda,1,1,r}$, $w_{\lambda,1,1,r}$ and $u_{\lambda,1,1,r}$ for $d = 2$, $s = 0.1$ and different λ. As a reference, the corresponding limits if the exchange of limits with integration/differentiation is permitted, are drawn as lines. For the numerical evaluation we used the function `NIntegrate` from MATHEMATICA. With $\lambda = 6$ we get the following behavior:



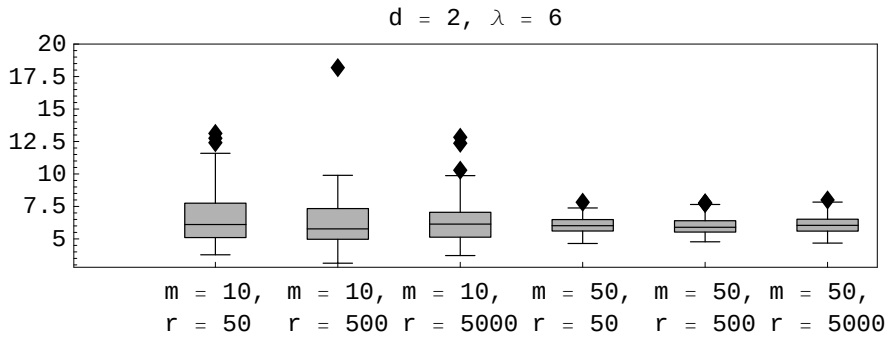
We have fast monotonous convergence to constants, which seem to be the values after the exchange of the limit $r \to \infty$ with the differentiation and integration. Next we look at $\lambda = 1.2$.

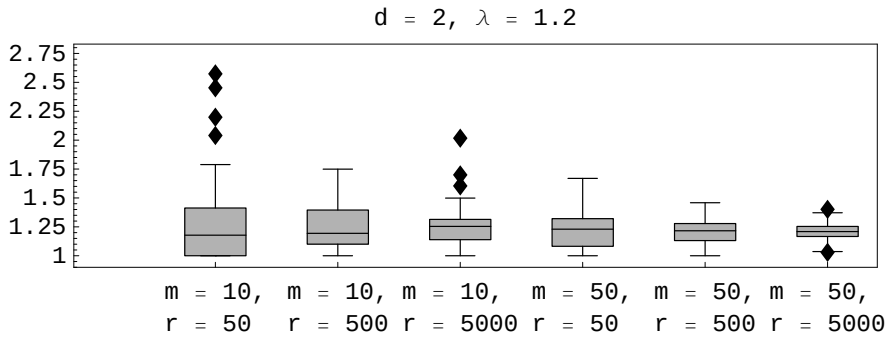


The convergence here is much slower but also monotonous. The quantities $v_{\lambda,1,1,r}$, $w_{\lambda,1,1,r}$ and $u_{\lambda,1,1,r}$ have a similar behavior as the function χ, see Example 5.1.5. χ also converged very slowly to $d_\lambda^* = d$ for λ close to 1. In Example 6.2.2 we show evaluations of $v_{\lambda,1,1,r}$ for very high r supporting the conjecture that $v_{\lambda,1,1,r}$ converges actually to $v_{\lambda,1,1}$ of the Asymptotic ML method. Thus the conditions (6.29) are, in all, not unrealistic requirements and seem to be fulfilled in the logistic case.

$\Diamond$

Example 6.1.7
We want to test the estimator, which maximizes $\tilde{\Upsilon}(\lambda)$ as in Example 6.1.3. The following plots are done for $d = 2$, the other parameters are chosen as in Example 6.1.3.





In the case of high dependence nothing has changed, in the case of low dependence, the bias we have seen in Example 6.1.3 is not existing any more. Remark also that in the low dependence case often a value of 1 is estimated, which would mean independence although we have dependence in our simulated data due to $\lambda > 1$.

We do not give graphics for higher dimensions, since the generation of a graphic like the one above for $d = 3$ can take days already. This is due to the numerical inefficiency of the Conditional ML method described above.

$\Diamond$

This procedure is hardly suited for practical purposes in higher dimensions, since the computation takes too long. A possible way to make this work in practice would be to find easily evaluable functions $f(s, r, \lambda)$ which approximate $\chi_\lambda(r, s)$, and then to maximize

$$\sum_{i=1}^{m} \log\left(l_\lambda\left(Z^{(i)}\right)\right) - \sum_{i=1}^{m} \log\left(f(s, r_i, \lambda)\right)$$

instead of $\tilde{\Upsilon}(\lambda)$. But such functions are yet unknown.

We will compare the two procedures presented here together with a few others, which will be introduced in the next sections, in Section 6.4 in detail.

6.2 Maximum Likelihood Estimation with the Pickands Density

As in the previous section, we assume that we have n independent copies $\tilde{X}^{(1)}, \ldots, \tilde{X}^{(n)}$ of a random vector X, which follows a generalized Pareto distribution W_λ from a k–parametric family with $U = K_s$ by representation (2.5). By ϕ_λ we denote the corresponding Pickands density, and this time we denote by $\tilde{Z}^{(i)}$ and $\tilde{C}^{(i)}$ the corresponding standard Pickands coordinates, $i = 1, \ldots, n$. We choose a threshold $r < 0$ close enough to 0 for the conditions of Theorem 2.2.5 to hold, and consider only those observations with $\tilde{C}^{(i)} > r$. We denote these by $X^{(1)}, \ldots, X^{(m)}$. They are independent (see Sections 4.1 and 5.2) and have, by Theorem 2.2.5, the density $\frac{\phi_\lambda(z)}{\mu_\lambda}$, independent of r. Again we can do a maximum likelihood estimation of $\lambda = (\lambda_1, \ldots, \lambda_k)$ by choosing $\hat{\lambda}_m$ such that the expression

$$\prod_{i=1}^{m} \frac{\phi_\lambda\left(Z^{(i)}\right)}{\mu_\lambda} \qquad \text{or} \qquad \Omega(\lambda) := \sum_{i=1}^{m} \log\left(\phi_\lambda\left(Z^{(i)}\right)\right) - m\log(\mu_\lambda)$$

is maximized in λ.

Since here, in contrast to Section 6.1, we insert the observations into their exact densities, we can refer for the proof of the asymptotic normality of the maximum likelihood estimation to the corresponding literature, see below.

Theorem 6.2.1
Let $r < 0$ be chosen such that the conditions of Theorem 2.2.5 hold. Let $X_1, \ldots, X_m$ be independent random vectors, distributed according to a generalized Pareto distribution W_λ, with $C^{(i)} > r$. Let the function

$$\varrho(\lambda, z) := \frac{\phi_\lambda(z)}{\mu_\lambda}$$

be three times continuously differentiable in λ, where $\phi_\lambda(z)$ is the Pickands density. In addition, we assume that for every $\tilde{\theta} \in \Lambda$ there exist functions $g(z)$, $h(z)$ and $H(z) > 0$, possibly depending on $\tilde{\theta}$, such that the relations

$$\left|\frac{\partial}{\partial\theta_j}\varrho(\theta, z)\right| \le g(z), \qquad \left|\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}}\varrho(\theta, z)\right| \le h(z), \qquad \left|\frac{\partial^3}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}}\log\left(\varrho(\theta, z)\right)\right| \le H(z),$$

hold for θ in a neighborhood $N(\tilde{\theta})$ of $\tilde{\theta}$ and for $j, j_1, j_2, j_3 = 1, \ldots, k$ and for all $z \in R_{d-1}$. Next we demand that

$$\int_{R_{d-1}} g(z)\,dz < \infty, \qquad \int_{R_{d-1}} h(z)\,dz < \infty,$$
$$\int_{R_{d-1}} H(z)\,dz < \infty, \qquad \int_{R_{d-1}} H(z)\varrho(\theta, z)\,dz < \infty \text{ for } \theta \in N(\tilde{\theta}),$$

and

$$v_{\lambda,j_1,j_2} := E_\lambda\left(\left(\frac{\partial\log\left(\varrho\left(\lambda, Z\right)\right)}{\partial\lambda_{j_1}}\right)\left(\frac{\partial\log\left(\varrho\left(\lambda, Z\right)\right)}{\partial\lambda_{j_2}}\right)\right) < \infty$$

hold for $j_1, j_2 = 1, \ldots, k$, $Z = Z^{(i)}$. Let the matrix

$$V_\lambda := (v_{\lambda,j_1,j_2})_{j_1,j_2=1,\ldots,k}$$

be invertible. Then there exists a sequence of the maximum likelihood estimators $\hat{\lambda}_m$, which solve $\nabla_\theta \Omega(\theta) = 0$ and for whom the convergence

$$\sqrt{m}(\hat{\lambda}_m - \lambda) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, V_\lambda^{-1}\right)$$

holds for $m \to \infty$. The matrix V_λ is the Fisher information matrix. Thus the estimation is asymptotically efficient.

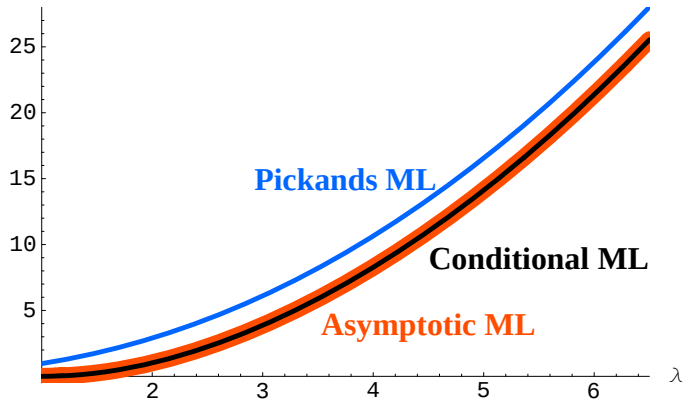
Proof:
See Section 4.2 in Serfling (1980, [63]), Section 33.2 in Cramer (1963, [12]) and Section 2.5 in Witting and Nölle (1970, [75]).

□

Due to the fact that by Theorem 6.2.1 this estimation procedure is asymptotically normal, independent of r, it may seem superior to the estimation procedures in Section 6.1. But with regard to the Asymptotic ML method it has the practical disadvantage of the computation of μ_λ. In contrast to d^*_λ, no analytical expressions of μ_λ are known, even in the logistic case. Thus, the only possibility to overcome this problem is in a numerical way, and this leads to higher runtimes, since we have to maximize (again by numerical means) with regard to λ. This corresponds to the problem of evaluating $\chi_\lambda(r_i, s)$ for $\tilde{\Upsilon}(\lambda)$ in Section 6.1, but, in contrast, we have here only one integral in the function which is to be maximized, whereas previously we had m different integrals for the Conditional ML method.

A mathematical disadvantage are the higher asymptotic variances in comparison to the Asymptotic and the Conditional ML methods, see the next example for a numerical evaluation in the logistic case.

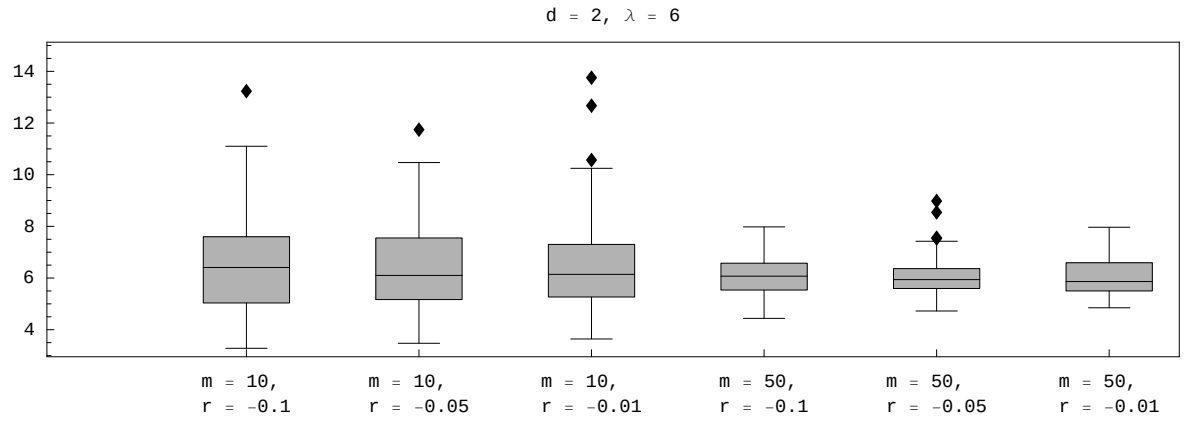
Example 6.2.2
We can compare the three ML methods considered so far by their asymptotic variances, computed in Theorems 6.1.2, 6.1.4 and 6.2.1. In the following plot, these asymptotic variances are plotted for the logistic case with $d = 2$. The asymptotic variance $\gamma^{-1}_{\lambda,1,1}$ of the Conditional ML method is thereby evaluated by the insertion of a high r into $v_{\lambda,1,1,r}$, see Theorem 6.1.4.



We note that the asymptotic variance of the Asymptotic and the Conditional ML seem to be identical. This leads to the conjecture that $v_{\lambda,1,1} = \gamma_{\lambda,1,1} = \lim_{r\to\infty} v_{\lambda,1,1,r}$ with the notations from Theorems 6.1.2 and 6.1.4. The asymptotic variance of the Pickands ML is generally larger.

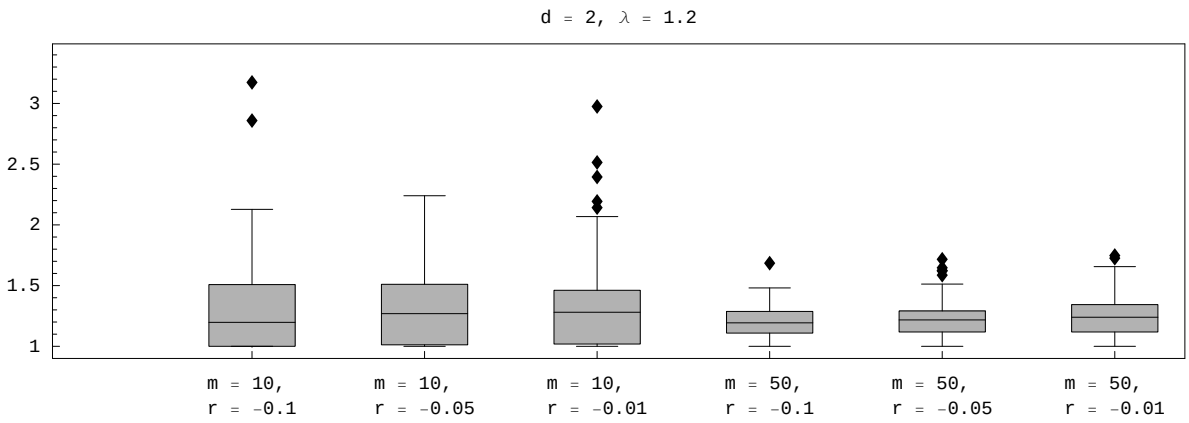
For simulation of the Pickands ML estimator we choose as before the logistic distribution W_λ as the distribution underlying the simulated data. In this example, we use again the functions `NMaximize` and `NIntegrate` of the software package MATHEMATICA, version 5.2, for the maximization of $\Omega(\lambda)$ and the computation of μ_λ.

With Algorithm 3.1.11, $m = 10$ and $m = 50$ data, which fell over the thresholds $r = -0.1, -0.05, -0.01$ were generated for the parameter values $d = 2$ and $\lambda = 6$. In the following graphics, 100 estimations $\hat{\lambda}_m$ for λ are represented by a boxplot as in Examples 6.1.3 and 6.1.7.



As was to be expected, the variance of the estimations reduces, when they are based on more data, and the results are independent of r.

In the next step, λ is set to 1.2. This gives



leading basically to the same results as for $\lambda = 6$. Quite often, however, the estimation is 1, i.e., independence, which is an underestimation leading to misinterpretation, see also Example 6.1.7.

In contrast to the Asymptotic ML method with the angular density, we do not have a bias close to the independence case. So this method, in spite of the higher asymptotic variance, gives more

reliable results. But the price for that is the higher numerical complexity.

$\Diamond$

6.3 Estimation via Relative Frequencies

In this section, we present another way to estimate in parametric multivariate generalized Pareto models. The idea for this method results from the fact that the number of observations, which fall into a certain area, can be asymptotically sufficient for the parameters of the model, see Falk (1998, [20]).

As in Section 6.1, we assume that we have independent and identically distributed random vectors $X_1, \ldots, X_n$ following a GPD $W_{\lambda_1,\ldots,\lambda_k}$ with an in $\lambda_1, \ldots, \lambda_k$ continuously differentiable angular density $l_{\lambda_1,\ldots,\lambda_k}(z)$ and $d^*_{\lambda_1,\ldots,\lambda_k} > 0$. For simplicity of notation, we put again $\lambda := (\lambda_1, \ldots, \lambda_k)$, where $\lambda \in \Lambda \subseteq \mathbb{R}^k$. We, furthermore, assume that the parameter space Λ is an open nonempty subset of $\mathbb{R}^k$.

For $v > d$ put

$$Q_v := \left\{ (z_1, \ldots, z_{d-1}) \in R_{d-1} \,\middle|\, z_i > \frac{1}{v}, i = 1, \ldots, d-1, \sum_{i=1}^{d-1} z_i < 1 - \frac{1}{v} \right\}.$$

The restriction $v > d$ takes care of the fact that the set Q_v is not empty. This corresponds to the condition $r > \frac{d}{s}$ in Lemma 5.1.2. We have $Q_{r,s} = Q_{rs}$ for the set $Q_{r,s}$ considered so far in (5.2) and (5.3). Furthermore, we put

$$B_{r,v} := \left\{ x \in (-\infty, 0)^d | c < -r, z \in Q_v \right\},$$

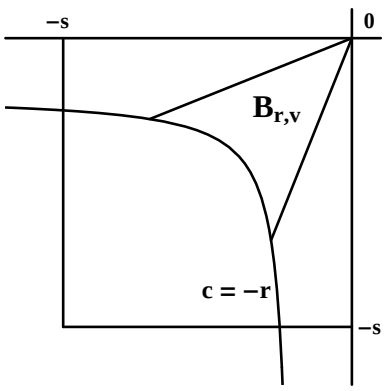
where c and z are the Pickands coordinates of x with regard to Fréchet margins.

In the following, it is crucial that $B_{r,v} \subseteq K_s$ holds for $0 < s < 1$. Thereby s is to be fixed such that W_λ possesses the representation (2.5) on K_s. Next v and r have to be chosen such that the inequality $v < sr$ holds. Then $B_{r,v} \subseteq K_s$ is fulfilled, since for $x \in B_{r,v}$ we have

$$\begin{aligned} ||x||_\infty &= \left|\left| T_F^{-1}(z, c) \right|\right|_\infty = \frac{1}{|c|} \max \left\{ \frac{1}{z_1}, \ldots, \frac{1}{z_{d-1}}, \frac{1}{1 - \sum_{i=1}^{d-1} z_i} \right\} \\ &< \frac{1}{r} \max\{v, \ldots, v, v\} = \frac{v}{r} < s, \end{aligned}$$

and, thus, $x \in K_s$.

The set $B_{r,v}$ can be illustrated in the bivariate case as follows:



The parameter v describes thereby the opening angle of $B_{r,v}$. It is small for v close to d, and it converges to a right angle for $v \to \infty$. Remark that $B_{r,v}$ is the same set as $A_{r,s,\frac{v}{s}}$ in (5.11).

We have

$$P(X_1 \in B_{r,v}) = \int_{-\infty}^{-r} \int_{Q_v} c^{-2} l_\lambda(z)\, dz\, dc = \int_{-\infty}^{-r} c^{-2}\, dc \cdot \underbrace{\int_{Q_v} l_\lambda(z)\, dz}_{=:\chi_\lambda(v)} = \frac{\chi_\lambda(v)}{r}. \tag{6.38}$$

Thereby $\chi_\lambda(r,s) = \chi_\lambda(rs)$ for χ from Definition 5.1.3. With

$$h(B_{r,v}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{B_{r,v}}(X_i)$$

we denote the relative frequency of the occurrence of the event $B_{r,v}$ with the indicator function $\mathbf{1}$ as in (2.18). By the law of large numbers (see for example Serfling (1980, [63]), Theorem 1.8B), this relative frequency converges for $n \to \infty$ to the probability of occurrence $P(X_1 \in B_{r,v}) = \frac{\chi_\lambda(v)}{r}$.

Choose now v_j, $j = 1, \ldots, k$, such that $d < v_1 < \ldots < v_k < sr$. With these define the function $\psi : \Lambda \to [0,1]^k$ by

$$\psi(\lambda) = \begin{pmatrix} \frac{\chi_\lambda(v_1)}{r} \\ \vdots \\ \frac{\chi_\lambda(v_k)}{r} \end{pmatrix}. \tag{6.39}$$

During the following we assume that ψ is injective.

We estimate the parameter λ by $\hat{\lambda}$ such that

$$\psi(\hat{\lambda}) = \begin{pmatrix} h(B_{r,v_1}) \\ \vdots \\ h(B_{r,v_k}) \end{pmatrix} \tag{6.40}$$

holds. For this we have to assume that

$$\begin{pmatrix} h(B_{r,v_1}) \\ \vdots \\ h(B_{r,v_k}) \end{pmatrix} \in \mathrm{Im}(\psi), \tag{6.41}$$

with $\mathrm{Im}(\psi)$ being the image of the function ψ, otherwise define $\hat{\lambda}$ arbitrary. For large n, this holds by the law of large numbers with a probability converging to 1, i.e.,

$$P\left(\begin{pmatrix} h(B_{r,v_1}) \\ \vdots \\ h(B_{r,v_k}) \end{pmatrix} \notin \mathrm{Im}(\psi) \right) \to_{n\to\infty} 0. \tag{6.42}$$

if $\psi(\lambda)$ is no boundary element of the image, which must be assumed. With real data sets, it can and will happen that (6.41) does not hold, even if all other assumptions are fulfilled. We will see this in the simulations in Example 6.3.8. One should then change the parameters r and v if possible, see Example 6.3.5 and Section 6.4, where we will try to find appropriate r and v.

By the law of large numbers we get

$$\psi(\hat{\lambda}) = \begin{pmatrix} h(B_{r,v_1}) \\ \vdots \\ h(B_{r,v_k}) \end{pmatrix} \to_{n\to\infty} \begin{pmatrix} \frac{\chi_\lambda(v_1)}{r} \\ \vdots \\ \frac{\chi_\lambda(v_k)}{r} \end{pmatrix} = \psi(\lambda).$$

The function ψ is, by the assumptions on l_λ, continuously differentiable and, thus, we conclude from $\psi(\hat{\lambda}) \to_{n\to\infty} \psi(\lambda)$ and the injectivity of ψ, the convergence $\hat{\lambda} \to \lambda$ for $n \to \infty$.

We also want to show the asymptotic normality of the estimator. Due to the continuous differentiability of ψ noted above, this function possesses a Taylor series expansion of first order (see Theorems 168.4 and 168.5 in Heuser (1998, [38])), i.e., there exists a $\vartheta \in \mathbb{R}^k$ with $||\vartheta - \lambda||_\infty < ||\hat{\lambda} - \lambda||_\infty$, such that

$$\psi(\hat{\lambda}) = \psi(\lambda) + J_\psi(\vartheta)\left(\hat{\lambda} - \lambda\right)$$

holds, where $J_\psi(\vartheta)$ is the Jacobian matrix of ψ at ϑ, as in Lemma 2.1.5.

We conclude

$$J_\psi(\vartheta)\left(\hat{\lambda} - \lambda\right) = \psi(\hat{\lambda}) - \psi(\lambda) = \begin{pmatrix} h(B_{r,v_1}) \\ \vdots \\ h(B_{r,v_k}) \end{pmatrix} - \begin{pmatrix} \frac{\chi_\lambda(v_1)}{r} \\ \vdots \\ \frac{\chi_\lambda(v_k)}{r} \end{pmatrix} \longrightarrow_{n\to\infty} 0 \tag{6.43}$$

by the law of large numbers. Since $\hat{\lambda} \to_{n\to\infty} \lambda$, we also have $J_\psi(\vartheta) \to_{n\to\infty} J_\psi(\lambda)$ due to the continuous differentiability of ψ.

Lemma 6.3.1
Let the random vector X have distribution W_λ with $d_\lambda^* > 0$, and choose $v_1, \ldots, v_k$ with $d < v_1 < \ldots < v_k < sr$. Define the k–dimensional random vector

$$\Xi := \begin{pmatrix} \Xi_1 \\ \vdots \\ \Xi_k \end{pmatrix} := \begin{pmatrix} \mathbf{1}_{B_{r,v_1}}(X) - \frac{\chi_\lambda(v_1)}{r} \\ \vdots \\ \mathbf{1}_{B_{r,v_k}}(X) - \frac{\chi_\lambda(v_k)}{r} \end{pmatrix}.$$

It has expectation 0 and covariance matrix $\Sigma = (\varsigma_{ij})_{1\le i,j\le k}$ with

$$\varsigma_{ij} := \frac{\chi_\lambda\left(v_{\min(i,j)}\right)}{r} - \frac{\chi_\lambda(v_i)\chi_\lambda(v_j)}{r^2}, \qquad 1 \le i,j \le k.$$

Proof:
We have for $i = 1, \ldots, k$

$$E(\Xi_i) = E\left(\mathbf{1}_{B_{r,v_i}}(X) - \frac{\chi_\lambda(v_i)}{r}\right) = P(X \in B_{r,v_i}) - \frac{\chi_\lambda(v_i)}{r} \overset{(6.38)}{=} 0,$$

and for $1 \le i,j \le k$

$$\varsigma_{ij} \quad = \quad E\left(\Xi_i \Xi_j\right) = E\left(\left(\mathbf{1}_{B_{r,v_i}}(X) - \frac{\chi_\lambda(v_i)}{r}\right)\left(\mathbf{1}_{B_{r,v_j}}(X) - \frac{\chi_\lambda(v_j)}{r}\right)\right)$$

$$
\begin{aligned}
&= E\left(\mathbf{1}_{B_{r,v_i}}(X)\cdot\mathbf{1}_{B_{r,v_j}}(X) - \mathbf{1}_{B_{r,v_i}}(X)\frac{\chi_\lambda(v_j)}{r} - \mathbf{1}_{B_{r,v_j}}(X)\frac{\chi_\lambda(v_i)}{r} + \frac{\chi_\lambda(v_i)}{r}\cdot\frac{\chi_\lambda(v_j)}{r}\right)\\
&= E\left(\mathbf{1}_{B_{r,v_i}\cap B_{r,v_j}}(X)\right) - P(X\in B_{r,v_i})\frac{\chi_\lambda(v_j)}{r} - P(X\in B_{r,v_j})\frac{\chi_\lambda(v_i)}{r} + \frac{\chi_\lambda(v_i)\cdot\chi_\lambda(v_j)}{r^2}\\
&\overset{(6.38)}{=} E\left(\mathbf{1}_{B_{r,\min(v_i,v_j)}}(X)\right) - \frac{\chi_\lambda(v_i)\cdot\chi_\lambda(v_j)}{r^2} = P\left(X\in B_{r,\min(v_i,v_j)}\right) - \frac{\chi_\lambda(v_i)\cdot\chi_\lambda(v_j)}{r^2}\\
&\overset{(6.38)}{=} \frac{\chi_\lambda\left(v_{\min(i,j)}\right)}{r} - \frac{\chi_\lambda(v_i)\chi_\lambda(v_j)}{r^2},
\end{aligned}
$$

since the v_i are ordered by definition.

□

Remark 6.3.2
The variance of the random variables Ξ_i from Lemma 6.3.1 is

$$\mathrm{Var}(\Xi_i) = \frac{\chi_\lambda(v_i)}{r}\left(1 - \frac{\chi_\lambda(v_i)}{r}\right),$$

and the matrix Σ vanishes for $r\to\infty$ with fixed λ and v.

◊

Lemma 6.3.3
Let ψ be injective and $\psi(\lambda)\notin\partial\mathrm{Im}(\psi)$ hold. Then we have with $t\in\mathbb{R}^k$ and the matrix Σ from Lemma 6.3.1,

$$P\left(\sqrt{n}J_\psi(\vartheta)(\hat\lambda-\lambda)\le t\right)\longrightarrow_{\mathcal{D}}\mathcal{N}(0,\Sigma)$$

for $n\to\infty$, where the inequality is meant componentwise.

Proof:
By the independence and the identical distribution of the X_i, as well as (6.43), Lemma 6.3.1 and the multivariate central limit theorem (see for example Gänssler and Stute (1977, [30]), Theorem 8.8.1), we have

$$
\begin{aligned}
&P\left(\sqrt{n}J_\psi(\vartheta)(\hat\lambda-\lambda)\le t, \begin{pmatrix} h(B_{r,v_1})\\ \vdots\\ h(B_{r,v_k})\end{pmatrix}\in\mathrm{Im}(\psi)\right)\\
&\overset{(6.43)}{=} P\left(\sqrt{n}\left(\begin{pmatrix} h(B_{r,v_1})\\ \vdots\\ h(B_{r,v_k})\end{pmatrix} - \begin{pmatrix} \frac{\chi_\lambda(v_1)}{r}\\ \vdots\\ \frac{\chi_\lambda(v_k)}{r}\end{pmatrix}\right)\le t\right)\\
&= P\left(\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{B_{r,v_1}}(X_i) - \frac{\chi_\lambda(v_1)}{r}\\ \vdots\\ \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{B_{r,v_k}}(X_i) - \frac{\chi_\lambda(v_k)}{r}\end{pmatrix}\le t\right)\\
&= P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\begin{pmatrix} \mathbf{1}_{B_{r,v_1}}(X_i) - \frac{\chi_\lambda(v_1)}{r}\\ \vdots\\ \mathbf{1}_{B_{r,v_k}}(X_i) - \frac{\chi_\lambda(v_k)}{r}\end{pmatrix}\le t\right)\\
&\longrightarrow_{\mathcal{D}} \mathcal{N}(0,\Sigma)
\end{aligned}
$$

for $n \to \infty$. Thus, by (6.42)

$$\begin{aligned} & P\left(\sqrt{n} J_\psi(\vartheta)(\hat{\lambda} - \lambda) \le t\right) \\ = \quad & P\left(\sqrt{n} J_\psi(\vartheta)(\hat{\lambda} - \lambda) \le t, \begin{pmatrix} h(B_{r,v_1}) \\ \vdots \\ h(B_{r,v_k}) \end{pmatrix} \in \mathrm{Im}(\psi)\right) \\ & \underbrace{+ P\left(\sqrt{n} J_\psi(\vartheta)(\hat{\lambda} - \lambda) \le t, \begin{pmatrix} h(B_{r,v_1}) \\ \vdots \\ h(B_{r,v_k}) \end{pmatrix} \notin \mathrm{Im}(\psi)\right)}_{\to_{n\to\infty} 0} \\ \longrightarrow_{\mathcal{D}} \quad & \mathcal{N}(0, \Sigma) \end{aligned}$$

for $n \to \infty$.

□

Theorem 6.3.4
Under the assumptions of Lemma 6.3.3, we have

$$P\left(\sqrt{n}\left(\hat{\lambda} - \lambda\right) \le t\right) \longrightarrow_{\mathcal{D}} \mathcal{N}\left(0, \left(J_\psi(\lambda)\right)^{-1} \Sigma \left(J_\psi^T(\lambda)\right)^{-1}\right)$$

for $n \to \infty$.

Proof:
Since the function ψ is injective by assumption, the inverse of its Jacobian matrix exists at the point λ. The assertion follows immediately from Lemma 6.3.3 and Lemma 6.1.1.

□

Example 6.3.5
In this example, $X_1, \ldots, X_n$ follow a logistic distribution with parameter $\lambda > 1$. The angular density $l_\lambda(z)$ is then continuously differentiable, and the parameter space $\Lambda = (1, \infty) \subseteq \mathbb{R}$ is an open set. We choose now r, $s > 0$ and $d < v < sr$. Then we have for the function $\psi : (1, \infty) \to \mathbb{R}$ the representation

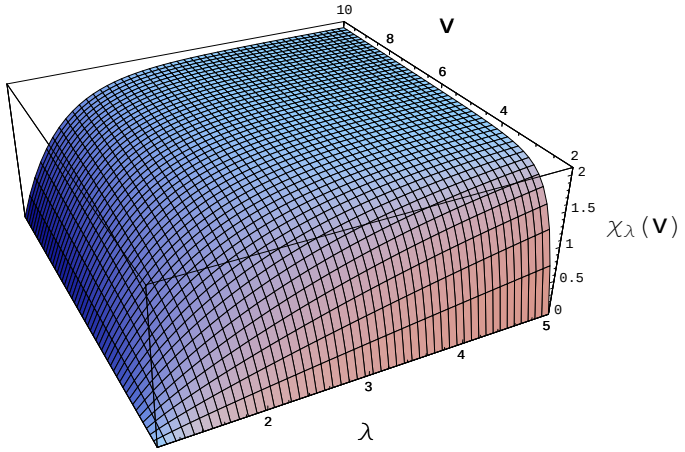
$$\psi(\lambda) = \frac{1}{r}\chi_\lambda(v) = \frac{1}{r}\int_{Q_v} l_\lambda(z)\, dz$$

with the angular density l_λ known from Example 2.3.10. If this function is strongly monotonous in λ, the other two conditions, which ψ should fulfill (injectivity and every image point is an inner point), are given. For the property of monotonicity in λ the parameter r does not play a role. An analytical proof of the monotonicity is very difficult due to the complex structure of l_λ. Numerical experiments, such as the following figure (done again with the MATHEMATICA function `NIntegrate`), however, lead to the conjecture that the condition of monotonicity is fulfilled. In the following picture, the function

$$\chi_\lambda(v) = \int_{Q_v} l_\lambda(z)\, dz$$

was plotted for the parameters $v \in (2, 10)$ and $\lambda \in (1, 5)$ for the case $d = 2$. With v fixed the

function is increasing in λ.

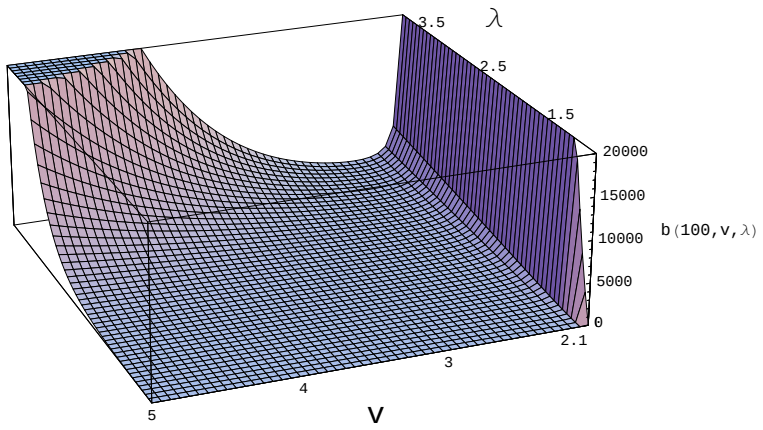


Remark that due to the structure of the function, small fluctuations in the data, leading to slightly different relative frequencies, can have large effects on the estimator, since one computes it by taking the inverse function. This property is also reflected in the high variance for certain parameters below.

The variance of the normal approximation given in Theorem 6.3.4 for this case is

$$\frac{\psi(\lambda)(1-\psi(\lambda))}{(\psi'(\lambda))^2} = \frac{\chi_\lambda(v)(r-\chi_\lambda(v))}{\left(\frac{\partial}{\partial\lambda}\chi_\lambda(v)\right)^2} =: b(r,v,\lambda).$$

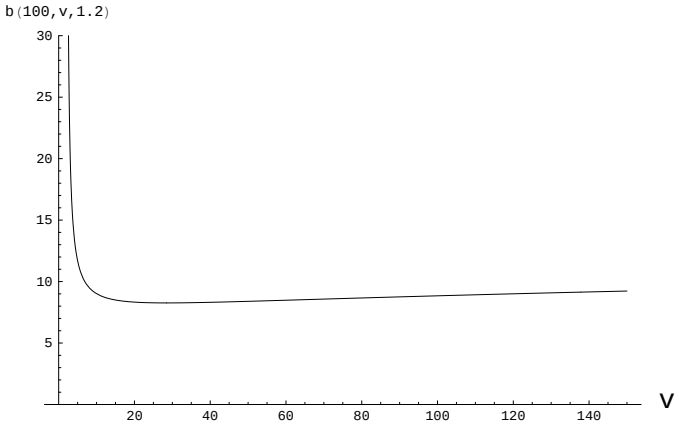
Note that $\chi_\lambda(v) \leq d$ and $r > \frac{d}{s}$ with $s < 1$, by the general assumptions of this estimation procedure, certify that the variance above is always positive. The parameter r is usually chosen by the user and expresses, which observations are considered to be extreme. We note that by the above formula a smaller r leads to lower variance. The dependence of this variance on λ and v is much more complex. In the following graphic, we plot the function $b(100, v, \lambda)$, again depending on v and λ, for $d = 2$.



When v is fixed the variance increases with λ such that we expect the procedure to deliver good results for small λ (close to independence), whereas we expect higher variability for larger λ.

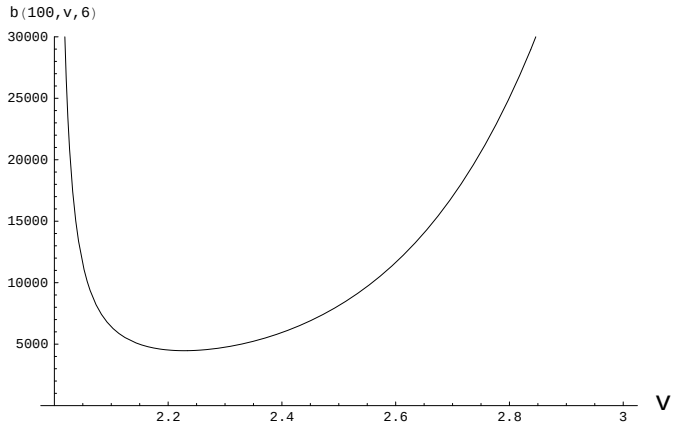
Keeping λ fixed, the variance seems to converge for $v \to d$ and $v \to \infty$ to ∞ respectively. In between there seems to exist a v with minimal variance. This would be the optimal v, on which the estimation procedure should be based. Since this optimum, however, depends directly on the parameter which is to be estimated it cannot be computed in practice. In Section 6.4, a method will be presented, with which approximations of this optimal v can be gained.

Here we restrict ourselves to computing the optimal v for the parameters $\lambda = 1.2$ and $\lambda = 6$, which we consider in our examples. Next the function $b(100, v, 1.2)$ is plotted for $v \in (2, 150)$.



Through numerical minimization (analogously to Section 6.1, the function `NMinimize` from MATHEMATICA, version 5.2, was used) we find as optimal v the value 28.51. Observe that $v < sr$ must hold, which is fulfilled here. In situations, where this is not the case, sr should be considered as the best value.

In the same manner we plot the function $b(100, v, 6)$ for $v \in (2, 3)$.



Here the optimal v is approximately 2.233.

Since $r \gg \chi_\lambda(v)$ and, thus, $r - \chi_\lambda(v) \approx r$, one gets similarly good approximations by just minimizing the function

$$\frac{\chi_\lambda(v)}{\left(\frac{\partial}{\partial\lambda}\chi_\lambda(v)\right)^2}.$$

It does not depend on r anymore and is preferable for computational and efficiency reasons. The approximations for v gained in this manner for the above examples are 27.96 and 2.227. The values of the function b at these points, and thus the variance of the corresponding estimation procedure, are numerically identical, i.e., the computer returns the same values of b for both approaches.

$\Diamond$

Remark 6.3.6
In the case $k = 1$, the covariance matrix $(J_\psi(\lambda))^{-1} \Sigma \left(J_\psi^T(\lambda)\right)^{-1}$ can be computed and minimized with regard to v as in Example 6.3.5. But a minimization of the covariance matrix in a proper sense with regard to $v = (v_1, \ldots, v_k)$ is also necessary for the case $k > 1$. A common criterion is the minimization of the determinant, see Section 4.1.2 in Serfling (1980, [63]). It is known from linear algebra (see for example Chapter XI of Lang (1966, [48])) that for symmetric matrices the determinant is the product of the eigenvalues, i.e., one has to minimize the product of the eigenvalues of $(J_\psi(\lambda))^{-1} \Sigma \left(J_\psi^T(\lambda)\right)^{-1}$ with regard to v.

$\Diamond$

Remark 6.3.7
The definition of $B_{r,v}$ and especially Q_v given above should be modified if one examines models like the asymmetric logistic model. With this definition of $B_{r,v}$ and Q_v it may be possible that no injective function ψ can be found. Then the set $B_{r,v}$ should be modified in such a manner that it is not anymore symmetric to the bisecting line of the negative quadrant.

$\Diamond$

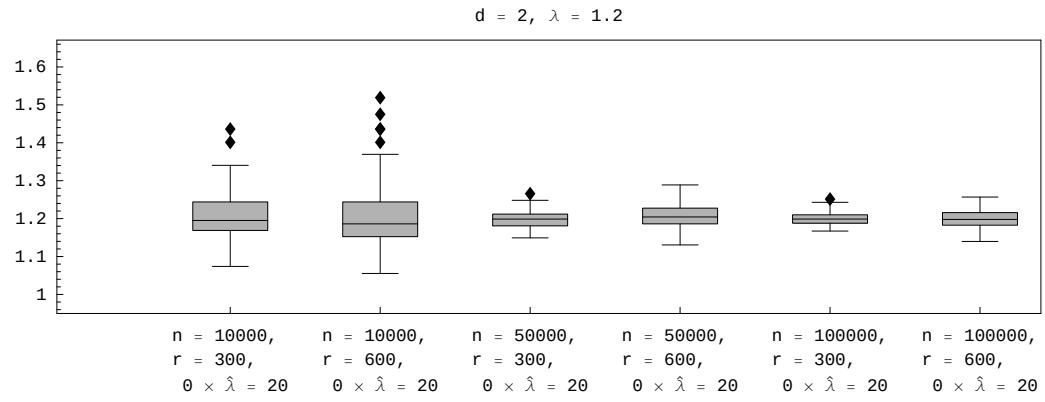
Example 6.3.8
With the help of Algorithm 3.3.4 it is possible to evaluate the estimation procedure presented in this section with simulated data. As an example we use again the logistic distribution. We also want to check with these simulated data the insights gained through the considerations in Example 6.3.5 concerning the reliability of the procedure. Example 6.3.5 suggests that the procedure works well in cases of low dependence, whereas in cases of high dependence, it should work less well and depend highly on the choice of the parameter v.

With real data sets, we often have the problem that condition (6.41) is not fulfilled. In the logistic case ψ increases and we have $\psi(\lambda) \to 0$ for $\lambda \to 1$. Since large values of $\psi(\lambda)$ correspond to large values of λ, due to the monotonicity of ψ, a violation of (6.41) suggests a high degree of dependence. Actually an estimation of $\hat{\lambda} = \infty$ would then be reasonable. Since this would lead to a difficult graphical representation of the results of the estimation procedure below, we set in these cases $\hat{\lambda}$ to 20 during the simulations.

The evaluation of the simulations is done as in Examples 6.1.3, 6.1.7 and 6.2.2 by creating boxplots of 100 computed estimators for a fixed set of parameters. In addition we note how often the case $\hat{\lambda} = 20$ happened, i.e., how often the condition (6.41) was violated. As in Example 6.3.5 we only examine the two-dimensional case here.

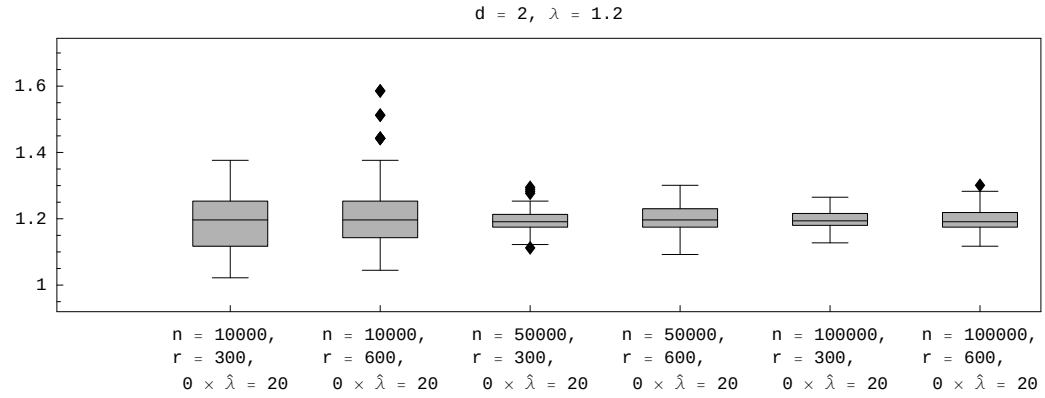
We consider the sample sizes of $n = 10000$, 50000, 100000 as well as $r = 300$ and $r = 600$. Remark that only a small fraction of these seemingly gigantic data sets are extreme observations, on which the estimations are actually based. In a first simulation the underlying parameter is

set to 1.2 and $v = 20$ is chosen.



One realizes that the condition (6.41) was always fulfilled and that the estimation aimed at the right parameter on average. As was to be expected the fluctuation of the estimator decreases with growing n. It increases, however, with growing r. This is a phenomenon that was predicted in Example 6.3.5.

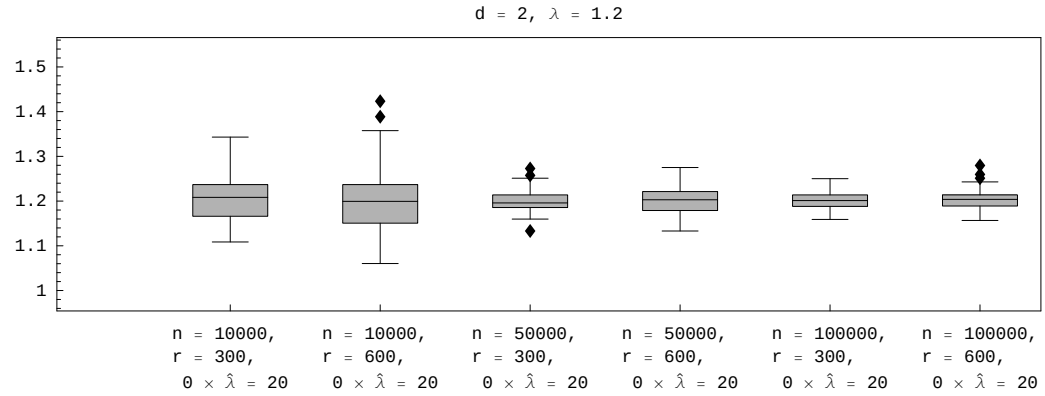
Next we reduce v to 3, i.e., we diminish the set $B_{r,v}$, whose relative frequency is the base of our estimation procedure, via the opening angle.



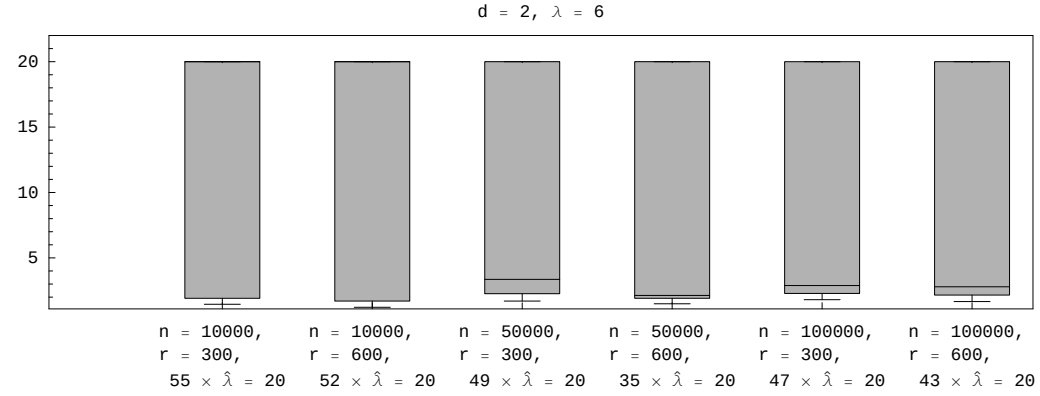
The estimator still works quite well.

The optimal v for the case $\lambda = 1.2$, was according to Example 6.3.5, at $v \approx 27.96$. With this parameter the estimation gives the following results, which do not differ much from those gained

above.

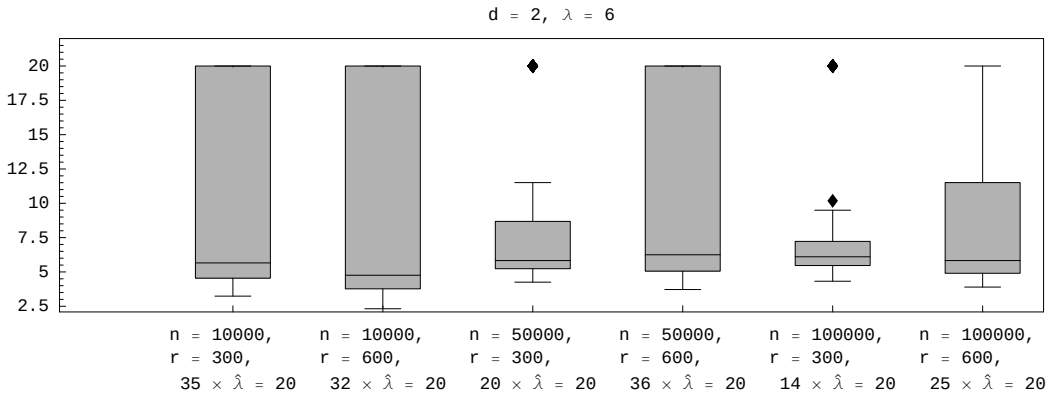


Next we examine the case $\lambda = 6$ for $v = 20$.



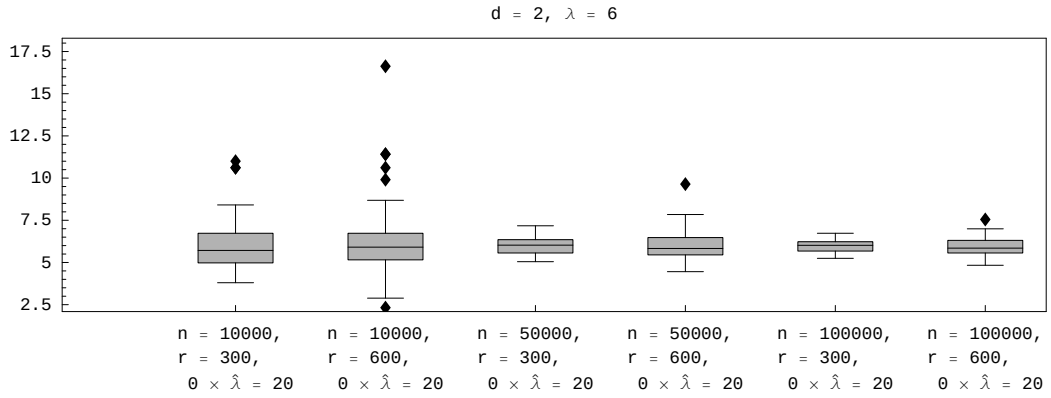
Observing the corresponding frequencies of $\hat{\lambda} = 20$, we immediately see that for about a third to a half of the data, no estimator could be computed at all. This leads of course to an extremely high variance of the estimator, which would be ∞ if $\hat{\lambda}$ would have been set to ∞ in these cases. So in this case the estimator is useless.

We reduce the opening angle again to $v = 3$.



Here the number of clearly mistaken results decreases, however, condition (6.41) is, still, too often violated to consider this a useful method.

According to Example 6.3.5 the optimal v lies at about 2.227. With $v = 3$ one is already quite close to this value, nevertheless, we gain a considerable improvement by taking the optimal v, as the next graphic shows.



In no case condition (6.41) was violated here.

We have seen that the procedure works quite well close to the independence case even with different choices of v, close to the dependence case the procedure, however, reacts very sensitively to changes of v. This is exactly reciprocal to the asymptotic maximum likelihood estimation in Section 6.1, which reacted in the independence case very sensitively to changes of the parameter r. This leads to the idea to combine both approaches to a common estimation procedure which is to be presented in the next section.

◊

6.4 Comparison of the Estimation Procedures

The results of the previous sections suggest a combination of some of the estimation procedures.

We begin with a simple iteration.

Algorithm 6.4.1

1. *Determine $\lambda^{(0)}$ by the Asymptotic ML method, maximizing Υ.*
2. *Determine $v \in (d, rs]^k$ such that the determinant of the covariance matrix of the estimator via relative frequencies with underlying parameter $\lambda^{(0)}$ becomes minimal.*
3. *Determine the estimator $\lambda^{(1)}$ through the procedure via relative frequencies with parameter v.*
4. *Return $\lambda^{(1)}$.*

◊

The first estimation of $\lambda^{(0)}$ can also be done by some other estimation procedures, but maximizing Υ by the Asymptotic ML method is typically the computationally fastest way. And since

we only want a reasonable starting value the Asymptotic ML method suffices, since we have seen in Example 6.3.8 that the estimation procedure is robust to choices of v in cases close to independence, where the Asymptotic ML is biased. In cases of high dependence, Asymptotic ML is not biased and, thus, also a reasonable starting value.

There is also the possibility to iterate this procedure, this leads to a multiple iteration.

Algorithm 6.4.2
Let $i = 1$, $\eta > 0$ small, and $I \in \mathbb{N}$.

1. *Determine $\lambda^{(0)}$ by the Asymptotic ML method, maximizing Υ.*
2. *Determine $v^{(i)} \in (d, rs]^k$ such that the determinant of the covariance matrix of the estimator via relative frequencies with underlying parameter $\lambda^{(i-1)}$ becomes minimal.*
3. *Determine the estimator $\lambda^{(i)}$ through the procedure via relative frequencies with parameter $v^{(i)}$.*
4. *If $\left|\frac{\lambda^{(i-1)}-\lambda^{(i)}}{\lambda^{(i-1)}}\right| \leq \eta$ or $i = I$, return $\lambda^{(i)}$, else increase i by 1 and go to 2.*

$\Diamond$

The choice of the break off parameters η and I depends on the underlying problem and cannot be specified in general.

We now want to compare the procedures presented in this chapter for the estimation of the model parameters by using them with comparable parameters on identical data sets. We examine the following procedures:

- Asymptotic ML: the parameters are estimated by maximizing Υ.
- Conditional ML: the parameters are estimated by maximizing $\tilde{\Upsilon}$.
- Pickands ML: the parameters are estimated by maximizing Ω.
- Simple Iteration: the parameters are estimated by Algorithm 6.4.1.
- Multiple Iteration: the parameters are estimated by Algorithm 6.4.2, where $\eta = 0.01$ and $I = 10$.

Since Asymptotic ML, Conditional ML, Simple and Multiple Iteration are based on Pickands coordinates with regard to Fréchet margins, whereas Pickands ML is based on standard Pickands coordinates, the estimation procedures use different (random) sample sizes. Thus a comparison with identical parameters becomes difficult. Especially the extreme areas are different in those cases. Thus the choice of the respective threshold must be adjusted such that roughly the same number of observations falls over the respective thresholds. This we did by

$$r_P = -1.5 \cdot \frac{d^2}{r_F}, \tag{6.44}$$

where r_F is the chosen threshold for the Pickands coordinates with regard to Fréchet margins, and r_P the threshold for standard Pickands coordinates. It turns out that with this choice, the Pickands ML method gets slightly more data in cases with high dependence and slightly less

data than the other methods in cases close to independence. This should be taken into account when interpreting the simulations. In practice one usually chooses the thresholds by considering the area, which is thought to be extreme.

Before we compute the simulations for the logistic case, we want to add some short considerations on the computational efficiency of the methods.

For the Asymptotic ML method we only have to maximize a quickly evaluable function, which is usually done with an iteration procedure and is very efficient with the corresponding software packages. This is the only method which is usable in high dimensions (≥ 5) from an efficiency point of view, since no numerical evaluation of an integral has to be done if an analytical expression of d_λ^* is known.

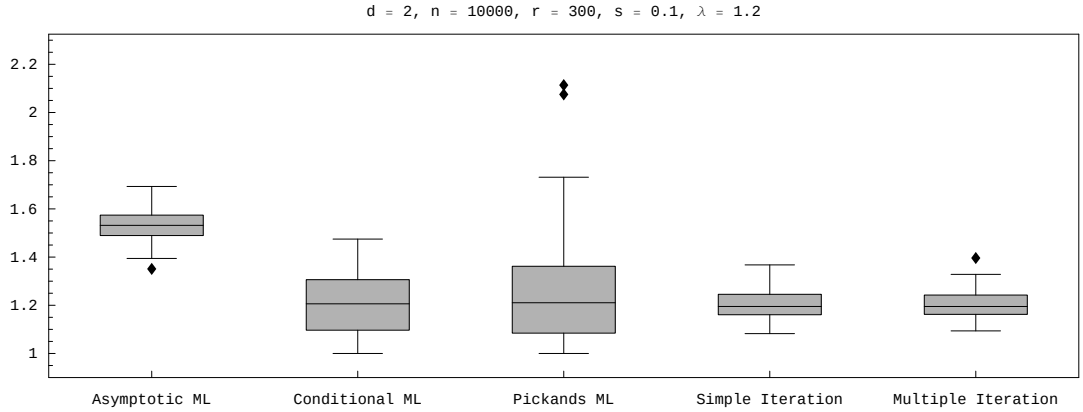
For the Conditional ML method we also have to maximize a function, which contains, however, m integrals, which are in practice only computable by numerical methods. In every step of the maximization procedure, m integrals have to be evaluated, which makes this method very ineffective.

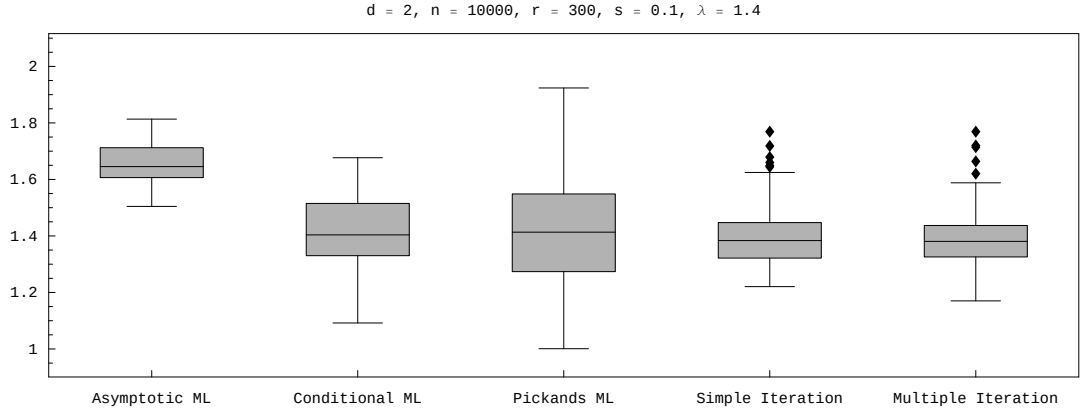
For the Pickands ML method, we have to maximize a function with one numerical integral in general. This is the same effort as for the Asymptotic ML method if d_λ^* is unknown. But in most practical cases, Asymptotic ML is faster due to the known analytical representation of d_λ^*.

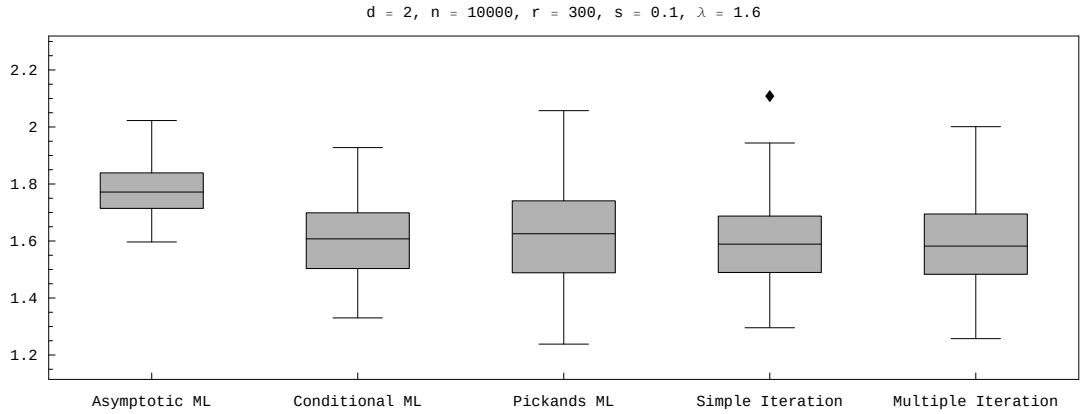
For the Simple Iteration we have to do an Asymptotic ML estimation first, then, to determine v, we have to maximize a function, which contains two numerical integrals and one numerical derivative. Subsequently, we have to solve an equation numerically which contains one integral. Thus this method is more costly than the Asymptotic ML and the Pickands ML method, but a lot less numerical integrals have to be determined than with the Conditional ML method. So this procedure is more efficient than Conditional ML and is from this point of view preferable to it.

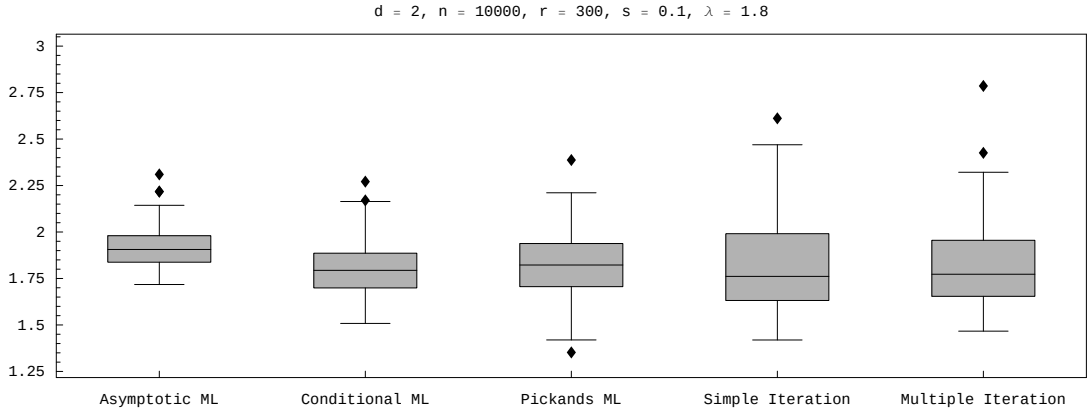
Since for the Multiple Iteration the Simple Iteration has to be executed repeatedly, it is more costly than this but most of the time, still, less costly than Conditional ML.

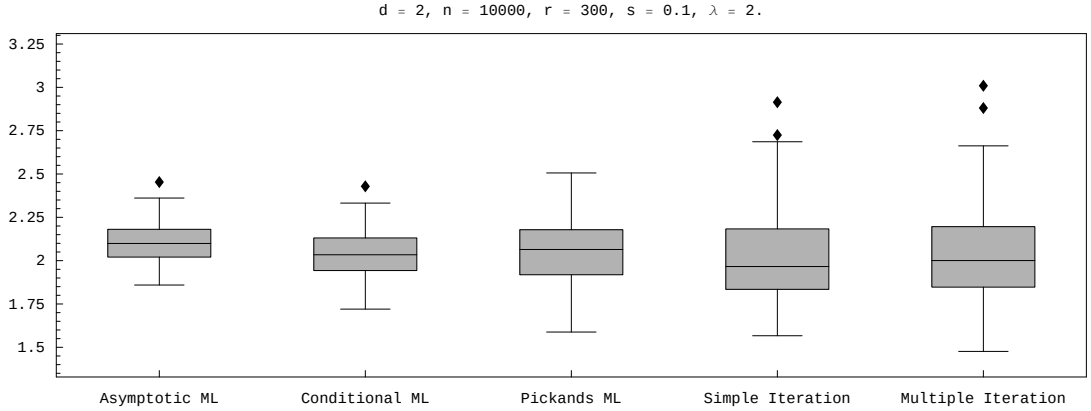
Example 6.4.3
The five procedures introduced above are now compared by their results on identical data sets. For each setting of parameters, i.e, for each of the graphics below, 100 data sets, whose observations follow a logistic GPD with parameter λ, are generated. For every procedure the estimator for the 100 data sets was computed. These estimations are summarized in the following plots as usual as boxplots. At first the parameters were set to $d = 2$, $n = 10000$, $r = 300$ and $s = 0.1$, the threshold for the Pickands ML methods was computed as in (6.44). For these parameter settings the behavior of the estimators was examined for $\lambda = 1.2$, 1.4, 1.6, 1.8, 2, 3, 4, 5 and 6.

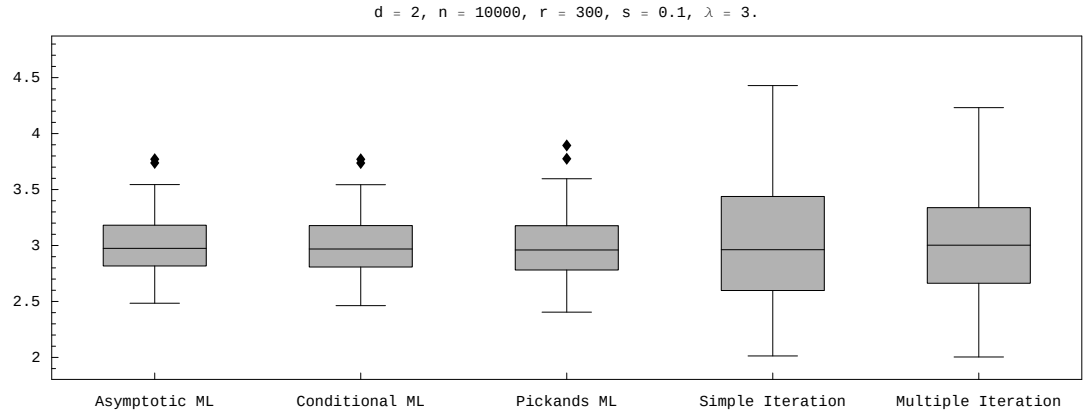

d = 2, n = 10000, r = 300, s = 0.1, λ = 1.2
2.2
2
1.8
1.6
1.4
1.2
1
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration

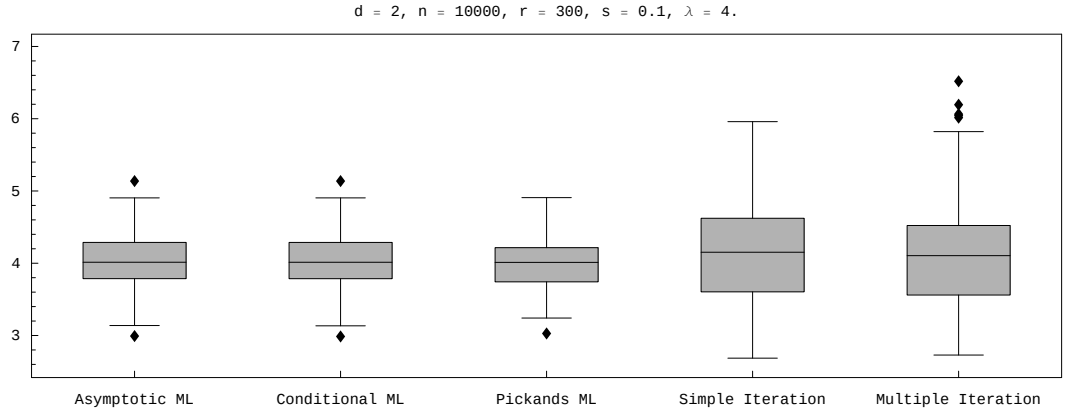

d = 2, n = 10000, r = 300, s = 0.1, λ = 1.4
2
1.8
1.6
1.4
1.2
1
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration

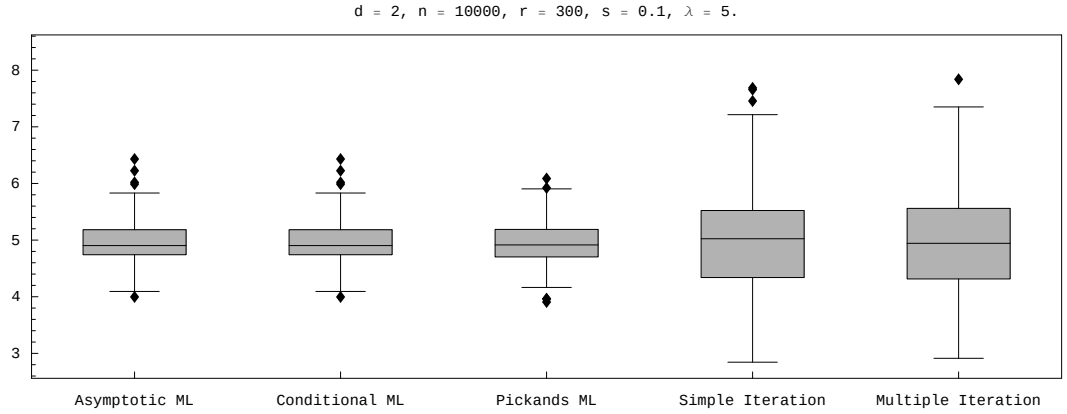

d = 2, n = 10000, r = 300, s = 0.1, λ = 1.6
2.2
2
1.8
1.6
1.4
1.2
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration

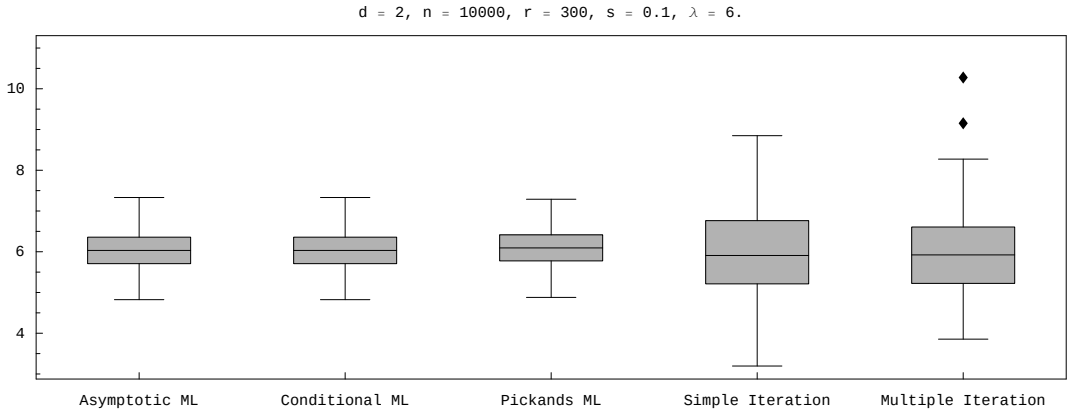

d = 2, n = 10000, r = 300, s = 0.1, λ = 1.8
3
2.75
2.5
2.25
2
1.75
1.5
1.25
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration


d = 2, n = 10000, r = 300, s = 0.1, λ = 2.
3.25
3
2.75
2.5
2.25
2
1.75
1.5
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration


d = 2, n = 10000, r = 300, s = 0.1, λ = 3.
4.5
4
3.5
3
2.5
2
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration



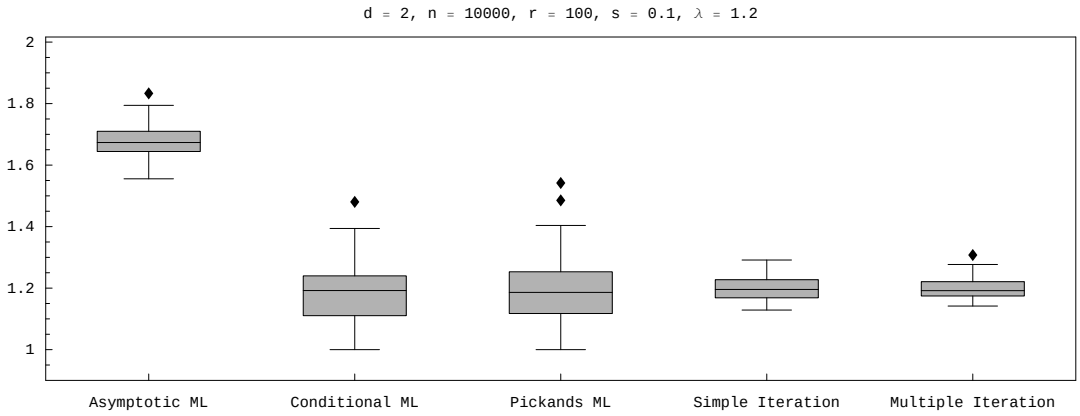


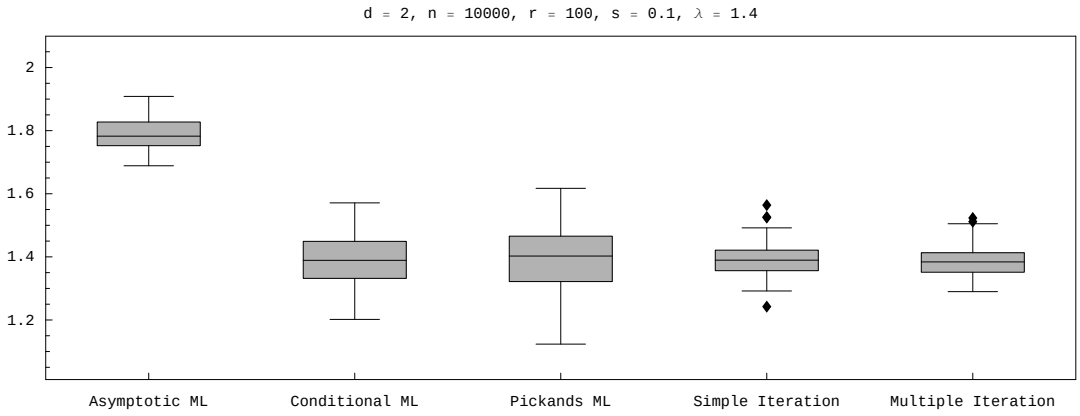


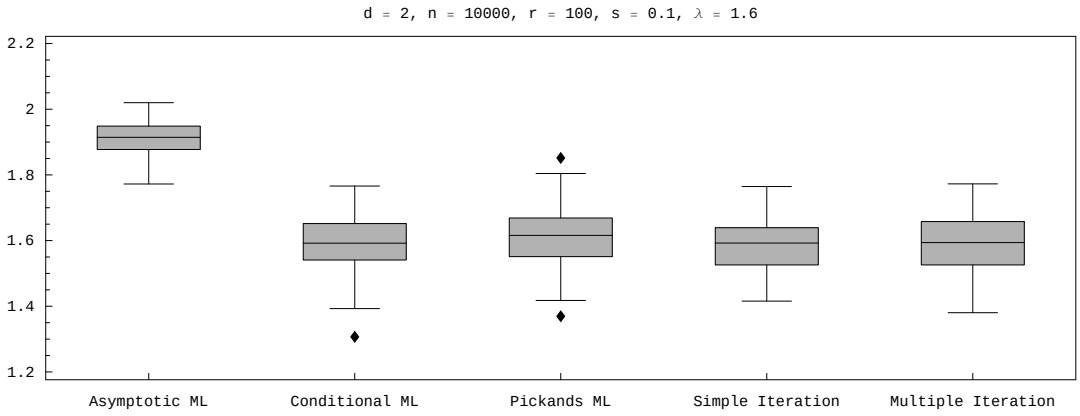
First of all we observe that the two iteration methods hardly differ in their behavior so that the most efficient method of Simple Iteration should be used, since there is not much gained in other iteration steps. Also the Conditional ML and the Pickands ML do not show big differences. So here the more efficient Pickands ML is preferable. Very noticeable is the bias of the efficient Asymptotic ML method for small λ. In these cases an additional correction by another procedure is necessary to reduce the bias.

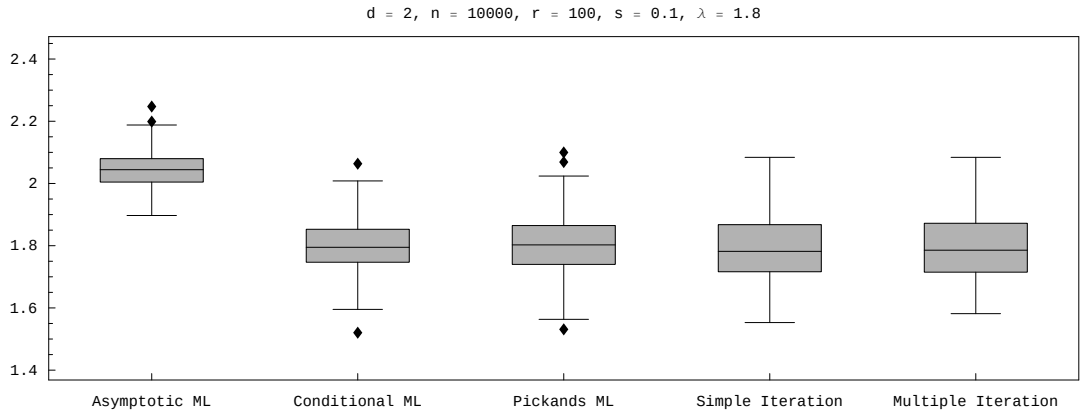
For small λ, the iteration methods are better, i.e., the estimations have smaller variances, for large λ the ML methods. We can read from the graphics that with these parameter values one should add a correction to the Asymptotic ML estimator if the original estimation is smaller than 2. Also an iteration procedure should used if the Conditional ML or the Pickands ML return an estimate smaller than 2 to reduce the variance of the estimator. This correction is most reasonably done with the Simple Iteration.

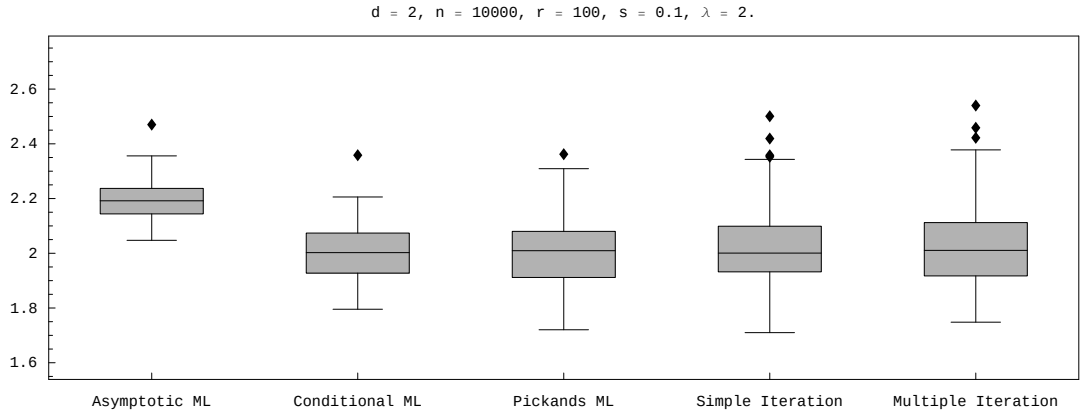
Next we put $r = 100$.

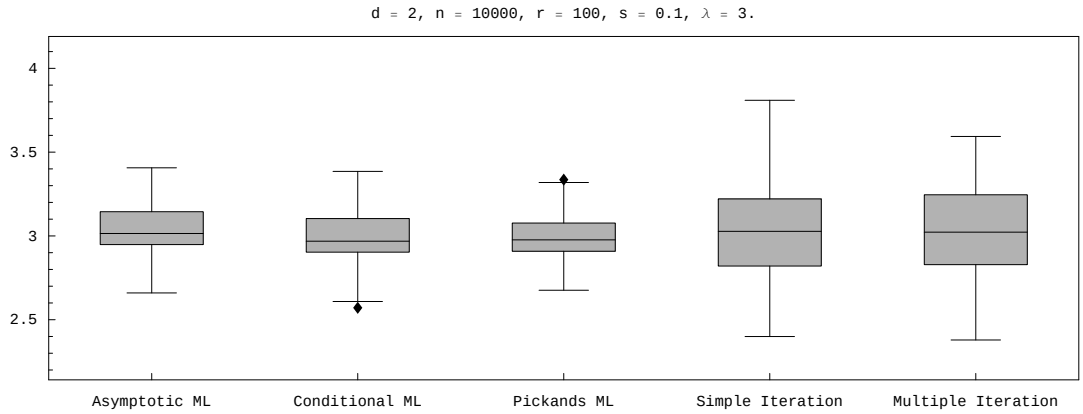


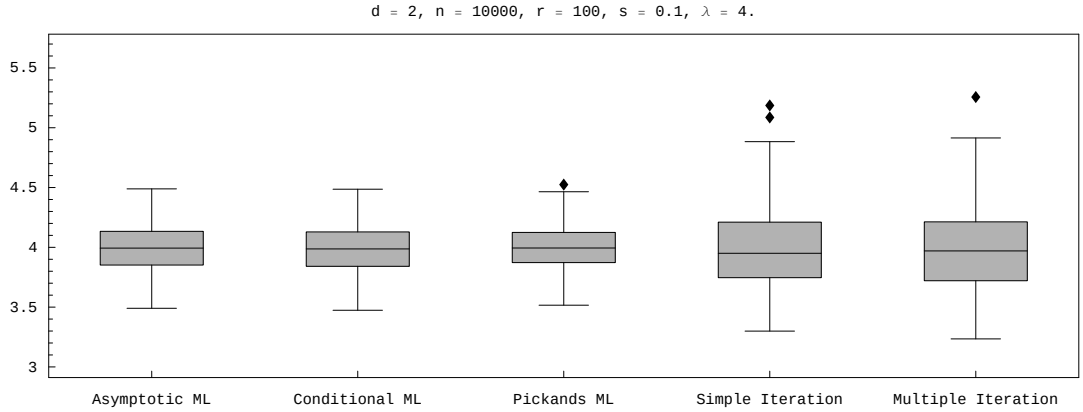


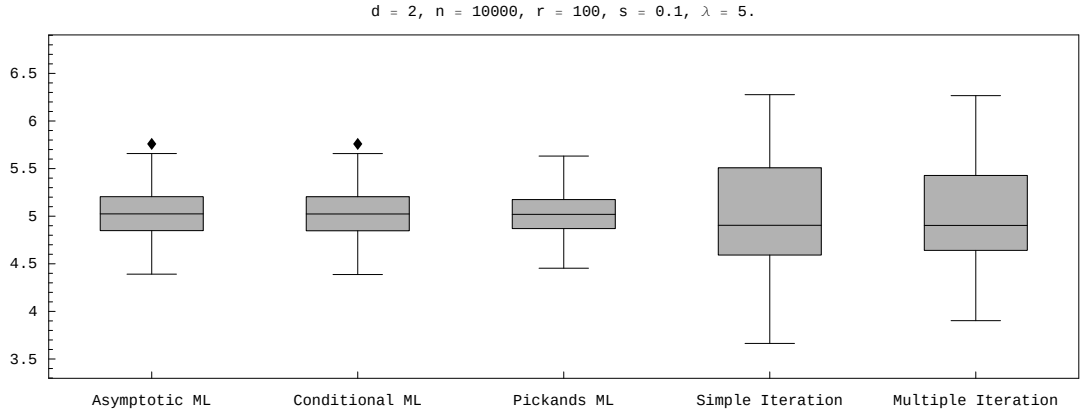


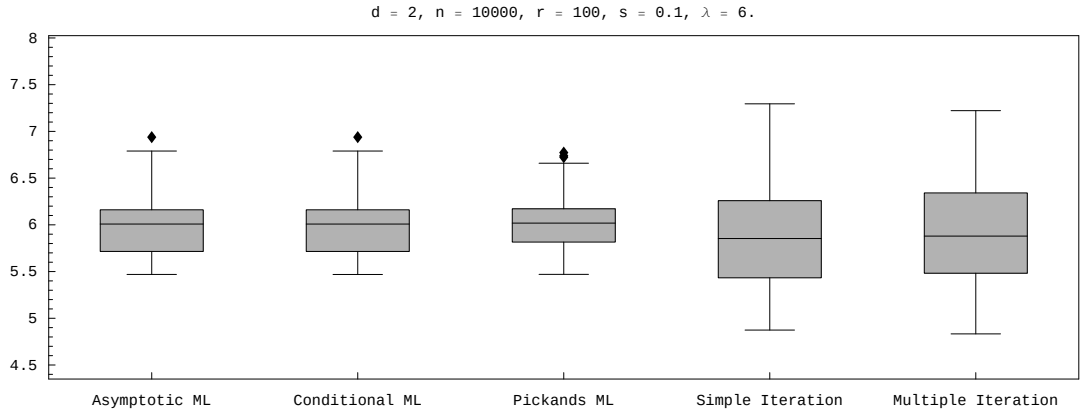

d = 2, n = 10000, r = 100, s = 0.1, λ = 1.8
2.4
2.2
2
1.8
1.6
1.4
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration


d = 2, n = 10000, r = 100, s = 0.1, λ = 2.
2.6
2.4
2.2
2
1.8
1.6
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration


d = 2, n = 10000, r = 100, s = 0.1, λ = 3.
4
3.5
3
2.5
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration

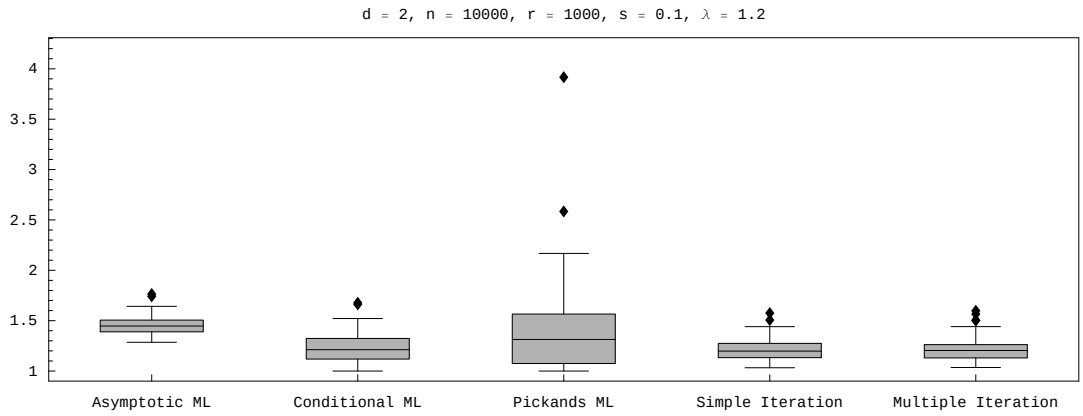


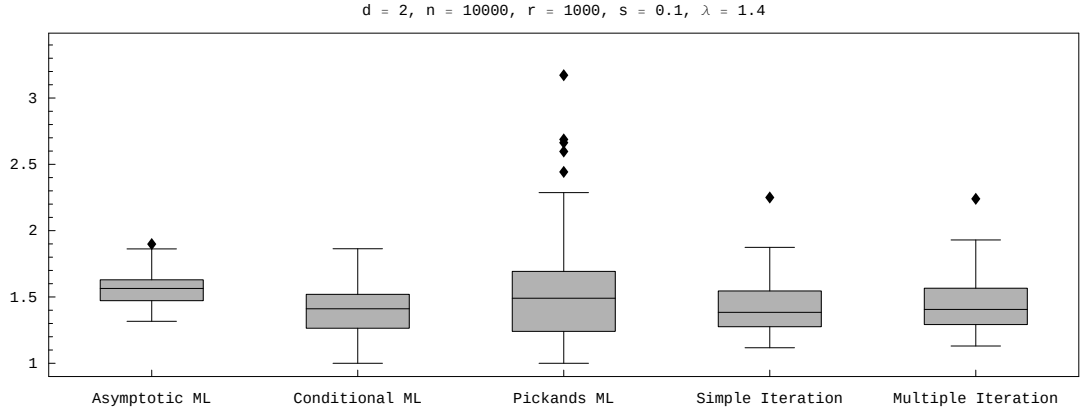


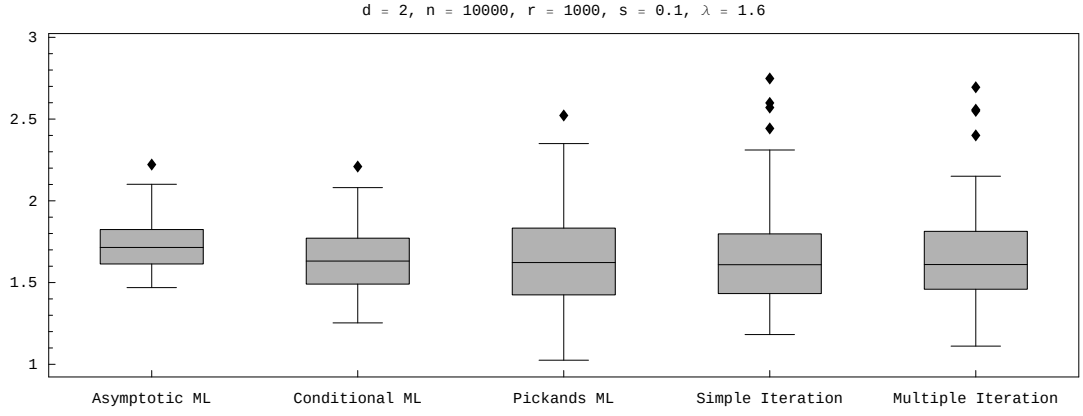


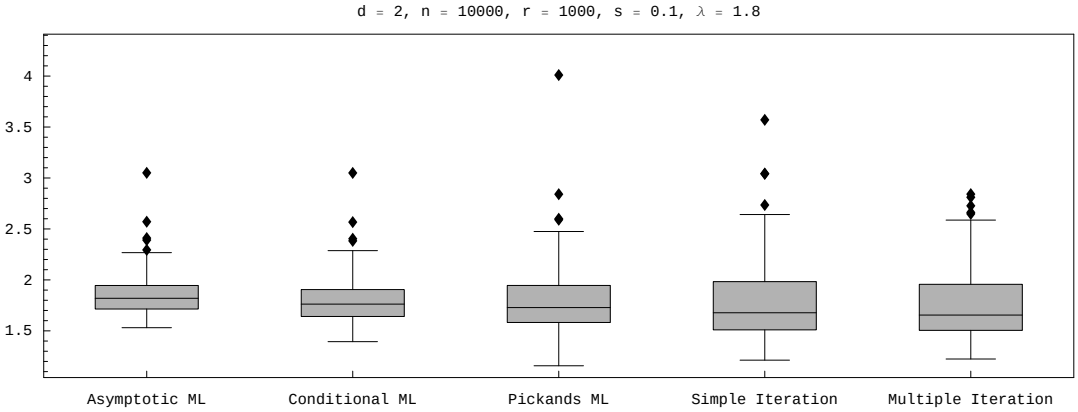
Here one should make a correction with the Simple Iteration when the Asymptotic ML estimation is smaller than 2.5. We observe that a small r reduces the variances of the estimators via relative frequencies, so they are superior to the ML methods in more cases than before. r must, however, be at least d/s, what would be 20 in this case. So one should choose r not too small securing that one has still enough freedom for the choice of v. In all, the behavior of the methods is as before.

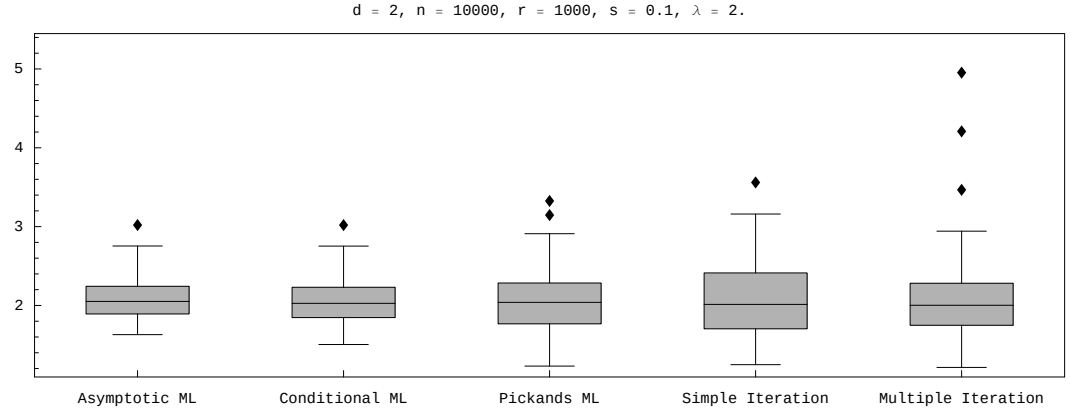
Finally we put $r = 1000$.

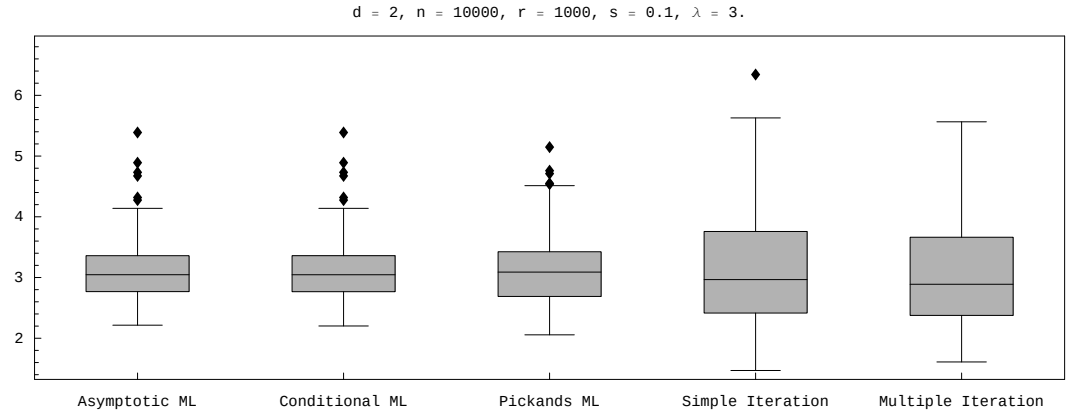


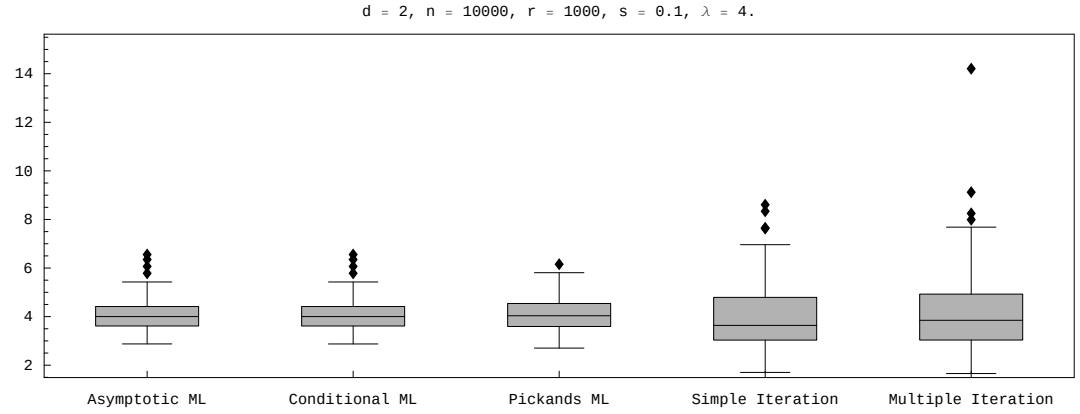


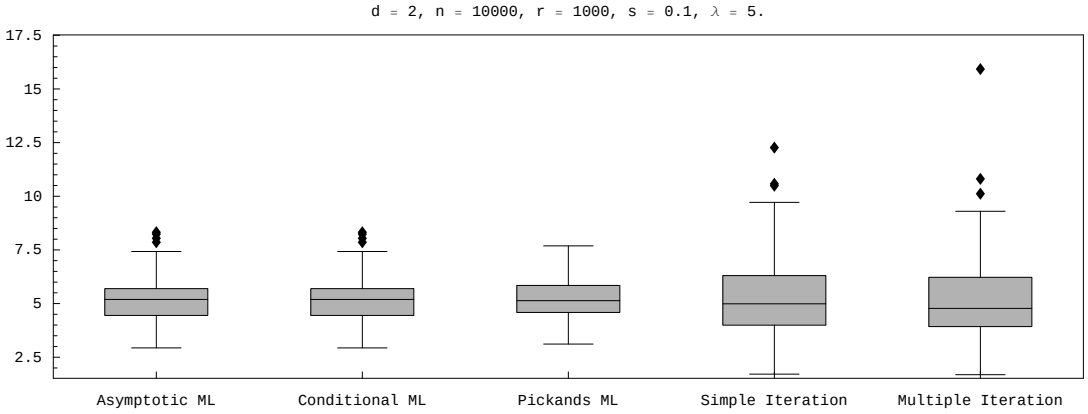


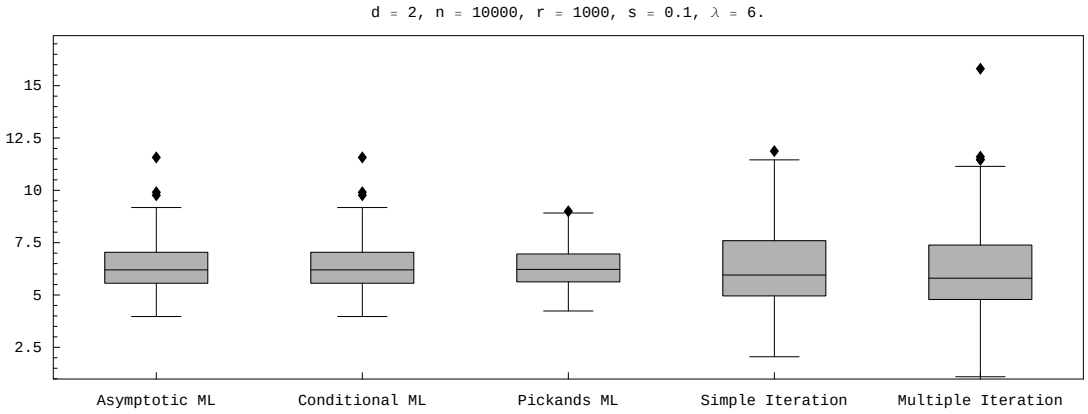

d = 2, n = 10000, r = 1000, s = 0.1, λ = 1.8
4
3.5
3
2.5
2
1.5
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration


d = 2, n = 10000, r = 1000, s = 0.1, λ = 2.
5
4
3
2
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration


d = 2, n = 10000, r = 1000, s = 0.1, λ = 3.
6
5
4
3
2
Asymptotic ML
Conditional ML
Pickands ML
Simple Iteration
Multiple Iteration

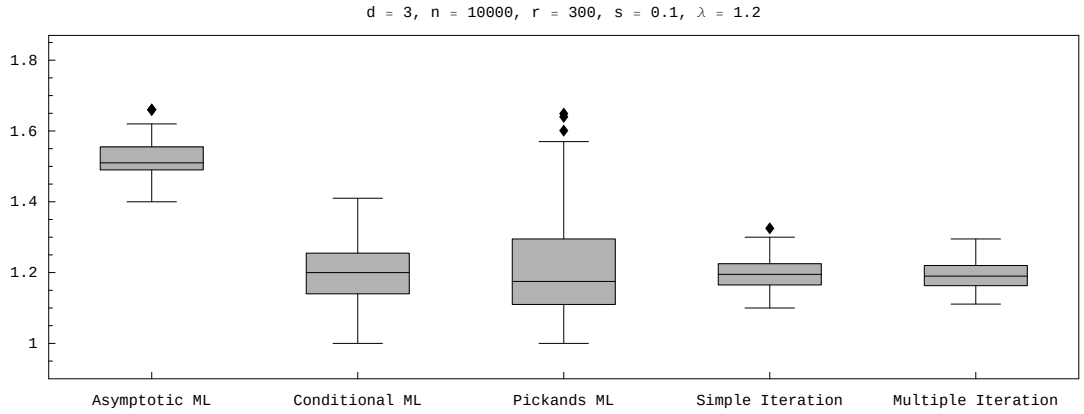


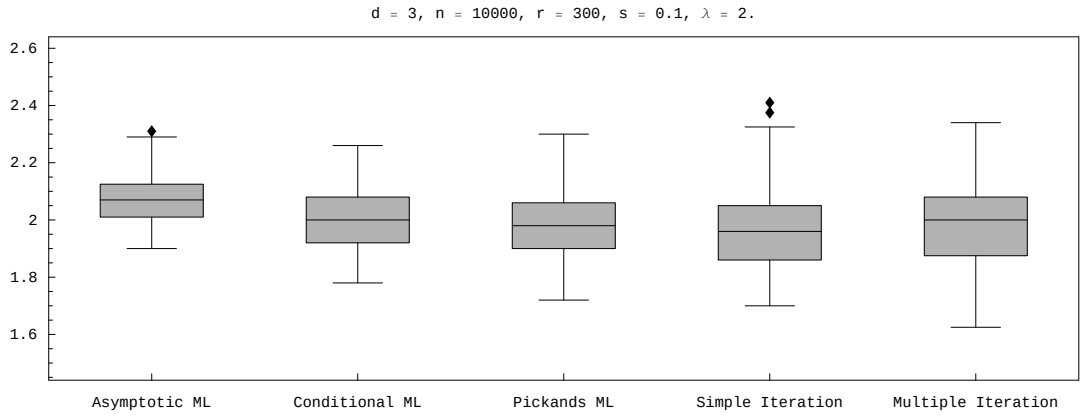


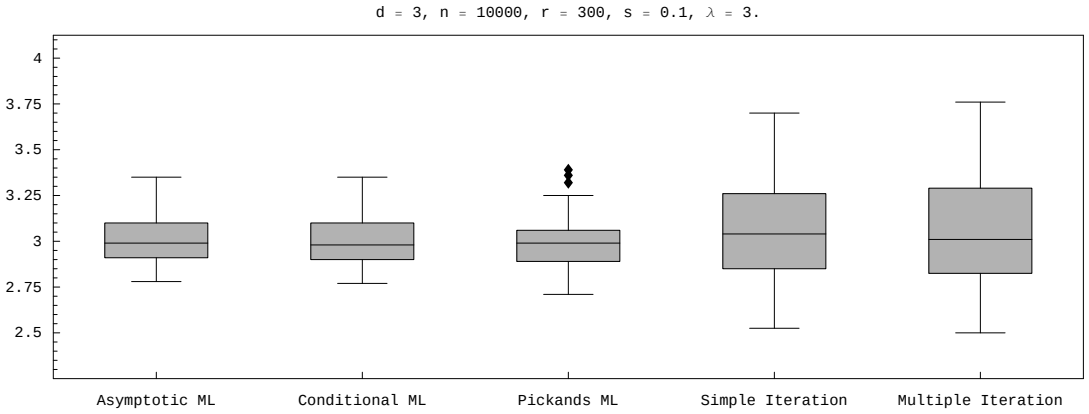


The acceptance level for the Asymptotic ML method could here be at 1.8. A value between 2 and 2.5 seems to be a good value to decide the question if one should do an additional iteration with the estimator via relative frequencies or not, when one has used the efficient Asymptotic ML method. The new value should only be accepted if it is smaller than the old one to prevent the possibly higher variance of the simple iteration from entering into the estimation on a larger scale.

The Conditional ML and the Pickands ML method seem to be working quite well in all cases. We also examine the behavior for $d = 3$, however, we restrict ourselves to a few graphics here.







The behavior is identical to the bivariate cases shown above. For small λ, i.e., close to independence, the iteration procedures are superior, for $\lambda = 2$ all procedures are roughly equivalent,

for $\lambda = 3$, the boundary case between dependence and independence, see Lemma 2.3.11, the ML methods are preferable. A value a little larger than 2 also seems to be a good value here, to decide the question whether another iteration is necessary after executing the Asymptotic ML method.

In all, also taking into consideration the asymptotic variances from Example 6.2.2, the Conditional ML method seems to be the best estimator, but since it takes a very long time to compute an estimation with it, it might not be the best suited estimation procedure for practical purposes, especially in dimensions larger than 2. In these cases one should turn to one of the other estimation procedures. Asymptotic ML is the only method, which can efficiently be used in high dimensions, but is possibly biased for finite sample sizes.

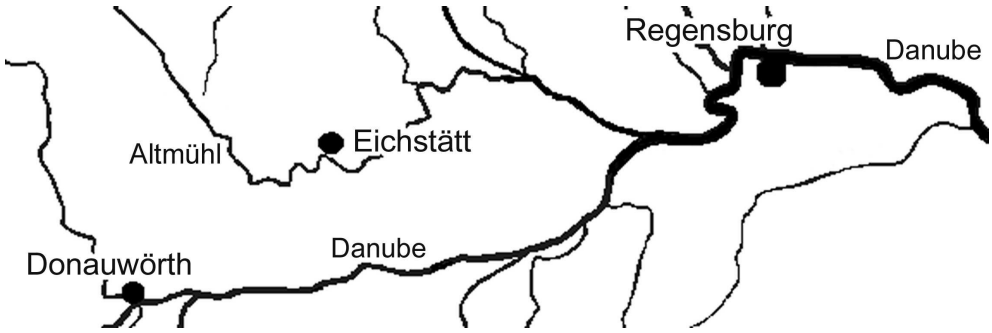
$\Diamond$

Chapter 7

Application to a Hydrological Data Set

In this chapter we want to apply the methods presented in the previous chapters to a real hydrological data set. In recent years southern Bavaria has been hit by floods with damages amounting to several million euros. Naturally an investigation in the dynamics behind these floods is essential for the understanding of their impacts and for possible flood protection measures taken by the authorities. We want to give an example here, how extreme value theory can help in these analyses.

The data we investigate consist of water discharges, which are measured in cubic meters per second (m^3/sec). We look at the measurements from three different locations in southern Bavaria. The first site is Eichstätt situated on the Altmühl, the other two are Donauwörth and Regensburg, which are lying on the Danube with Regensburg being downstream, see the following sketch.



The data were provided by the Bavarian State Office for the Environment (Bayerisches Landesamt für Umwelt). For the sites in Donauwörth and Regensburg the data were available from the 1st of November 1923 to the 31st of December 2004, for the site in Eichstätt the data were available from the 1st of November 1929 to the 31st of December 2004. For the seasonal adjustment all of the data were used, for the multivariate analysis the data from 1923 to 1929 for the locations at the Danube were dropped to have balanced data.

We will investigate the problem, whether extreme water levels tend to appear together at these sites. Naturally we would expect a high degree of dependence between the sites in Donauwörth and Regensburg, since they are lying along the same river. But is there, for example, an extremal

dependence between Eichstätt and Donauwörth?

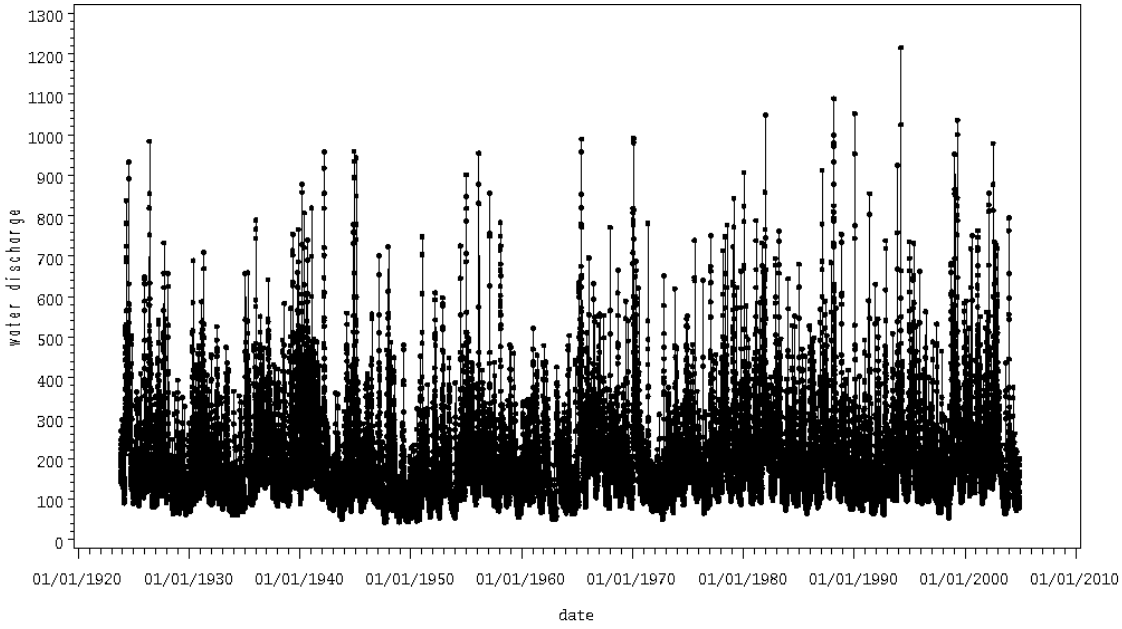
As suggested in Section 8.2.2 of Coles (2001, [8]) we begin in Section 7.1 with a univariate analysis of the data at each site. After transforming the data to suitable margins a bivariate and trivariate analysis, i.e., a modelling of the extremal dependence structure with generalized Pareto distributions is done in Sections 7.2 and 7.3. We compute nonparametric estimators of the angular density and give the degrees of dependence by means of the logistic models from Section 2.3 and the estimation procedures from Chapter 6.

7.1 Univariate Analysis

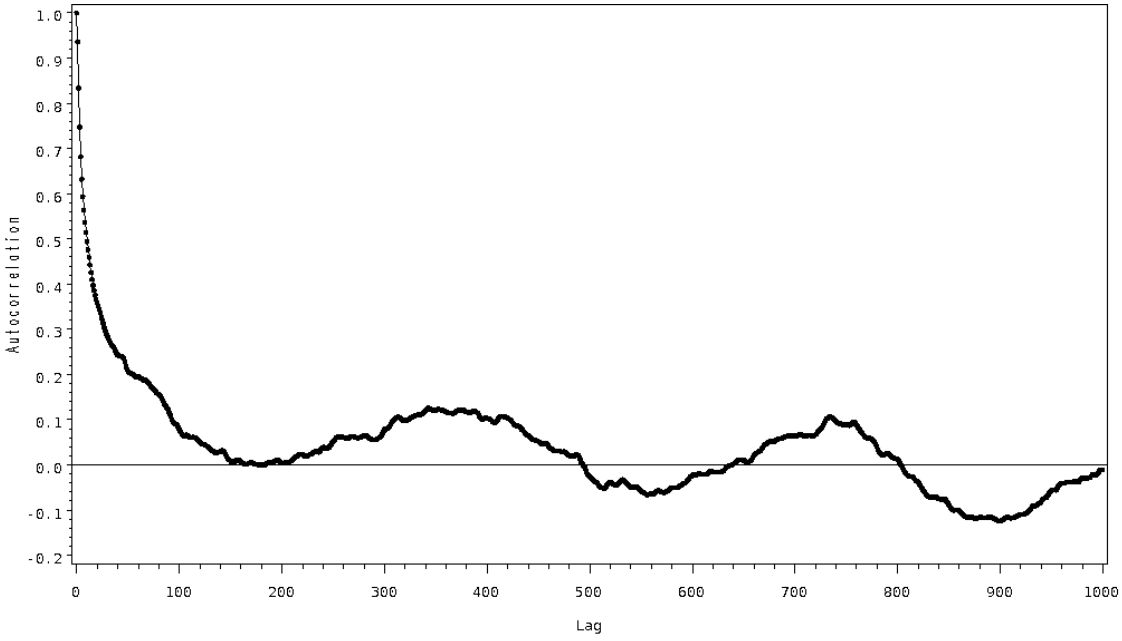
Before we can begin the multivariate analysis we have to do an univariate analysis for each of the three locations. In this section we restrict ourselves to the data collected at Donauwörth, the other two were treated in a similar manner. In the course of our investigation we will frequently use terms from time series analysis. For an introduction to time series analysis we refer to Falk et al. (2006, [22]), especially to Chapters 1 and 3, where seasonal adjustment, the autocorrelation function and the periodogram are explained. A summary of the tools used here can also be found in Section 2.5 of Reiss and Thomas (2001, [57]). The plots in this section were done with SAS, version 9.1.3 (using the programs from Falk et al. (2002, [23]) and Falk et al. (2006, [22]) as templates), the seasonal adjustment was done with STATISTICA, version 7.0, and the rest of the calculations with MATHEMATICA, version 5.2.

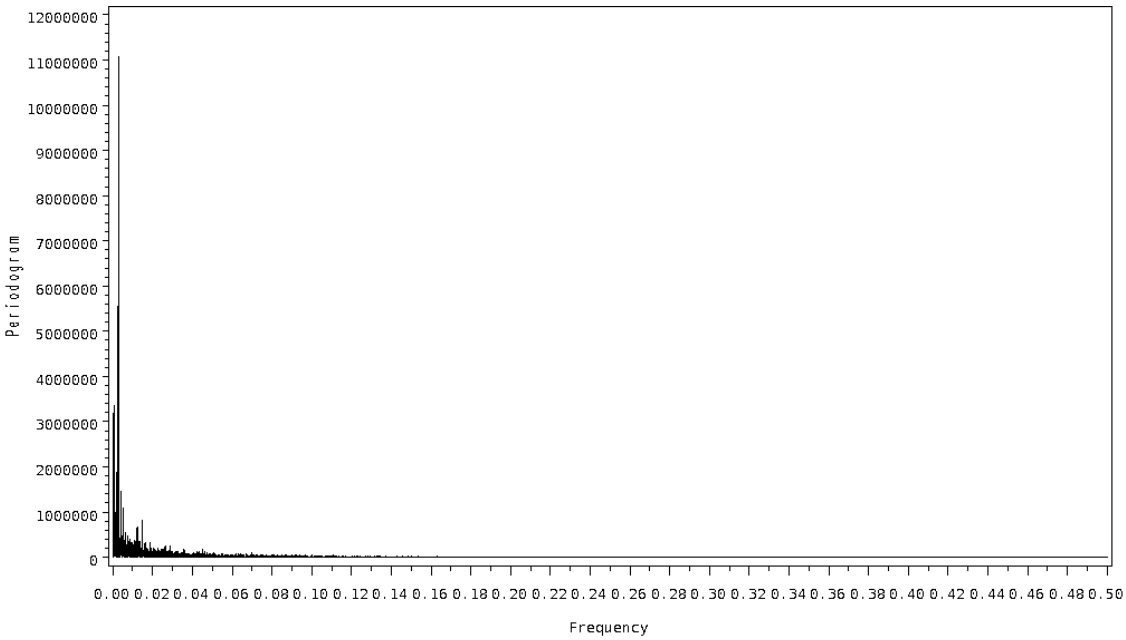
Every 15 minutes the water discharge is measured at the site in Donauwörth, and for each day the arithmetic mean of the respective observations is computed. This daily arithmetic mean is the target measurement of our investigation. The only exception is in leap years where for reasons of seasonal adjustment (to get a fixed cycle of 365 days) the observation for the 28th of February contains the arithmetic mean of the data from the 28th and 29th of February. There is, thus, no observation for the 29th of February.

We begin by taking a simple look at the data of daily means plotted over the time axis.



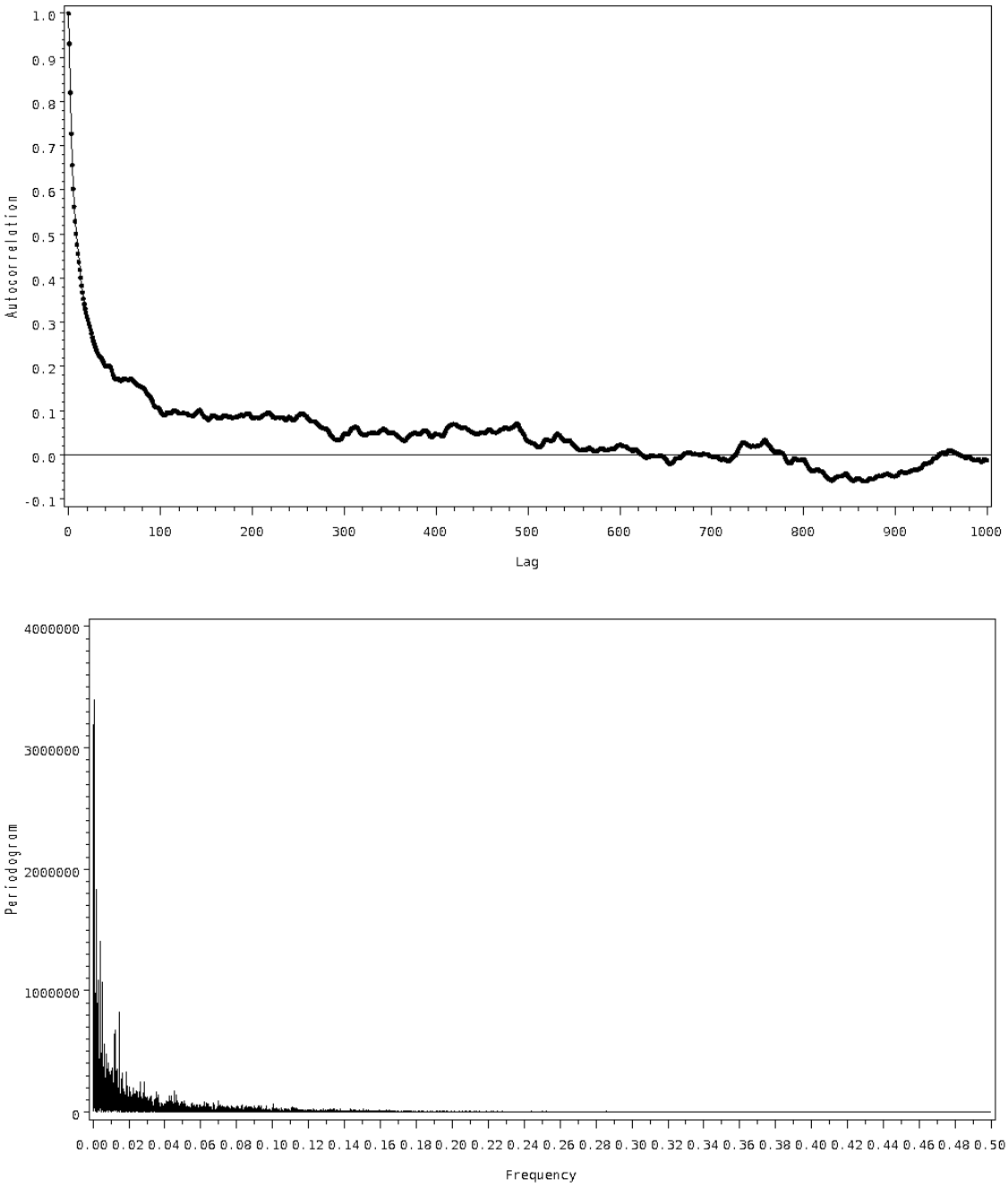
There seems to be no visible trend, but there might be a seasonal component. Tools to investigate temporal dependence are the autocorrelation function and the periodogram. So next we plot the autocorrelation function and the periodogram of the Donauwörth data.





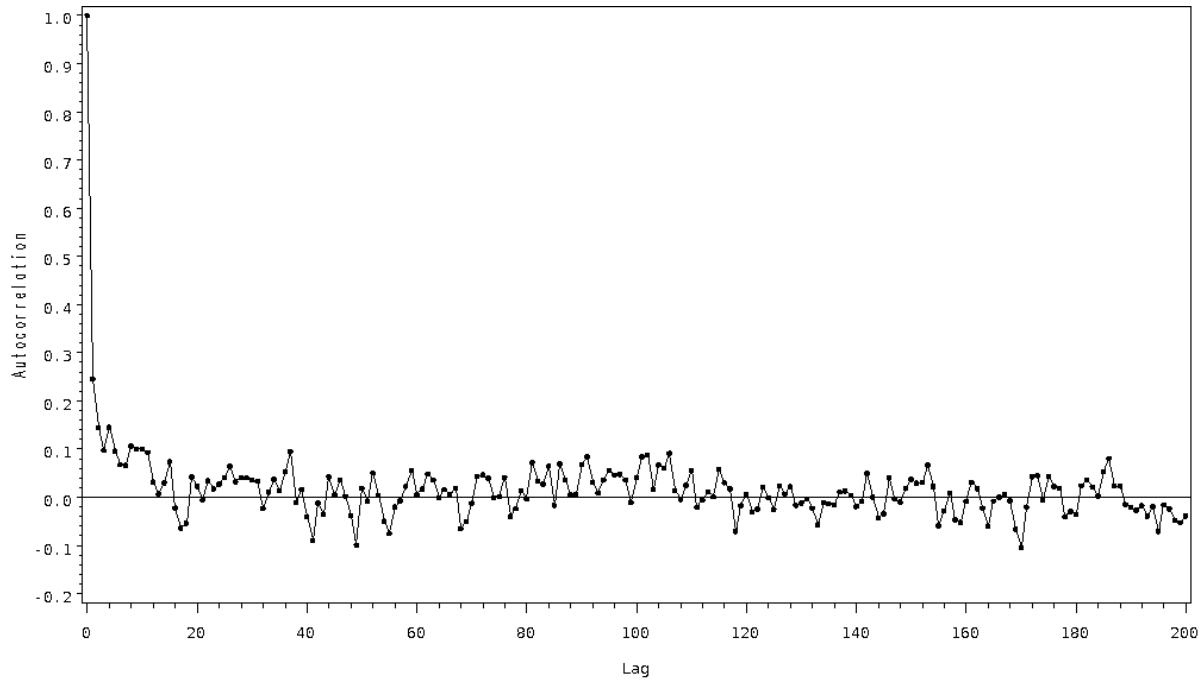
Both graphics point to a high seasonality in the data with a period of 365 observations, since the periodogram has one clear peak, which is at $0.002734 \approx \frac{1}{365}$. The autocorrelation function is also varying very regularly with this period. So a seasonal adjustment with a period of 365 was

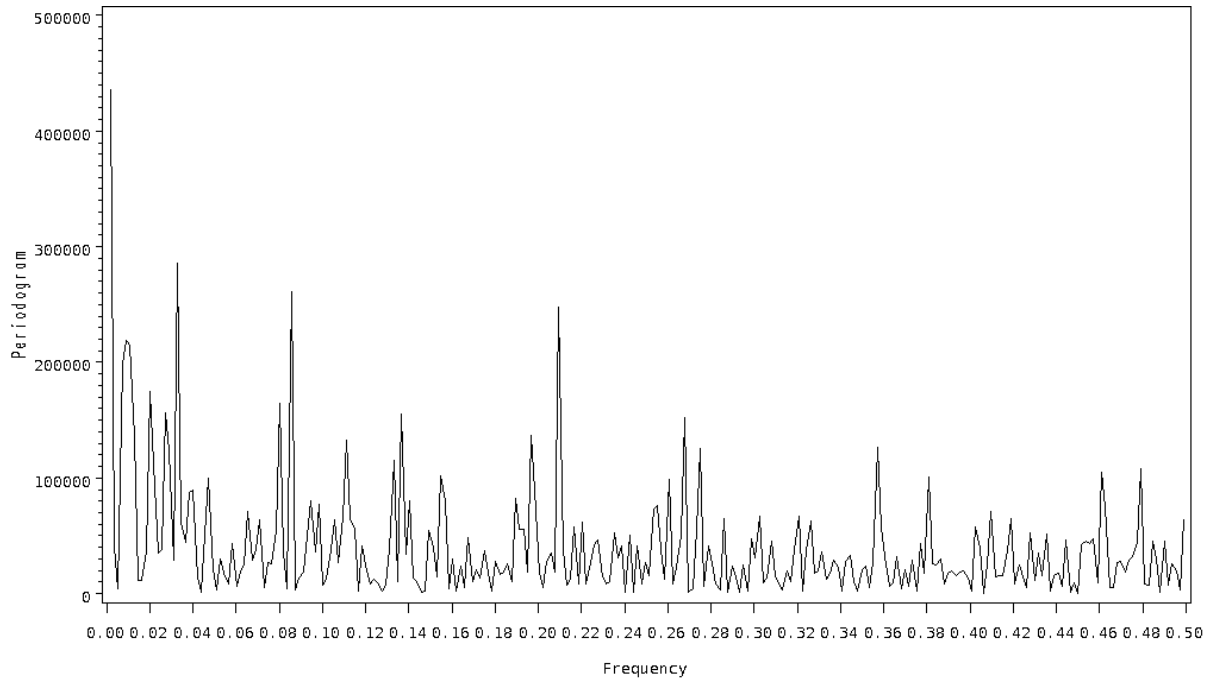
done, and the autocorrelation function and the periodogram of the adjusted data are plotted.



The high seasonality is now taken out. However, there is, still, a clear temporal dependence in the data, i.e., observations at a given date have a high correlation with the previous observations. Since we are only interested in the extremes of the data, and since we want to the model the data as coming from independent random variables, we now take blockwise maxima of order 50. I.e., the time axis is divided into intervals of 50 days, and for each interval only the maximum is considered further. This procedure is suggested by the fact that observations with a lag of only a few days show a high correlation. It is also the standard method in extreme value analysis of time series, see Sections 9.3 and 10.2 of Beirlant et al. (2005, [3]). The traditional approach would be

to take annual block maxima, however, to get more data it is also an advisable procedure to take maxima of smaller block sizes, see Section 11.1 of Reiss and Thomas (2001, [57]). The necessity to move away from daily observations is also given, since it takes the water up to several days to go from Donauwörth or Eichstätt to Regensburg and, thus, a multivariate analysis based on the daily observations could lead to wrong interpretations of the results. Altogether our data set is, thus, reduced to 549 observations, consisting of maxima of blocks with a length of 50 days from the daily observations. We plot again the autocorrelation function and the periodogram of the block maxima data.





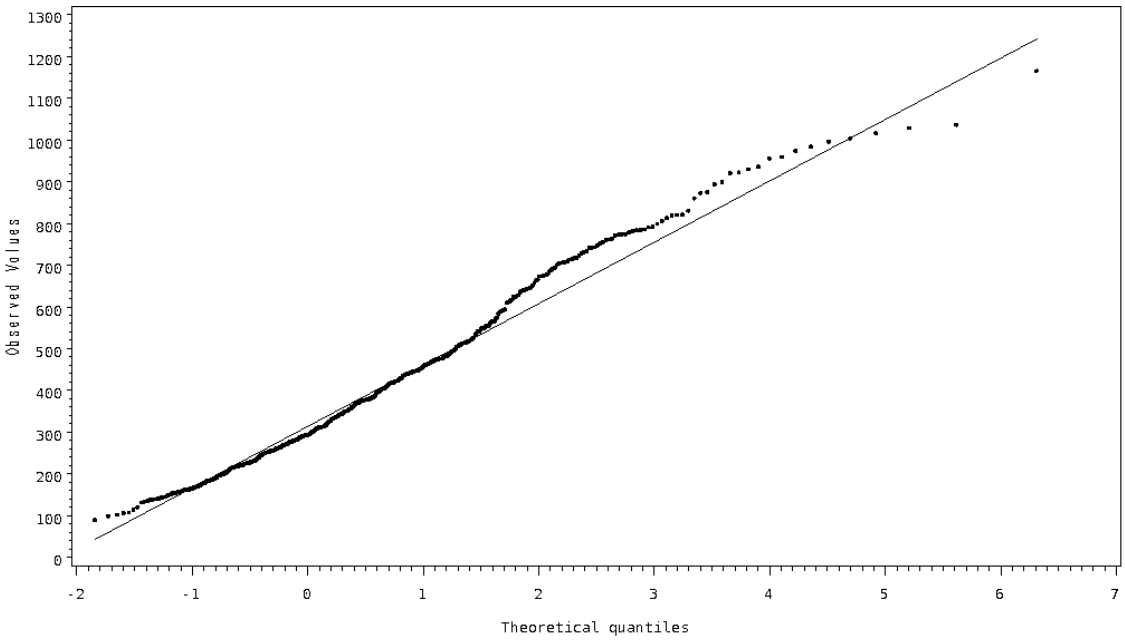
There might be still some low level temporal dependence in the data, but for simplicity reasons we model the blockwise maxima data as coming from independent and identically distributed random variables.

A possible alternative to taking block maxima of fixed size is using clusters of variable lengths by defining appropriate thresholds, see Section 5.3 in Coles (2001, [8]).

The question we consider next is, which distribution function could possibly underly our block maxima data. Remember that our original data were daily means, and as such it is reasonable to assume that they are approximately normally distributed. These data were seasonally adjusted, and blockwise maxima were taken afterwards. Since the normal distribution lies in the maximum domain of attraction of the Gumbel distribution (see for example Section 2.2 in Reiss and Thomas (2001, [57]) or Section 2.3.2 in Galambos (1978, [29])), it seems reasonable to assume that the data follow a Gumbel distribution. I.e., they have the distribution function

$$F_{\mu,\sigma}(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$$

with a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$. This convergence of normal maxima to the Gumbel distribution remains true, even if the underlying random variables are dependent, see Sections 5.2 and 5.3 in Coles (2001, [8]). We check this distribution assumption with the following quantile plot. We refer to Section 1.6 of Falk et al. (2002, [23]) for more information on quantile plots.



Since this plot is close to a line the assumption of the Gumbel distribution is a reasonable one. This is in accordance with Section 2.2 of Beirlant et al. (2005, [3]), as well as Sections 4.1 and 11 of Reiss and Thomas (2001, [57]), where it is noted that the Gumbel distribution is commonly used in hydrology for the modelling of river discharge maxima.

Next we have to estimate the parameters μ and σ, which we do by the standard maximum-

likelihood method, i.e., we maximize the loglikelihood function

$$\begin{aligned}\Lambda(\mu,\sigma) &= \log\left(\prod_{i=1}^{n}\frac{1}{\sigma}\exp\left(-\frac{x_i-\mu}{\sigma}\right)\exp\left(-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right)\right)\\ &= -n\log\sigma-\frac{n}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right)-\sum_{i=1}^{n}\exp\left(-\frac{x_i-\mu}{\sigma}\right)\end{aligned}$$

with regard to μ and σ, where $x_1,\ldots,x_n$ are our data. This is an approach of estimation suggested in Section 5.1 of Beirlant et al. (2005, [3]) and Section 3.3 of Coles (2001, [8]). For the case of the Donauwörth data this leads to the estimators

$$\hat{\mu}_1 = 312.207 \qquad \text{and} \qquad \hat{\sigma}_1 = 147.07.$$

These were again computed by the function `NMaximize` of MATHEMATICA. Since we will need them later we also give the corresponding values for the Regensburg and Eichstätt data:

$$\hat{\mu}_2 = 664.872 \qquad \text{and} \qquad \hat{\sigma}_2 = 294.72,$$

$$\hat{\mu}_3 = 14.6549 \qquad \text{and} \qquad \hat{\sigma}_3 = 13.8436.$$

The function $F_{\hat{\mu}_1,\hat{\sigma}_1}$ is then the estimated distribution function of the Donauwörth data. Therefore, we can transform the data onto $(-1,0)$ by computing

$$y_i := F_{\hat{\mu}_1,\hat{\sigma}_1}(x_i) - 1, \qquad i = 1,\ldots,n. \tag{7.1}$$

With this transformation we can model the y_i to come from a uniform distribution on $(-1,0)$, see Corollary 1.6.4 in Falk et al. (2002, [23]). This is, clearly, only an approximation. The ordering of the data remains untouched by the transformation, so that the extreme observations will now be close to 0.

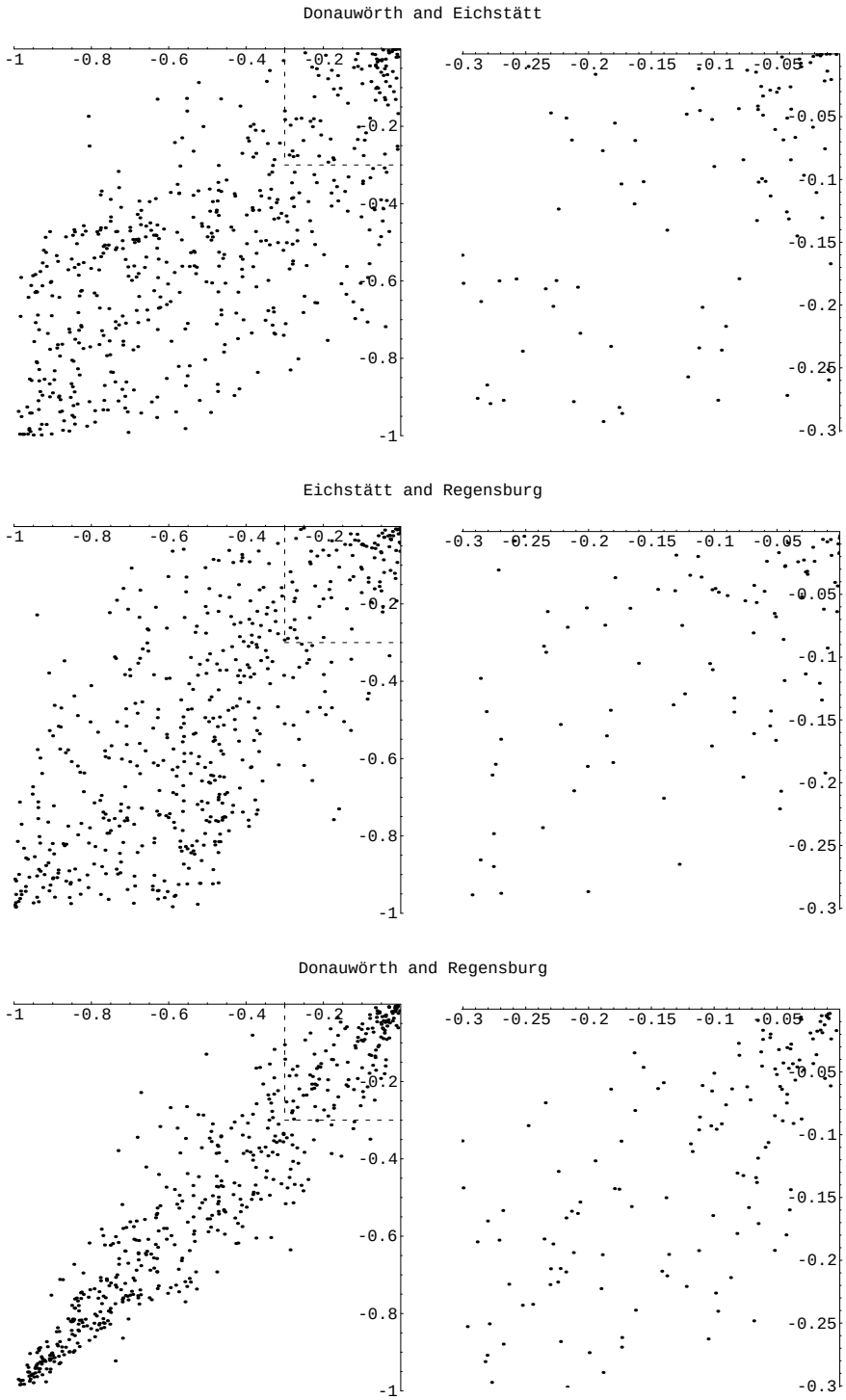
This method of first estimating the margins and then estimating the dependence parameters of the suitably transformed data is often referred to as piecing-together estimates (PTEs), see for example Section 9.3 of Reiss and Thomas (2001, [57]).

By estimating the margins in the way we did we followed the classical approach of extreme value theory. This was done to be able to give a closed formula at the end of Section 7.3 for the probability of events, which have not been recorded yet. Instead of using the parametric assumption of the Gumbel distribution we could also use recently introduced methods to transform the data with the empirical distribution function to $(-1,0)$ as it is proposed in Capéraà et al. (1997, [6]), Coles et al. (1999, [9]), Einmahl et al. (2001, [18]) or Section 9.3 of Beirlant et al. (2005, [3]). This corresponds to the rank transformation. Another possibility would be to use a smoothed version of the empirical distribution function, see Section 2.1 of Reiss and Thomas (2001, [57]).

7.2 Bivariate Analysis

After having computed the univariate analysis in the previous section, we can now come to the bivariate analysis. By denoting the resulting transformed data with do_i for Donauwörth, ei_i for Eichstätt and re_i for Regensburg, $i = 1,\ldots,549$ we look at each combination separately. First we

plot the bivariate data (do_i, ei_i), (ei_i, re_i) and (do_i, re_i) and the extreme area $K_{0.3} = (-0.3, 0)^2$, in which we will be especially interested in the sequel, to get a first visual impression.



One immediately sees that for the last data set a clear dependence between the components is present. For the other two there also seems to be some weak dependence. But what about the extreme area? Is there some dependence present, and if so, are there differences in the dependencies between the data sets?

The left picture for each data set can also be interpreted as the estimated copula of the data. In fact, there is a close connection between GPDs and copulas. According to Section 5.1 of Falk et al. (2004, [21]) GPDs are, after a certain shift, quasi-copulas.

The observations falling into $K_{0.3}$, which can be seen in the right graphics above are now modelled as coming from independent and identically distributed random variables following a bivariate GPD.

This is justified, since we look at these data under the condition that they exceed some high threshold (each transformed component is larger than -0.3 in the extreme area). We know from (2.13) or Theorems 2.2 and 2.3 in Rootzén and Tajvidi (2005, [60]) that such exceedances follow asymptotically a GPD close to 0. Since we transformed the data to uniform margins in the univariate analysis in Section 7.1, a bivariate GPD with uniform margins is the natural choice here.

There is also another way of justifying this approach of modelling the extremes. The componentwise maxima taken in Section 7.1 are reasonably modelled by a bivariate EVD when the corresponding bivariate observations are put together. To standardize the margins, the univariate data should be transformed to a negative exponential distribution, i.e.,

$$\tilde{y}_i := \log F_{\hat{\mu},\hat{\sigma}}(x_i), \qquad i = 1, \ldots, 549, \tag{7.2}$$

instead of $y_i = F_{\hat{\mu},\hat{\sigma}}(x_i) - 1$ as in (7.1). But from the known relation $\log(1+\varepsilon) \approx \varepsilon$ for ε close to 0, which can easily be seen by a Taylor expansion, we get for extreme observations, i.e., for $F_{\hat{\mu},\hat{\sigma}}$ close to 1, the approximation

$$\tilde{y}_i = \log F_{\hat{\mu},\hat{\sigma}}(x_i) = \log\left(1 + F_{\hat{\mu},\hat{\sigma}}(x_i) - 1\right) \approx F_{\hat{\mu},\hat{\sigma}}(x_i) - 1 = y_i.$$

The bivariate data, transformed to negative exponential margins, should reasonably be modelled by an EVD with negative exponential margins. But such an EVD can close to the origin be approximated by a GPD with uniform margins, since an EVD is in the spectral 1-neighborhood of a GPD, see Section 5.3 of Falk et al. (2004, [21]) and Section 5.3 in this manuscript.

Thus, with both marginal transformations, which do not differ much for the extreme observations we are interested in, a GPD approximation of the underlying distribution of the extreme data is reasonable.

The first justification has the advantage that it can also be applied when the data originally do not follow asymptotically an EVD, which will be the case in most applications.

Before coming to the estimations under the GPD model assumption, we have to make sure that we are not in the case of independence

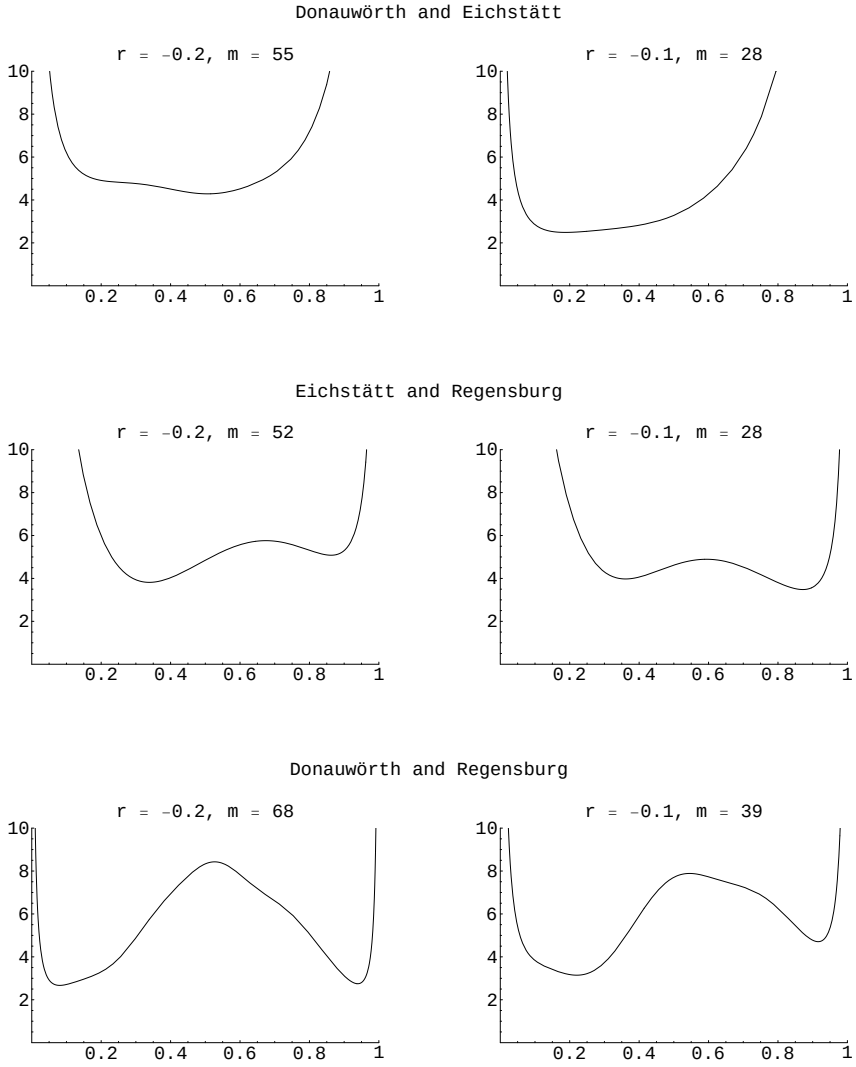
$$W(x_1, x_2) = 1 + x_1 + x_2,$$

where we should asymptotically see no observations above our thresholds. Therefore we have to check, whether the distribution underlying our observations, which is to be approximated by a GPD, has tail independent margins. Our observations containing the 549 maxima originally follow asymptotically an EVD, and we transform the margins as in (7.2) to negative exponential margins, which does vary much from the transformation to uniform margins in the extreme area, see above. In Section 6.5 of Falk et al. (2004, [21]) and in Falk and Michel (2006, [24]) a test for tail independence in EVD models is introduced, which tests for $D(z) = 1$. We use the Neyman-Pearson version of this test to check the null hypothesis of tail independence for different thresholds c. The resulting p-values are given in the following table.

Data set	$c = -0.3$	$c = -0.2$	$c = -0.1$
Donauwörth & Eichstätt	$1.01 \cdot 10^{-28}$	$9.60 \cdot 10^{-14}$	$5.28 \cdot 10^{-7}$
Eichstätt & Regensburg	$5.37 \cdot 10^{-28}$	$5.86 \cdot 10^{-22}$	$6.45 \cdot 10^{-12}$
Donauwörth & Regensburg	$2.09 \cdot 10^{-29}$	$3.12 \cdot 10^{-16}$	$9.34 \cdot 10^{-4}$

We see that the null hypothesis of tail independence is rejected at the 5% level, even after a Bonferroni correction, in all cases. Thus, the GPD approximation does not lead to the case of independence. We refer to the above references for details of the testing procedure.

By the GPD model assumption the methods presented in the previous chapters are applicable, and we begin by estimating the angular density of the data with the estimator (4.3) from Chapter 4. Thereby we chose -0.2 and -0.1 as thresholds r for the radial component of the standard Pickands coordinates. The corresponding number of observations which fall over the threshold is given in the graphics as m.



First we notice that the results seem to be independent of the choice of the threshold, when one takes into account that the kernel density estimators are based on less data in the right graphics, which naturally increases the variance. This supports the model assumption of generalized

Pareto distributions since by the results of Chapter 4 the distribution of the angular components is independent of the threshold, see especially Remark 4.1.7.

Between the extremes of Donauwörth and Eichstätt there seems to be a high degree of independence present, whereas a little more dependence seems to be between Eichstätt and Regensburg. But the clearest dependence is suggested by the graphics to be between Donauwörth and Regensburg. This is not surprising since the Danube contains much more water than the Altmühl. Thus we could argue that an extreme water discharge at Donauwörth would effect Regensburg much harder than an extreme water discharge in Eichstätt which is more easily swallowed by the Danube.

Next we assume the data to come from a logistic model, and we want to estimate the parameters, thereby giving quantitative measures of dependence. This model assumption is not contradicted by the nonparametric estimators, although it cannot be excluded that some slight asymmetry may be present. We chose $s = 0.3$, $r_F = 80$ and $r_P = -0.1$, and with these parameters we computed the five estimators as in Example 6.4.3. Between 20 and 40 of the observations transformed in Section 7.1 fell over the respective thresholds each time. The results are given in the following table for each data set.

Data set	Asym. ML	Cond. ML	Pick. ML	Simple It.	Multiple It.
Donauwörth & Eichstätt	1.43	1.01	1.0	1.45	1.45
Eichstätt & Regensburg	1.45	1.08	1.0	1.57	1.49
Donauwörth & Regensburg	2.28	2.23	2.16	3.30	2.96

In the next table we give the estimated standard errors of the above estimators, using the normal approximations of Theorems 6.1.2, 6.1.4, 6.2.1 and 6.3.4. The underlying parameter was thereby replaced by its estimation. For the iteration methods the parameter v from the final iteration was used to compute the standard error by Theorem 6.3.4, so that these estimated standard errors have to be viewed with some care.

Data set	Asym. ML	Cond. ML	Pick. ML	Simple It.	Multiple It.
Donauwörth & Eichstätt	0.09	0.003	-	0.24	0.24
Eichstätt & Regensburg	0.10	0.02	-	0.25	0.26
Donauwörth & Regensburg	0.28	0.27	0.29	0.71	1.09

In two cases of the Pickands ML method the standard error could not be given due to the estimation of 1.0. Also the standard errors of the Conditional ML estimators have to interpreted carefully due to the estimation close to 1.

Note that in all these and the following estimations the uncertainties arising from having to estimate the margins are not taken into account. This is, however, not uncommon, see Section 9.3 of Beirlant et al. (2005, [3]).

For the two data sets including Eichstätt we get very similar results. Both Conditional and Pickands ML give values very close to 1, whereas Asymptotic ML and the iteration methods estimate around 1.5. We have seen in Examples 6.1.3, 6.2.2 and 6.4.3 that in cases of low dependence Conditional ML and Pickands ML tend to estimate very close to 1 in many cases, although the underlying parameter is higher. We have also seen in the test for tail independence above that the case of independence is not an appropriate model for these data. Since we have also noted in Example 6.4.3 that the Asymptotic ML method is possibly biased close to the

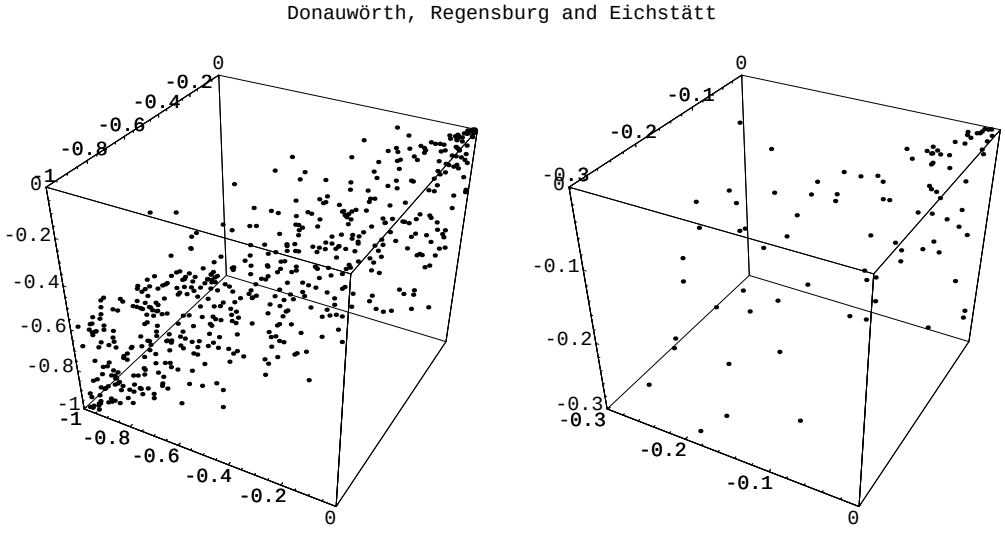
independence and we are not able to give reasonable standard errors for two of the ML methods, the iteration methods should be preferred here. They are also more reliable by Example 6.4.3 and, thus, be the ones we stick to. So in both data sets the parameter is estimated close to 1.5 with a standard error of about 0.25, thereby the values for the second one being a little higher.

The estimations for the Donauwörth & Regensburg data set are different, since the three ML methods seem to agree here, whereas the iteration methods give higher values and have higher standard errors either. This is a behavior known from Example 6.4.3. Since we are closer to the dependence case the ML methods should be preferred, giving estimations around 2.2 with a standard error of around 0.3. So there is stronger dependence here, but not as strongly as one might have guessed aforehand, since it is only little over the boundary case of $\lambda = 2$ between complete dependence and independence in the logistic model, see Lemma 2.3.11.

The result could be interpreted in the following way: Even for locations lying alongside the same river an extreme flood at one location does not necessarily entail an extreme flood at the other location. Either the extreme flood forms between the two sites or it may weaken between them so that one location experiences an extreme flood while the other may, still, have a flood but not at extreme levels. To determine the underlying reasons for that like the geology, the rainfall in the area or others, to formulate corresponding interpretations and suggest resulting consequences for flood protection measures is now the task of a hydrologist.

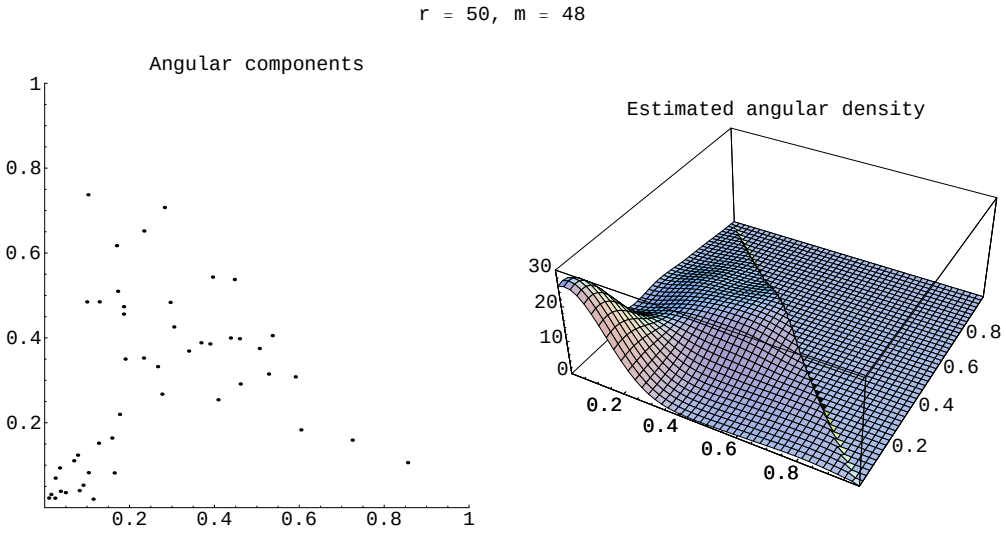
7.3 Trivariate Analysis

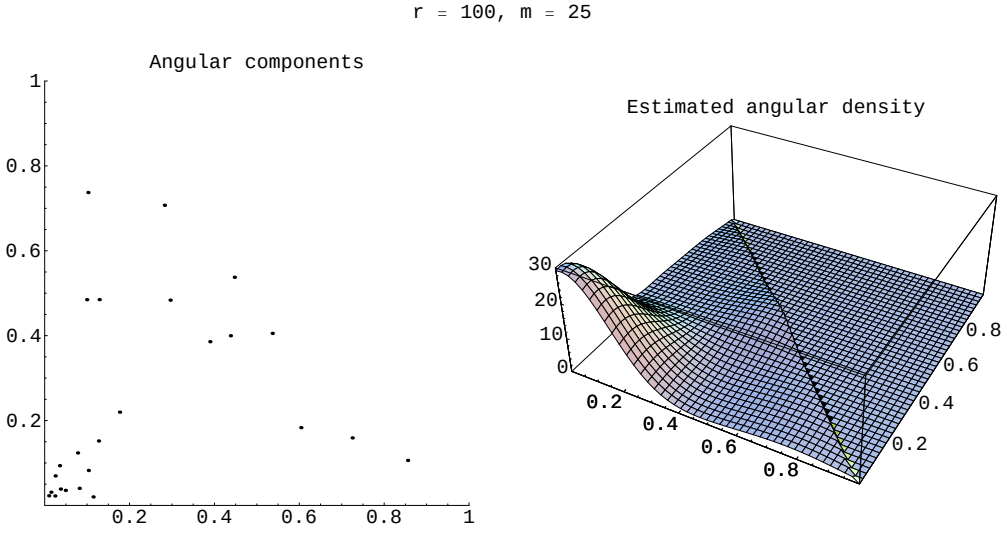
In this section we finally look at the full trivariate data set. The goal is here to find a suitable trivariate model for the data. We begin, as before, by taking a look at the data (do_i, re_i, ei_i) in full and in the extreme area.



Again there seems to be some dependence present, but to specify it from just looking at the data is very hard. As in Section 7.2 we model the extreme data in the right graphic to come from independent and identically distributed random variables following a trivariate GPD with uniform margins. Since we are not anymore in the bivariate case we have to use the methods from Chapter 5, especially estimator (5.7) for nonparametric estimation of the angular density.

We choose s to be 0.3 and plot the angular components with regard to Fréchet margins and the estimated angular density for two different thresholds. m denotes again the number of observations falling over the respective thresholds.





The angular density seems to be unbounded close to the origin, with mass along the bisector and close to the line $x_1 + x_2 = 1$. Such a behavior is typical for the nested logistic model, see the graphics in Example 2.3.15. This was to be expected since we found out in the previous section under the logistic model that between Eichstätt and the other two locations there seems to be the same kind of (weak) dependence, whereas between Donauwörth and Regensburg there was a higher degree of dependence. So the logistic model does not fit the data, but because of the bivariate margins mentioned above the nested logistic model seems reasonable, see Remark 2.3.14. This is confirmed by the nonparametric estimators shown above.

The final task is now to determine the parameters of the nested logistic model. We did this with the estimation procedures as above, where we chose $s = 0.3$, $r_F = 50$, and $r_P = -0.25$. The results are summarized in the following table.

Parameter	Asymptotic ML	Conditional ML	Pickands ML
$\hat{\lambda}_1$	2.40	2.28	2.07
$\hat{\lambda}_2$	1.63	1.14	1.18

The estimated standard errors for the ML methods by Theorems 6.1.2, 6.1.4 and 6.2.1 are given below.

Parameter	Asymptotic ML	Conditional ML	Pickands ML
$\hat{\lambda}_1$	0.21	0.28	0.60
$\hat{\lambda}_2$	0.10	0.04	0.08

The iteration methods could not be used since condition (6.41) did not hold even with the estimated optimal choices of the parameter v, see Example 6.3.5. This behavior could have different reasons. First it could mean that the sample size is not large enough, which is the most plausible explanation. Another possibility is that the nested logistic model might not be fitting the data as good as one could hope. Or finally it might be the case that the iteration procedure does in general not perform well in the nested logistic case. To clarify these questions detailed investigations on the choice of appropriate sample sizes and the performance of the iteration procedures for the nested logistic model would have to be made. Also methods for the checking of parametric model assumptions like the nested logistic model have yet to be investigated.

The estimations of the parameters with the ML methods give similar results as in the bivariate case with the parameter $\hat{\lambda}_1$ being around 2.2 and thereby being larger than the second parameter $\hat{\lambda}_2$ being somewhere around 1.4, when considering the standard errors. This indicates again stronger dependence between Donauwörth and Regensburg than between Eichstätt and the two other locations.

With these estimated values we can now give the estimated distribution function of the maxima of the seasonally adjusted data in its upper tail. It reads

$$\begin{aligned}
\hat{W}(x_1,x_2,x_3) &= 1-\left(\left((1-F_{\hat{\mu}_1,\hat{\sigma}_1}(x_1))^{\hat{\lambda}_1}+(1-F_{\hat{\mu}_2,\hat{\sigma}_2}(x_2))^{\hat{\lambda}_1}\right)^{\hat{\lambda}_2/\hat{\lambda}_1}+(1-F_{\hat{\mu}_3,\hat{\sigma}_3}(x_3))^{\hat{\lambda}_2}\right)^{1/\hat{\lambda}_2} \\
&= 1-\left(\left(\left(1-\exp\left(-\exp\left(-\frac{x_1-\hat{\mu}_1}{\hat{\sigma}_1}\right)\right)\right)^{\hat{\lambda}_1}\right.\right. \\
&\quad \left.\left.+\left(1-\exp\left(-\exp\left(-\frac{x_2-\hat{\mu}_2}{\hat{\sigma}_2}\right)\right)\right)^{\hat{\lambda}_1}\right)^{\hat{\lambda}_2/\hat{\lambda}_1} \\
&\quad +\left(1-\exp\left(-\exp\left(-\frac{x_3-\hat{\mu}_3}{\hat{\sigma}_3}\right)\right)\right)^{\hat{\lambda}_2}\Bigg)^{1/\hat{\lambda}_2}
\end{aligned}$$

with $\hat{\mu}_i$, $\hat{\sigma}_i$, $i=1,2,3$ from Section 7.1. It can, for example, be used to estimate the probability of events beyond the range of the data.

If we use the estimated values $\hat{\lambda}_1$ and $\hat{\lambda}_2$ from the Conditional ML method, we can give the probability that the three places experience high (seasonally adjusted) water discharges within a short time frame (50 days). For example we look at the probability that the seasonally adjusted water discharge at Donauwörth exceeds $1300\,m^3/sec.$, the seasonally adjusted water discharge

at Regensburg exceeds $1500\,m^3/sec.$ and the seasonally adjusted water discharge at Eichstätt exceeds $100\,m^3/sec.$ together within 50 days. I.e., we want to estimate the probability that the three places experience together a severe flood of a magnitude, which has never been recorded before. The estimation for the probability of such an event is

$$\hat{W}_S(1300, 1500, 100) = 0.000254545,$$

where $\hat{W}_S$ denotes the survivor function of $\hat{W}$. I.e., it is estimated that such a flood occurs every $\frac{1}{0.000254545} \cdot \frac{50}{365} \approx 538.2$ years on average. However, such numbers have to be viewed with some care since by taking $\hat{\lambda}_1 = 2.2$ and $\hat{\lambda}_2 = 1.4$, which cannot be excluded by the above shown estimators and standard errors, we get a probability for the above event of 0.000555271 and, thus, an average of approximately 246.7 years for the occurrence of such an event.

Chapter 8

Final Remarks

The preceding text dealt with the investigation of the multivariate generalized Pareto distributions and their role in the framework of extreme value theory. Some theoretical considerations, like the non-uniqueness of the angular measure for GPDs and the generalization of a known counterexample to an arbitrary dimension $d \geq 3$, showing that GP functions are not necessarily distribution functions, were made.

We have developed an algorithm for the simulation of multivariate generalized Pareto distributions of logistic type. This algorithm is easy to implement and generates random vectors for a few thousand dimensions within a reasonable time. The development of this algorithm was based on results by Shi (1995, [64]), Stephenson (2003, [70]) and Section 5.4 of Falk et al. (2004, [21]). We have also introduced algorithms for the generation of other GPDs in low dimensions. An unsolved problem, where much work is to be done in the future, is the efficient simulation of arbitrary GPDs in high dimensions. Except Shi and Zhou (1999, [65]), where a possible generalization of the Shi transformation, with some drawbacks, for the trivariate nested logistic EVD case is given, there is not much known in this area. It is an open problem to find a method similar to the one presented in Section 3.1 for other (parametric) cases than the logistic one.

Ways to estimate nonparametrically and parametrically in generalized Pareto models were presented, following ideas of Coles and Tawn (1991, [10]), Coles and Tawn (1994, [11]) or Coles et al. (1999, [9]) from the extreme value setup. It is surely possible to improve these methods and invent others which must be compared to those presented here.

One point not treated in this manuscript is the problem of checking the generalized Pareto model assumption. We did this briefly for the hydrological data set in Chapter 7 by graphical means, but statistical methods for that are yet unknown. Another open problem remains the testing of parametric model assumptions and the discrimination between different parametric models as stated in Tawn (1988, [72]) or Paper B of Tajvidi (1996, [71]). One way to proceed here could be to use condition (6.41) as a check of the model assumption. It would have to be investigated to what extent condition (6.41) will be violated even if the parametric model is truly underlying the data.

Then a lot of practical questions, which could be treated much more thoroughly than done in Chapter 7, have to be considered when applying the estimation methods in GPD models to real data sets. Examples are the transformation of the margins or the choice of the sample size and appropriate thresholds.

We have mainly considered extreme value distributions with negative exponential margins and

generalized Pareto distributions with uniform margins in this manuscript. The choice of these margins is somewhat arbitrary, since by simple marginal transformations one can consider any desired marginal form. Another popular choice of margins preferred by various authors are the Fréchet margins for EVDs and, thus, the standard Pareto margins for GPDs. There have been discussions, which kind of marginal form should be preferred for theoretical considerations. The result of this manuscript to that question is something like a solomonic verdict, since we have seen both advantages and disadvantages for both choices of margins during the course of this text.

In the simulations, the uniform margins lead to easier definitions of the Shi coordinates whose radial component is just the λ-norm of the desired random vector. Working with Fréchet or Pareto margins, one has to introduce the transformation $x \mapsto \frac{1}{x}$ before taking the λ-norm to compute the radial component. This can be seen at the beginning of Section 2 in Stephenson (2003, [70]), who works with Fréchet margins and where the radial component is the λ-norm of the reciprocal of the desired vector. Thus, the uniform margins seem more natural here, since with the other margins one uses implicitly the uniform margins. Also the uniform margins lead to the independence of the radial and angular component of the Pickands coordinates which was used in the simulations in Section 3.2. The components of Pickands coordinates of GPD random vectors with Pareto margins are not independent, only after the transformation $x \mapsto -\frac{1}{x}$ to uniform margins. And the uniform margins have shown their advantages in the bivariate nonparametric estimations where the angular density could efficiently be estimated, since the distribution of the angular components of the Pickands coordinates was exactly known and independent of the corresponding threshold.

But there have also been disadvantages of the uniform margins. The nonparametric estimation of the angular density in higher dimensions was not possible with the standard Pickands coordinates. Instead, one had to go to Pickands coordinates with regard to Fréchet margins and, thus, implicitly go to Pareto margins by the transformation $x \mapsto -\frac{1}{x}$.

So we see that both marginal choices have their justification, and both should be considered. There does not seem to be a natural choice of the margins, but, instead, the choice of the margins should depend on the problem which is to be considered.

The above mentioned problems are just a small sample of those which have to be worked on for a better understanding of GPDs in the future. In a nutshell, the investigation of the generalized Pareto distributions in the framework of extreme value theory is, still, in its beginning and will surely be an active research area in the coming years.

Bibliography

[1] **B. Abdous, K. Ghoudi, A. Khoudraji** - Non-Parametric Estimation of the Limit Dependence Function of Multivariate Extremes - Extremes 2, pp. 245-268 - 1999

[2] **G. Balkema, P. Embrechts** - Multivariate Excess Distributions - http://www.math.ethz.ch/%7Ebaltes/ftp/guuspe08Jun04.pdf - 2004

[3] **J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels** - Statistics of Extremes, Theory and Applications - Jon Wiley & Sons Chicester - 2005

[4] **R. Bhattacharya, R. Rao** - Normal Approximation and Asymptotic Expansions - Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons New York - 1976

[5] **B. Brommundt** - Value at Risk Schätzung mit Hilfe elliptischer Copulae - Diploma Thesis, Zentrum Mathematik, Technische Universität München - http://www-m4.ma.tum.de/Diplarb/DA_Bernd_Brommundt.pdf - 2003

[6] **P. Capéraà, A.–L. Fougères, C. Genest** - A Nonparametric Estimation Procedure for Bivariate Extreme Value Copulas - Biometrika 84, pp. 567-577 - 1997

[7] **P. Capéraà, A.–L. Fougères** - Estimation of a Bivariate Extreme Value Distribution - Extremes 3, pp. 311-329 - 2000

[8] **S. Coles** - An Introduction to Statistical Modeling of Extreme Values - Springer Series in Statistics, Springer New York - 2001

[9] **S. Coles, J. Heffernan, J. Tawn** - Dependence Measures for Extreme Value Analysis - Extremes 2, pp. 339-365 - 1999

[10] **S. Coles, J. Tawn** - Modelling Extreme Multivariate Events - Journal of the Royal Statistical Society 53, pp. 377-392 - 1991

[11] **S. Coles, J. Tawn** - Statistical Methods for Multivariate Extremes - Applied Statistics 43, pp. 1-48 - 1994

[12] **H. Cramer** - Mathematical Methods of Statistics - Princeton University Press - 1963

[13] **P. Deheuvels** - On the Limiting Behavior of the Pickands Estimator for Bivariate Extreme-Value Distributions - Statistics & Probability Letters 12, pp. 429-439 - 1991

[14] **P. Deheuvels, G. Martynov** - Cramér-von Mises-type Tests with Applications to Tests of Independence for Multivariate Extreme Value Distributions - Communications in Statistics, Theory and Methods 25, pp. 871-908 - 1996

[15] **P. Deheuvels, J. Tiago de Oliveira** - On the Non-Parametric Estimation of the Bivariate Extreme-Value Distributions - Statistics & Probability Letters 8, pp. 315-323 - 1989

[16] **H. Drees, X. Huang** - Best Attainable Rates of Convergence for Estimates of the Stable Tail Dependence Function - Technical Report, Universität Köln - 1995

[17] **H. Drees, X. Huang** - Best Attainable Rates of Convergence for Estimators of the Stable Tail Dependence Function - Journal of Multivariate Analysis 64, pp. 25-47 - 1998

[18] **J. Einmahl, L. deHaan, V. Piterbarg** - Nonparametric Estimation of the Spectral Measure of an Extreme Value Distribution - The Annals of Statistics 29, pp. 1401-1423 - 2001

[19] **J. Einmahl, L. deHaan, A. Sinha** - Estimating the Spectral Measure of an Extreme Value Distribution - Stochastic Processes and their Applications 70, pp. 143-171 - 1997

[20] **M. Falk** - Local Asymptotic Normality of Truncated Empirical Processes - The Annals of Statistics 26, pp. 692-718 - 1998

[21] **M. Falk, J. Hüsler, R.–D. Reiss** - Laws of Small Numbers: Extremes and Rare Events - 2nd edition - Birkhäuser Basel - 2004

[22] **M. Falk, F. Marohn, R. Michel, D. Hofmann, M. Macke, B. Tewes, P. Dinges** - A First Course on Time Series Analysis, Examples with SAS - http://statistik.mathematik.uni-wuerzburg.de/timeseries/ - 2006

[23] **M. Falk, F. Marohn, B. Tewes** - Foundations of Statistical Analyses and Applications with SAS - Birkhäuser Basel - 2002

[24] **M. Falk, R. Michel** - Testing for Tail Independence in Extreme Value Models - Annals of the Institute of Statistical Mathematics 58, pp. 261-290 - 2006

[25] **M. Falk, R.–D. Reiss** - Efficient Estimation of the Canonical Dependence Function - Extremes 6, pp. 61-82 - 2003

[26] **M. Falk, R.–D. Reiss** - On the Distribution of Pickands Coordinates in Bivariate EV and GP models - Journal of Multivariate Analysis 92, pp. 267-295 - 2005

[27] **M. Falk, R.–D. Reiss** - On Pickands Coordinates in Arbitrary Dimensions - Journal of Multivariate Analysis 92, pp. 426-453 - 2005

[28] **B. Fristedt, L. Gray** - A Modern Approach to Probability Theory - Birkhäuser Boston - 1997

[29] **J. Galambos** - The Asymptotic Theory of Extreme Order Statistics - 2nd edition - Krieger Malabar - 1978

[30] **P. Gänssler, W. Stute** - Wahrscheinlichkeitstheorie - Springer Berlin - 1977

[31] **C. Genest, K. Ghoudi, L.–P. Rivest** - A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions - Biometrika 82, pp. 543-552 - 1995

[32] **K. Ghoudi, A. Khoudraji, L.–P. Rivest** - Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles - Canadian Journal of Statistics 26, pp. 187-197 - 1998

[33] **L. deHaan, J. deRonde** - Sea and Wind: Multivariate Extremes at Work - Extremes 1, pp. 7-45 - 1998

[34] **P. Hall, N. Tajvidi** - Distribution and Dependence-Function Estimation for Bivariate Extreme-Value Distributions - Bernoulli 6, pp. 835-844 - 2000

[35] **H. A. Hauksson, M. Dacorogna, T. Domenig, U. Müller, G. Samorodnitsky** - Multivariate Extremes, Aggregation and Risk Estimation - Quantitative Finance 1, pp. 79-95 - 2001

[36] **J. Heffernan** - Nonparametric Estimation of the Bivariate Extreme Value Distribution Using a Limiting Conditional Representation - http://www.maths.lancs.ac.uk/∼currie/Papers/bev.cond.pdf - submitted to Extremes - 2004

[37] **H. Heuser** - Lehrbuch der Analysis 1 - 12th edition - B. G. Teubner Stuttgart - 1998

[38] **H. Heuser** - Lehrbuch der Analysis 2 - 10th edition - B. G. Teubner Stuttgart - 1998

[39] **D. Hofmann** - Über die Darstellung multivariater Extremwertverteilungen mittels Normen - Diploma Thesis, Institut für Mathematik, Bayerische Julius–Maximilians–Universität Würzburg - http://www.danielhofmann.com/downloads/Diplomarbeit.pdf - 2006

[40] **T. Hsing, C. Klüppelberg, G. Kuhn** - Dependence Estimation and Visualization in Multivariate Extremes with Applications to Financial Data - Extremes 7, pp. 99-121 - 2004

[41] **J. Jiménez, E. Villa-Diharce, M. Flores** - Nonparametric Estimation of the Dependence Function in Bivariate Extreme Value Models - Journal of Multivariate Analysis 76, pp. 159-191 - 2001

[42] **H. Joe** - Multivariate Extreme-Value Distributions with Applications to Environmental Data - Canadian Journal of Statistics 22, pp. 47-64 - 1994

[43] **H. Joe, R. Smith, I. Weissman** - Bivariate Threshold Methods for Extremes - Journal of the Royal Statistical Society 54, pp. 171-183 - 1992

[44] **M. Jones, J. Marron, S. Sheather** - A Brief Survey of Bandwidth Selection for Density Estimation - Journal of the American Statistical Association 91, pp. 401-407 - 1996

[45] **E. Kaufmann, R.–D. Reiss** - Approximation Rates for Multivariate Exceedances - Journal of Statistical Planning and Inference 45, pp. 235-245 - 1995

[46] **C. Kelley** - Iterative Methods for Optimization - SIAM, Frontiers in Applied Mathematics 18 - 1999

[47] **S. Kotz, S. Nadarajah** - Extreme Value Distributions, Theory and Applications - Imperial College Press - 2000

[48] **S. Lang** - Linear Algebra - Addison-Wesley, Reading - 1966

[49] **A. Ledford, J. Tawn** - Statistics for Near Independence in Multivariate Extreme Values - Biometrika 83, pp. 169-187 - 1996

[50] **K. Mardia, J. Kent, J. Bibby** - Multivariate Analysis - Academic Press London - 1979

[51] **S. Nadarajah** - Simulation of Multivariate Extreme Values - Journal of Statistical Computation and Simulation 62, pp. 395-410 - 1999

[52] **J. Pfanzagl** - Parametric Statistical Theory - de Gruyter Berlin - 1994

[53] **J. Pickands** - Multivariate extreme value distributions - Proc. 43th Session ISI (Buenos Aires), pp. 859-878 - 1981

[54] **J. Powell** - Notes on Nonparametric Density Estimation - University of Califormia, Berkeley - http://emlab.berkeley.edu/users/powell/e241a_sp05/ndnotes.pdf - 2002

[55] **B. L. S. Prakasa Rao** - Nonparametric Functional Estimation - Academic Press, Orlando - 1983

[56] **R.–D. Reiss** - A Course on Point Processes - Springer Series in Statistics, Springer New York - 1993

[57] **R.–D. Reiss, M. Thomas** - Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields - 2nd edition - Birkhäuser Basel - 2001

[58] **A. Rényi** - Wahrscheinlichkeitsrechnung - VEB Verlag der Wissenschaften, Berlin - 1966

[59] **S. Resnick** - Extreme Values, Regular Variation, and Point Processes - Springer Series of the Applied Probability Trust, Springer New York - 1987

[60] **H. Rootzén, N. Tajvidi** - Multivariate Generalized Pareto Distributions - Preprint 2005:38, Chalmers Tekniska Högskola Göteborg - http://www.math.chalmers.se/Math/Research/Preprints/2005/38.pdf - 2005

[61] **R. Saucier** - Computer Generation of Statistical Distributions - Technical Report ARL-TR-2168, Army Research Laboratory - http://ftp.arl.mil/random/random.pdf - 2000

[62] **M. Schlather, J. Tawn** - A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference - Biometrika 90, pp. 139-156 - 2003

[63] **R. Serfling** - Approximation Theorems of Mathematical Statistics - Jon Wiley & Sons New York - 1980

[64] **D. Shi** - Multivariate Extreme Value Distribution and its Fisher Information Matrix - ACTA Mathematicae Applicatae Sinica 11, pp. 421-428 - 1995

[65] **D. Shi, S. Zhou** - Moment Estimation for Multivariate Extreme Value Distributions in a Nested Logistic Model - Annals of the Institute of Statistical Mathematics 51, pp. 253-264 - 1999

[66] **S. Sheather, M. Jones** - A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation - Journal of the Royal Statistical Society 53, pp. 683-690 - 1991

[67] **J. Simonoff** - Smoothing Methods in Statistics - Springer New York - 1996

[68] **R. Smith, J. Tawn, H. Yuen** - Statistics of Multivariate Extremes - International Statistical Review 58, pp. 47-58 - 1990

[69] **R. Smith, J. Tawn, S. Coles** - Markov Chain Models for Threshold Exceedances - Biometrika 84, pp. 249-268 - 1997

[70] **A. Stephenson** - Simulating Multivariate Extreme Value Distributions of Logistic Type - Extremes 6, pp. 49-59 - 2003

[71] **N. Tajvidi** - Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalised Pareto Distributions - Dissertation, Department of Mathematics, Chalmers Tekniska Högskola Göteborg - http://www.math.chalmers.se/∼nader/thesis.ps - 1996

[72] **J. Tawn** - Bivariate Extreme Value Theory: Models and Estimation - Biometrika 75, pp. 397-415 - 1988

[73] **J. Tawn** - Modelling Multivariate Extreme Value Distributions - Biometrika 77, pp. 245-253 - 1990

[74] **W. Wertz** - Statistical Density Estimation, A Survey - Vandenhoeck & Ruprecht - 1978

[75] **H. Witting, G. Nölle** - Angewandte Mathematische Statistik - Teubner Stuttgart - 1970