Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Lehrstuhl für Mathematik VIII
Statistik

Dissertation zur Erlangung des Akademischen Grades eines

Doktors der Naturwissenschaften

Some Applications of D-Norms to Probability and Statistics

vorgelegt von

Florian Wisheckel

im September 2020 in Würzburg

"A change in perspective is worth 80 IQ points."

- Alan Key (1982) -

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

The topic at hand is a rather recent one. To be fair, already mathematicians like Nicolaus Bernoulli were interested in the investigation of maxima, the key subject of extreme value theory. Specifically, he wanted to figure out the mean duration of the last survivor's life from a population of n people. In the 19th century, scientists were interested in detecting outliers (e.g., Peirce's criterion) and generally large observations coming from a normal distribution. Some time later, other distributions were considered, too, but it was not until mid 20th century that the asymptotic univariate types of extreme value distributions were discovered: the crucial theorem is by Gnedenko (1943), building on earlier work by Fisher and Tippett (1928) and Fréchet (1927). Gumbel (1958) compiled lots of this knowledge in his book and more on this history can be found there as well.

We still have to wait some more years, until, in the 1970ies, the attention of the extreme value community shifted towards multivariate extremes. This step is important to us as now we have to deal with dependence structures between the different components of the limiting distribution. In fact, quite a few views how to characterize this dependence developed over the next years, e.g., the spectral or angular measure (Einmahl et al. (1997)), Pickands dependence function (Pickands (1981)) or the stable tail dependence function (Huang (1992)). Interpreting the latter as D-norms, we can gather lots of interesting insights. This is why we make this view the key element of this thesis.

In the next section of this introductory chapter, we first establish some notation and other preliminaries needed throughout the remaining chapters. Each chapter itself will introduce the respective setup and more theory that is needed for that part, but not for the others.

The following three chapters are adapted versions of existing articles. They are unified in the sense that common topics were compiled together in this introduction and contain some remarks that did not make it into the final version of the published papers.

The second chapter, based on Falk and Wisheckel (2017), is an investigation of the asymptotic dependence behavior of the components of bivariate order statistics. We find that the two components of the order statistics become asymptotically independent for certain combinations of (sequences of) indices that are selected, and it turns out that no further assumptions on the dependence of the two components in the underlying sample are necessary. To establish this, an explicit representation of the conditional distribution of bivariate order statistics is derived.

Chapter 3 is from Falk et al. (2019*a*) and deals with the conditional distribution of an Archimedean copula, conditioned on one of its components. We show that its tails are independent under minor conditions on the generator function, even if the unconditional tails were dependent. The theoretical findings are underlined by a simulation study and can be generalized to Archimax copulas.

Falk et al. (2019*b*) lead to Chapter 4 where we introduce a nonparametric approach to estimate the probability that a random vector exceeds a fixed threshold if it follows a Generalized Pareto copula. To this end, some theory underlying the concept of Generalized Pareto distributions is presented first, the estimation procedure is tested using a simulation and finally applied to a dataset of air pollution parameters in Milan, Italy, from 2002 until 2017.

While the previous three chapters are adaptations of already completed and published papers, the fifth chapter collects several additional results and considerations.

Finally, the sixth chapter provides a short review of the whole thesis as well as an outlook to possible further work.

1.2 Preliminaries

General Notation

As announced, we want to clarify some notations first: random vectors like $\boldsymbol{X}$ that take values in $\mathbb{R}^d$, $d > 1$, will be printed in bold face and using upper case letters to indicate that we are in a multivariate context. Non-random vectors like $\boldsymbol{a} \in \mathbb{R}^d$ also appear in bold face, but are lower case.

If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ are two d-dimensional vectors, we define operations and relations between them component-wise. In particular we have

$$\begin{aligned} \boldsymbol{x}\boldsymbol{y} &:= (x_1 y_1, \dots, x_d y_d)^T \\ \frac{\boldsymbol{x}}{\boldsymbol{y}} &:= \left(\frac{x_1}{y_1}, \dots, \frac{x_d}{y_d}\right)^T \\ \max(\boldsymbol{x}, \boldsymbol{y}) &:= (\max(x_i, y_i))_{i=1}^d \\ \boldsymbol{x} \le \boldsymbol{y} &:\Leftrightarrow x_1 \le y_1, \dots, x_d \le y_d \end{aligned}$$

and so on. The same goes for random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ as long as they are defined on the same probability space. We assume that this is the case for random vectors in a common context. Furthermore $\mathbf{1} = (1, \dots, 1)^T$, $\mathbf{0} = (0, \dots, 0)^T \in \mathbb{R}^d$ denote the vector of constant ones or zeros of the appropriate dimension for the given context. As a notational shorthand we sometimes use scalars $c \in \mathbb{R}$ to mean $c\mathbf{1}$. Note however that we still have $\boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^d x_i y_i$ as usual.

Matrices are denoted by upper case letters. Their components are accessed by double-subscripts like $\Sigma = (\sigma_{ij})$ with $i, j = 1, \dots, d$. For some vector $\boldsymbol{c} \in \mathbb{R}^d$, $C = \operatorname{diag}(\boldsymbol{c}) \in \mathbb{R}^{d \times d}$ is a diagonal matrix with the entries of $\boldsymbol{c}$ on the main diagonal and 0 elsewhere.

When we have a collection of (random) vectors, we sometimes put the index as a superscript to leave the subscript for the components, i.e.,

$$\boldsymbol{X}^{(i)} = \left(X_1^{(i)}, \dots, X_d^{(i)}\right)^T.$$

As long as there is no risk of confusion with exponents, we may leave out the parentheses and just write $\boldsymbol{X}^i$ instead of $\boldsymbol{X}^{(i)}$.

Convergence of x to x_0 is denoted by $x \to x_0$, the one-sided limit from above or from the right, i.e., for $x > x_0$, is written as $x \downarrow x_0$. Likewise, $x \uparrow x_0$ is the limit from below or from the left.

Order Statistics and Extreme Value Distributions

Consider a sample $X_1, \dots, X_n$ of $n \in N$ random variables (rv). Most of the time, we will require them to be an independent and identically distributed (iid) sample, but this is not necessary for a general definition of order statistics. Let $X_{m:n}$ for $m = 1, \dots, n$ denote the m-th order statistic (os) of our sample, which is the m-th smallest observation. This means that we can sort the random variables in ascending order by $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ (for fixed ω in the sample space Ω, which means the order statistics are rvs, too). Therefore, $X_{n:n} = M_n := \max(X_1, ..., X_n)$ is the maximum and $X_{1:n} = m_n := \min(X_1, ..., X_n)$ is the minimum. $X_{\lceil \frac{n}{2} \rceil :n}$ is the median where $\lceil \frac{n}{2} \rceil$ is the ceiling function, i.e., the smallest integer greater than or equal to $\frac{n}{2}$. Finally, $X_{n-k+1:n}$ is the k-th largest order statistic, so $k = 1$ leads back to the maximum.

For the multivariate case, let $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)} \in \mathbb{R}^d$ be a sample of n d-variate random vectors (rv). We define $\boldsymbol{X}_{\boldsymbol{m}:n}$ with $\boldsymbol{m} \in \{1, \dots, n\}^d$ as the vector of component-wise order statistics, i.e., $\boldsymbol{X}_{\boldsymbol{m}:n} = (X_{m_1:n,1}, \dots, X_{m_d:n,d})^T$ where $X_{m_i:n,i}$ is the m_ith order statistic of $X_i^{(1)}, \dots, X_i^{(n)}$. Note that in most cases $\boldsymbol{X}_{\boldsymbol{m}:n}$ is not an element of the sample as its components typicallycome from different observations.

Order statistics themselves are mainly investigated in Chapters 2 and 5, but a special one - the maximum - is the key subject of extreme value theory: let us assume that the sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \in \mathbb{R}^d$ is iid with common distribution function (df) F. Then the distribution of $\boldsymbol{M}_n$ is given by

$$P\left(\boldsymbol{M}_n \leq \boldsymbol{x}\right) = P\left(\boldsymbol{X}_i \leq \boldsymbol{x},\, i = 1, \dots, n\right) = P\left(\boldsymbol{X}_1 \leq \boldsymbol{x}_n\right)^n = F^n(\boldsymbol{x}).$$

If $n \to \infty$, this can only converge to 0 or 1. To have a chance for convergence to a non-degenerate df, we consider an affine transformation. We use sequences $\boldsymbol{a}_n > 0, \boldsymbol{b}_n \in \mathbb{R}^d$ and get

$$P\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{x}\right) = F^n\left(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}\right).$$

We say that the df F is *in the domain of attraction* of a multivariate, non-degenerate df G, denoted by $F \in \mathcal{D}(G)$, if and only if (iff) there exist vectors $\boldsymbol{a}_n > \boldsymbol{0} \in \mathbb{R}^d$, $\boldsymbol{b}_n \in \mathbb{R}^d$, $n \in \mathbb{N}$, such that

$$F^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \to_{n \to \infty} G(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d. \tag{1.1}$$

The limit df G is necessarily *max-stable*, see, e.g., Resnick (1987, Proposition 5.9). This means that there exist vectors $\boldsymbol{a}_n > \boldsymbol{0} \in \mathbb{R}^d$, $\boldsymbol{b}_n \in \mathbb{R}^d$, $n \in \mathbb{N}$, such that

$$G^n(\boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n) = G(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

A characterization of multivariate max-stable df was established by de Haan and Resnick (1977) and Vatan (1985); for an introduction to multivariate extreme value theory see, e.g., Falk et al. (2011, Chapter 4).

The univariate margins G_i, $1 \leq i \leq d$, of a multivariate max-stable df G belong necessarily to the family of univariate max-stable df, which is a parametric family $\{G_\alpha : \alpha \in \mathbb{R}\}$ with

$$G_\alpha(x) = \begin{cases} \exp\left(-(-x)^\alpha\right), & x \leq 0, \\ 1, & x > 0, \end{cases} \qquad \text{for } \alpha > 0,$$

$$G_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^\alpha), & x > 0, \end{cases} \qquad \text{for } \alpha < 0,$$

and

$$G_0(x) := \exp(-e^{-x}), \qquad x \in \mathbb{R}, \tag{1.2}$$

being the family of reverse Weibull, Fréchet and Gumbel distributions. Note that $G_1(x) = \exp(x)$, $x \leq 0$, is the standard *negative exponential* df. We refer e.g., to Galambos (1987, Section 2.3) or Resnick (1987, Chapter 1).

Copulas

By Sklar's theorem (Sklar (1959, 1996)), there exists a rv $\boldsymbol{U} = (U_1, \dots, U_d)$ with the property that each component U_i follows the uniform distribution on $(0,1)$, such that

$$\boldsymbol{X} =_D \left(F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)\right),$$

where F_i is the df of X_i and $F_i^{-1}(u) = \inf\{t \in \mathbb{R} : F_i(t) \geq u\}$, $u \in (0,1)$, is the common generalized inverse or quantile function of F_i, $1 \leq i \leq d$. By $=_D$ we denote equality in distribution.

The rv $\boldsymbol{U}$, therefore, follows a *copula*, say C_F. If F is continuous, then the copula C_F is uniquely determined by $C_F(\boldsymbol{u}) = F\left(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\right)$, $\boldsymbol{u} = (u_1, \dots, u_d) \in (0,1)^d$.

For a general introduction to copulas, we refer to Nelsen (2006). The significance of copulas to extreme value theory can be seen as follows: Deheuvels

(1984) and Galambos (1987) showed that $F \in \mathcal{D}(G)$ iff this is true for each univariate margin F_i *and* for the copula C_F. Precisely, they established the following result.

Theorem 1.1 (Deheuvels (1984), Galambos (1987))**.** *The df F satisfies $F \in \mathcal{D}(G)$ iff this is true for the univariate margins of F together with the convergence of the copulas:*

$$C_F^n\left(\boldsymbol{u}^{1/n}\right) \to_{n\to\infty} C_G(\boldsymbol{u}) = G\left(\left(G_i^{-1}(u_i)\right)_{i=1}^d\right), \tag{1.3}$$

$\boldsymbol{u} = (u_1, \dots, u_d) \in (0,1)^d$*, where G_i denotes the i-th margin of G, $1 \le i \le d$.*

Let $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \dots$ be independent copies of the rv $\boldsymbol{U}$, which follows the copula C_F. Then the copula $C_{\boldsymbol{M}_n}$ of

$$\boldsymbol{M}_n := \max_{1\le i\le n} \boldsymbol{U}^{(i)}$$

is $C_F^n\left(\boldsymbol{u}^{1/n}\right)$, where the maximum is also taken componentwise. The df of $\boldsymbol{M}_n$ is C_F^n and, thus, we have

$$C_F^n\left(\boldsymbol{u}^{1/n}\right) = C_{\boldsymbol{M}_n}(\boldsymbol{u}) = C_{C_F^n}(\boldsymbol{u}), \qquad \boldsymbol{u} \in [0,1]^d.$$

Therefore, condition (1.3) actually means pointwise convergence of the copulas

$$C_{\boldsymbol{M}_n}(\boldsymbol{u}) \to_{n\to\infty} C_G(\boldsymbol{u}),$$

where $C_G(\boldsymbol{u}) = G\left(\left(G_i^{-1}(u_i)\right)_{i=1}^d\right)$, $\boldsymbol{u} \in (0,1)^d$, is the copula of G. This is an *extreme value copula*. Note that each margin G_i of G is continuous, which is equivalent with the continuity of G (see, e.g., Reiss (1989, Lemma 2.2.6)).

Elementary arguments imply that condition (1.3) is equivalent with the condition

$$C_F^n\left(\boldsymbol{1} + \frac{\boldsymbol{y}}{n}\right) \to_{n\to\infty} G^*(\boldsymbol{y}) := C_G(\exp(\boldsymbol{y})), \qquad \boldsymbol{y} \le \boldsymbol{0} \in \mathbb{R}^d, \tag{1.4}$$

where $\boldsymbol{1} = (1, \dots, 1) \in \mathbb{R}^d$ and $G^*(\boldsymbol{y})$, $\boldsymbol{y} \le \boldsymbol{0} \in \mathbb{R}^d$, defines a max-stable df with standard negative exponential margins $G_i^*(y) = \exp(y)$, $y \le 0$, $1 \le i \le d$. Such a max-stable df will be called a *standard* one, abbreviated by SMS (standard max-stable).

While the condition on the univariate margins F_i in Theorem 1.1 addresses *univariate* extreme value theory, condition (1.3) on the copula C_F means by the equivalent condition (1.4) that the copula C_F is in the domain of attraction of a multivariate SMS df:

$$C_F^n\left(\mathbf{1}+\frac{\boldsymbol{y}}{n}\right) = P(n(\boldsymbol{M}_n-\mathbf{1})\le \boldsymbol{y}) \to_{n\to\infty} G^*(\boldsymbol{y}), \qquad \boldsymbol{y}\le \mathbf{0}\in\mathbb{R}^d.$$

Let C be an arbitrary copula on $\mathbb{R}^d$. Then condition (1.1) becomes

$$C\in\mathcal{D}(G) \iff C^n(\boldsymbol{a}_n\boldsymbol{x}+\boldsymbol{b}_n)\to_{n\to\infty} G(\boldsymbol{x}), \qquad \boldsymbol{x}\in\mathbb{R}^d,$$

where the norming constants $\boldsymbol{a}_n, \boldsymbol{b}_n$ are determined by the univariate margins C_i of C, i.e., the uniform distribution on $(0,1)$: With $a_n = 1/n$, $b_n = 1$ we obtain for large n

$$C_i(a_n x + b_n)^n = \left(1+\frac{x}{n}\right)^n \to_{n\to\infty} \exp(x), \qquad x\le 0.$$

We therefore obtain the conclusion: If a copula C satisfies $C\in\mathcal{D}(G)$, then the limiting df G has necessarily standard negative exponential margins:

$$G_i(x) = \exp(x), \qquad x\le 0,\ 1\le i\le d,$$

i.e., the limiting df G is necessarily a SMS df.

As a consequence we obtain that *multivariate* extreme value theory actually means extreme value theory for *copulas*.

D-Norms

A crucial characterization of SMS df due to Balkema and Resnick (1977), de Haan and Resnick (1977), Pickands (1981) and Vatan (1985) can be formulated as follows; see Falk (2019, Theorem 2.3.3).

Theorem 1.2 (Balkema and Resnick (1977), de Haan and Resnick (1977), Pickands (1981), Vatan (1985))**.** *A df G on $\mathbb{R}^d$ is an SMS df iff there exists a norm $\|\cdot\|$ on $\mathbb{R}^d$ such that*

$$G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|), \qquad \boldsymbol{x}\le\mathbf{0}\in\mathbb{R}^d. \tag{1.5}$$

Elementary arguments imply the following consequence.

Corollary 1.3. *A copula C satisfies $C\in\mathcal{D}(G)$ iff there exists a norm $\|\cdot\|$ on $\mathbb{R}^d$ such that*

$$C(\boldsymbol{u}) = 1-\|\mathbf{1}-\boldsymbol{u}\| + o(\|\mathbf{1}-\boldsymbol{u}\|) \tag{1.6}$$

as $\boldsymbol{u}\to\mathbf{1}\in\mathbb{R}^d$, uniformly for $\boldsymbol{u}\in[0,1]^d$.

Those norms, which can appear in the preceding result, can be characterized. Any norm $\|\cdot\|$ in equation (1.5) or (1.6) is necessarily of the following kind: There exists a rv $\boldsymbol{Z} = (Z_1, \dots, Z_d)$, whose components satisfy

$$Z_i \geq 0, \quad E(Z_i) = 1, \qquad 1 \leq i \leq d,$$

with

$$\|\boldsymbol{x}\| = E\left(\max_{1\leq i\leq d} (|x_i|\, Z_i)\right) =: \|\boldsymbol{x}\|_D\,,$$

$\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Such a norm $\|\cdot\|_D$ is called *D-norm*, with *generator* $\boldsymbol{Z}$. The additional index D means *dependence*. D-norms were first mentioned in Falk et al. (2004, equation (4.25)) and more elaborated in Falk et al. (2011, Section 4.4). Examples are:

- $\|\boldsymbol{x}\|_\infty = \max_{1\leq i\leq d} |x_i|$, with generator $\boldsymbol{Z} = (1, \dots, 1) \in \mathbb{R}^d$,
- $\|\boldsymbol{x}\|_1 = \sum_{i=1}^d |x_i|$, with generator $\boldsymbol{Z}$ being a random permutation of the vector $(d, 0, \dots, 0) \in \mathbb{R}^d$,
- each logistic norm $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, $p \in (1, \infty)$, with generator $\boldsymbol{Z} = (Z_1, \dots, Z_d) = (Y_1^{1/p}, \dots, Y_d^{1/p})/\Gamma(1 - 1/p)$, where $Y_1, \dots, Y_d$ are iid Fréchet-distributed rvs, i.e., $P(Y_i < y) = \exp(-1/y)$ for $y > 0$ and $j = 1, \dots, d$, and Γ denotes the usual gamma function.
- Let the rv $\boldsymbol{X} = (X_1, \dots, X_d)$ follow a multivariate normal distribution with mean vector zero, i.e., $E(X_i) = 0$, $1 \leq i \leq d$, and covariance matrix $\Sigma = (\sigma_{ij})_{1\leq i,j\leq d} = (E(X_i X_j))_{1\leq i,j\leq d}$. Then $\exp(X_i)$ follows a log-normal distribution with mean $\exp(\sigma_{ii}/2)$, $1 \leq i \leq d$, and, thus,

  $$\boldsymbol{Z} = (Z_1, \dots, Z_d) := \left(\exp\left(X_1 - \frac{\sigma_{11}}{2}\right), \dots, \exp\left(X_d - \frac{\sigma_{dd}}{2}\right)\right)$$

  is the generator of a D-norm, called *Hüsler-Reiss D-norm*. This norm only depends on the covariance matrix Σ and, therefore, it is denoted by $\|\cdot\|_{\mathrm{HR}_\Sigma}$.

Note that $\|\cdot\|_\infty$ is the smallest D-norm and $\|\cdot\|_1$ is the largest in the sense that we have $\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_D \leq \|\boldsymbol{x}\|_1$ for all $\boldsymbol{x} \in \mathbb{R}^d$ and arbitrary D-norm $\|\cdot\|_D$.

The generator of a D-norm is in general not uniquely determined, even its distribution is not. Take, for example, a rv $X > 0$ with $E(X) = 1$. Then $\boldsymbol{Z} = (Z_1, \dots, Z_d) = (X, \dots, X)$ generates the sup-norm $\|\cdot\|_\infty$. But this is also generated by the constant $\boldsymbol{1} = (1, \dots, 1) \in \mathbb{R}^d$. An account of the theory of D-norms is provided by Falk (2019).

Generalized Pareto Copulas

Corollary 1.3 stimulates the following idea. Choose an arbitrary D-norm $\|\cdot\|_D$ on $\mathbb{R}^d$ and put

$$C(\boldsymbol{u}) := \max\left(1 - \|\boldsymbol{1} - \boldsymbol{u}\|_D, 0\right), \qquad \boldsymbol{u} \in [0,1]^d.$$

Each univariate margin C_i of C, defined this way, satisfies for $u \in [0,1]$

$$\begin{aligned} C_i(u) &= C(1, \dots, 1, \underbrace{u}_{i\text{-th component}}, 1 \dots, 1) \\ &= 1 - \|(0, \dots, 0, 1-u, 0, \dots, 0)\|_D \\ &= 1 - (1-u)\underbrace{E(Z_i)}_{=1} = u, \end{aligned}$$

i.e., each C_i is the uniform df on $(0,1)$. But C does in general *not* define a df, see, e.g., Falk et al. (2011, Proposition 5.1.3). We require, therefore, the expansion

$$C(\boldsymbol{u}) = 1 - \|\boldsymbol{1} - \boldsymbol{u}\|_D$$

only for $\boldsymbol{u}$ close to $\boldsymbol{1} \in \mathbb{R}^d$, i.e., for $\boldsymbol{u} \in [\boldsymbol{u}_0, \boldsymbol{1}] \subset \mathbb{R}^d$ with some $\boldsymbol{0} < \boldsymbol{u}_0 < \boldsymbol{1} \in \mathbb{R}^d$. A copula C with this property will be called a *generalized Pareto copula* (GPC). These copulas were introduced in Aulbach et al. (2012); tests, whether data are generated by a copula in a δ-neighborhood of a GPC were derived in Aulbach et al. (2018), see Section 4.3 for the precise definition of this neighborhood. The multivariate generalized Pareto distributions defined in Section 4.2 show that GPC actually exist for any D-norm $\|\cdot\|_D$. The corresponding construction of a generalized Pareto distributed rv also provides a way to simulate data from an arbitrary GPC.

As a consequence, an *arbitrary* copula C satisfies the following equivalences:

$$\begin{aligned} &C \in \mathcal{D}(G) \\ &\iff C(\boldsymbol{u}) = 1 - \|\boldsymbol{1} - \boldsymbol{u}\|_D + o(\|\boldsymbol{1} - \boldsymbol{u}\|) \quad \text{for some } D\text{-norm } \|\cdot\|_D \\ &\iff C \text{ is in its upper tail close to that of a GPC.} \end{aligned}$$

In this case we have $G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_D)$, $\boldsymbol{x} \le \boldsymbol{0} \in \mathbb{R}^d$.

Archimedean Copulas

Take an arbitrary *Archimedean copula* on $\mathbb{R}^d$

$$C_\varphi(\boldsymbol{u}) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)),$$

where φ is a continuous and strictly decreasing function from $(0,1]$ to $[0,\infty)$ such that $\varphi(1) = 0$ (see, e.g., McNeil and Nešlehová (2009, Theorem 2.2)). As we focus on extreme value analysis, we can relax this condition a bit and require this representation only for $\boldsymbol{u}$ in a (left) neighborhood of $\mathbf{1}$. Suppose that

$$p := -\lim_{s\downarrow 0} \frac{s\varphi'(1-s)}{\varphi(1-s)} \text{ exists in } [1,\infty). \tag{1.7}$$

It follows from Charpentier and Segers (2009, Theorem 4.1) that C is in its upper tail close to the GPC with corresponding logistic D-norm $\|\cdot\|_p$.

Suppose that the generator function $\varphi : (0,1] \to [0,\infty)$ satisfies with some $s_0 \in (0,1)$

$$-\frac{s\varphi'(1-s)}{\varphi(1-s)} = p, \qquad s \in (0, s_0], \tag{1.8}$$

with $p \in [1,\infty)$. Then C_φ is a GPC, precisely,

$$C_\varphi(\boldsymbol{u}) = 1 - \|\mathbf{1} - \boldsymbol{u}\|_p = 1 - \left(\sum_{i=1}^d |1 - u_i|^p\right)^{1/p}, \qquad \boldsymbol{u} \in [1 - s_0, 1]^d.$$

This is readily seen as follows. Condition (1.8) is equivalent with the equation

$$(\log(\varphi(1-s)))' = \frac{p}{s}, \qquad s \in (0, s_0].$$

Integrating both sides implies

$$\log(\varphi(1-s)) - \log(\varphi(1-s_0)) = p\log(s) - p\log(s_0)$$

or

$$\log\left(\frac{\varphi(1-s)}{\varphi(1-s_0)}\right) = \log\left(\left(\frac{s}{s_0}\right)^p\right), \qquad s \in (0, s_0],$$

which implies

$$\varphi(1-s) = \frac{\varphi(1-s_0)}{s_0^p} s^p, \qquad s \in [0, s_0],$$

i.e.,

$$\varphi(s) = c(1-s)^p, \qquad s \in [1-s_0, 1],$$

with $c := \varphi(1-s_0)/s_0^p$. But this yields

$$\begin{aligned} C_\varphi(\boldsymbol{u}) &= \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)) \\ &= 1 - \left(\sum_{i=1}^d (1-u_i)^p\right)^{1/p}, \qquad \boldsymbol{u} \in [1-s_0, 1]^d. \end{aligned}$$

This should be enough to dive deeper into the topics of the individual chapters. Anything further will be introduced there.

Chapter 2

Asymptotic Independence of Bivariate Order Statistics

2.1 Introduction

Let $U_1, \dots, U_n$ be independent copies of a univariate rv U and denote by $U_{1:n} \le U_{2:n} \le \dots \le U_{n:n}$ the pertaining os. It follows from Theorem 1.3 in Falk and Reiss (1988) that there exists a universal constant such that for $1 \le r \le n-k+1 \le n$ and $n \in \mathbb{N}$

$$\sup_{x,y\in\mathbb{R}} |P(U_{r:n} \le x, U_{n-k+1:n} \le y) - P(U_{r:n} \le x)P(U_{n-k+1:n} \le y)| \le \text{const} \left(\frac{rk}{n(n-r-k+1)}\right)^{1/2}. \tag{2.1}$$

This upper bound converges to 0 if we consider a sequence $r = r(n)$ that satisfies $r/n \to_{n\to\infty} \lambda \in (0,1)$ together with $k = k(n) \to_{n\to\infty} \infty$, $k/n \to_{n\to\infty} 0$. Then $(U_{r:n})$ is a sequence of central os, $(U_{n-k+1:n})$ a sequence of intermediate os and the limiting 0 shows that they become asymptotically independent. The same holds for an intermediate sequence $r = r(n)$ together with fixed k, i.e., extreme os.

Starting with the work by Gumbel (1946) on extremes, the asymptotic independence of order statistics has been investigated in quite a few articles. For detailed references we refer to Galambos (1987, p. 150) and to Falk and Kohne (1986).

By the quantile transformation theorem (see, e.g. Reiss (1989, Lemma

1.2.4)) we can assume without loss of generality in the preceding result (2.1) that U follows the uniform distribution on $(0,1)$.

Let $(U_1,V_1),\dots,(U_n,V_n)$ be independent copies of the bivariate rv (U,V) that follows a copula C. Choose $r,k\in\{1,\dots,n\}$ and consider the bivariate vector $(U_{r:n},V_{k:n})$ of componentwise os. In this chapter we investigate the problem, whether asymptotic independence also holds for $(U_{r:n},V_{k:n})$ with proper sequences $r=r(n)$, $k=k(n)$.

Note that, for example, $U_{r:n}$ and $U_{n-r+1:n}$ with r fixed get by inequality (2.1) asymptotically independent, but $U_{r:n}$ and $V_{n-r+1:n}$ might not. Consider $(U,V):=(U,1-U)$. Then the joint distribution of (U,V) is a copula as well but $V_{n-r+1:n}=1-U_{r:n}$.

For $r=k=n$, the asymptotic joint distribution of $(U_{n:n},V_{n:n})$ is provided by multivariate extreme value theory. Precisely, if $n(U_{n:n}-1,V_{n:n}-1)$ has a non-degenerate limit distribution G, then this limit has the representation

$$G(x,y)=\exp\left(-\left\|(x,y)\right\|_D\right),\qquad x,y\le 0,$$

where $\left\|\cdot\right\|_D$ is a suitable *D-norm* on $\mathbb{R}^2$, see Theorem 1.2 and Falk (2019, Theorem 2.3.3).

The limit distribution of $n(U_{n-i+1:n}-1,V_{n-j+1:n}-1)$ with fixed i,j was established by Galambos (1975). The set of limiting distributions in the intermediate case $(U_{n-k:n},V_{n-r:n})$ with $k=k(n)$, $r=r(n)$ both converging to infinity as n increases, but $(k+r)/n\to_{n\to\infty}0$, was identified by Cheng et al. (1997). If in particular $n(U_{n:n}-1,V_{n:n}-1)$ converges in distribution to G as above, then $(n/\sqrt{k})\left(U_{n-k:n}-(n-k)/(n+1),V_{n-k:n}-(n-k)/(n+1)\right)$ follows asymptotically the bivariate normal distribution with mean vector $\mathbf{0}\in\mathbb{R}^2$ and covariance matrix $\begin{pmatrix}1 & 2-\|(1,1)\|_D\\ 2-\|(1,1)\|_D & 1\end{pmatrix}$ as shown by Falk and Wisheckel (2018). Asymptotic normality of $(U_{r:n},V_{k:n})$ in the central case, where $r/n\to_{n\to\infty}\lambda_1$, $k/n\to_{n\to\infty}\lambda_2$, $0<\lambda_1,\lambda_2<1$, is established in Reiss (1989).

In what follows we will establish

$$\sup_{x,y\in\mathbb{R}}\left|P(U_{r:n}\le x,V_{k:n}\le y)-P(U_{r:n}\le x)P(V_{k:n}\le y)\right|\to_{n\to\infty}0,$$

for various choices of $r=r(n)$ and $k=k(n)$. It turns out that for such sequences asymptotic independence holds with no further assumptions on the copula C. The main tool will be Lemma 2.2, in which the conditional distribution function $P(U_{r:n}\le x\mid V_{k:n}=y)$ is derived for arbitrary $r,k\in\{1,\dots,n\}$. This powerful tool should be of interest of its own.

2.2 Conditional Expectation of Bivariate OS

In this section we compute the conditional probability $P(U_{m:n} \leq x \mid V_{k:n} = y)$ for arbitrary $m, k \in \{1, \dots, n\}$ as a major tool. For the formulation of Lemma 2.2 and its proof it is quite convenient to explicitly quote Theorem 2.2.7 in Nelsen (2006).

Theorem 2.1 (Nelsen (2006))**.** *Let C be an arbitrary bivariate copula. For any $x \in [0,1]$, the partial derivative $\frac{\partial}{\partial y} C(x,y)$ exists for almost all y, and for such x and y*

$$0 \leq \frac{\partial}{\partial y} C(x,y) \leq 1. \tag{2.2}$$

Furthermore, the function $x \mapsto \frac{\partial}{\partial y} C(x,y)$ is defined and nondecreasing almost everywhere on $[0,1]$.

Now we are ready to state our major tool: we show that the conditional probability $P(U_{m:n} \leq x \mid V_{k:n} = y)$ is the linear combination of two probabilities concerning sums of independent Bernoulli rv. We set, as usual, $U_{0:n} = V_{0:n} = 0$ and $U_{n+1:n} = V_{n+1:n} = 1$

Lemma 2.2. *Let $(U_1, V_1), \dots, (U_n, V_n)$, $n \in \mathbb{N}$, be independent copies of a rv (U, V) that follows a copula C. Then we obtain for $1 \leq k, m \leq n$ and for almost every $x, y \in [0,1]$*

$$\begin{aligned} &P(U_{m:n} \leq x \mid V_{k:n} = y) \\ &= P\left(\sum_{i=1}^{k-1} 1_{[0,x]}\left(U_i^{(1)}\right) + \sum_{i=1}^{n-k} 1_{[0,x]}\left(U_i^{(2)}\right) \geq m \right) \\ &\qquad + \frac{\partial}{\partial y} C(x,y) P\left(\sum_{i=1}^{k-1} 1_{[0,x]}\left(U_i^{(1)}\right) + \sum_{i=1}^{n-k} 1_{[0,x]}\left(U_i^{(2)}\right) = m-1 \right) \end{aligned} \tag{2.3}$$

where $U_1^{(1)}, \dots, U_{k-1}^{(1)}, U_1^{(2)}, \dots, U_{n-k}^{(2)}$ are independent rv with

$$P\left(U_i^{(1)} \leq u\right) = P(U \leq u \mid V \leq y) = \frac{C(u,y)}{y}$$

and

$$P\left(U_i^{(2)} \leq u\right) = P(U \leq u \mid V > y) = \frac{u - C(u,y)}{1-y}, \qquad 0 \leq u \leq 1.$$

If we choose, for example, $m = k = n$, then we obtain from the preceding result the representation

$$P(U_{n:n} \le x \mid V_{n:n} = y) = \frac{\partial}{\partial y} C(x,y) P\left(\sum_{i=1}^{n-1} 1_{[0,x]}\left(U_i^{(1)}\right) = n-1\right) = \frac{\partial}{\partial y} C(x,y) \frac{C(x,y)^{n-1}}{y^{n-1}}.$$

Proof of Lemma 2.2. We have

$$\begin{aligned} &P(U_{m:n} \le x \mid V_{k:n} = y) \\ &= \lim_{\varepsilon \downarrow 0} \frac{P(U_{m:n} \le x,\, V_{k:n} \in [y, y+\varepsilon])}{P(V_{k:n} \in [y, y+\varepsilon])} \\ &= \lim_{\varepsilon \downarrow 0} \frac{P(U_{m:n} \le x,\, V_{k:n} \le y+\varepsilon) - P(U_{m:n} \le x,\, V_{k:n} \le y)}{\varepsilon} \frac{\varepsilon}{P(V_{k:n} \in [y, y+\varepsilon])}, \end{aligned}$$

where the second term on the right hand side above converges to $1/g_{k,n}(y)$ as $\varepsilon \downarrow 0$, where $g_{k,n}(\cdot)$ is the Lebesgue-density of $V_{k:n}$, see, e.g., Reiss (1989, Theorem 1.3.2).

In the next step we will break the set $\{U_{m:n} \le x,\, V_{k:n} \le y\}$ into disjoint subsets. By T, S we denote arbitrary subsets of $\{1, \dots, n\}$ and by $|T|$, $|S|$ their cardinalities, i.e., the numbers of their elements. Precisely, we have

$$\begin{aligned} &P(U_{m:n} \le x,\, V_{k:n} \le y) \\ &= P\left(\sum_{i=1}^{n} 1_{[0,x]}(U_i) \ge m,\, \sum_{i=1}^{n} 1_{[0,y]}(V_i) \ge k\right) \\ &= P\Bigg(\Big(\sum_{|T| \ge m} \Big\{U_i \le x, i \in T;\, U_i > x, i \in T^{\complement}\Big\}\Big) \\ &\qquad \cap \Big(\sum_{|S| \ge k} \Big\{V_i \le y, i \in S;\, V_i > y, i \in S^{\complement}\Big\}\Big)\Bigg) \\ &= \sum_{|T| \ge m} \sum_{|S| \ge k} P\Bigg(\Big\{U_i \le x, i \in T;\, U_i > x, i \in T^{\complement}\Big\} \\ &\qquad \cap \Big\{V_i \le y, i \in S;\, V_i > y, i \in S^{\complement}\Big\}\Bigg) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{|T|\geq m} \sum_{|S|\geq k} P(U_i \leq x, V_i \leq y,\, i \in T\cap S) P\left(U_i \leq x, V_i > y,\, i \in T \cap S^{\complement}\right) \\
&\qquad \times P\left(U_i > x, V_i \leq y,\, i \in T^{\complement} \cap S\right) P\left(U_i > x, V_i > y,\, i \in T^{\complement} \cap S^{\complement}\right) \\
&= \sum_{|T|\geq m} \sum_{|S|\geq k} P(U \leq x, V \leq y)^{|T\cap S|} P(U \leq x, V > y)^{\left|T\cap S^{\complement}\right|} \\
&\qquad \times P(U > x, V \leq y)^{\left|T^{\complement}\cap S\right|} P(U > x, V > y)^{\left|T^{\complement}\cap S^{\complement}\right|}.
\end{aligned}
$$

As a consequence and by writing $x = \exp(\log(x))$ for $x \geq 0$ we obtain

$$
\begin{aligned}
&P(U_{m:n} \leq x,\, V_{k:n} \leq y + \varepsilon) - P(U_{m:n} \leq x,\, V_{k:n} \leq y) \\
&= \sum_{|T|\geq m} \sum_{|S|\geq k} \Bigg\{ \exp\Bigg(|T\cap S| \log(P(U \leq x, V \leq y+\varepsilon)) \\
&\qquad + \left|T\cap S^{\complement}\right| \log(P(U \leq x, V > y+\varepsilon)) \\
&\qquad + \left|T^{\complement}\cap S\right| \log(P(U > x, V \leq y+\varepsilon)) \\
&\qquad + \left|T^{\complement}\cap S^{\complement}\right| \log(P(U > x, V > y+\varepsilon)) \Bigg) \\
&\quad - \exp\Bigg(|T\cap S| \log(P(U \leq x, V \leq y) \\
&\qquad + \left|T\cap S^{\complement}\right| \log(P(U \leq x, V > y)) \\
&\qquad + \left|T^{\complement}\cap S\right| \log(P(U > x, V \leq y)) \\
&\qquad + \left|T^{\complement}\cap S^{\complement}\right| \log(P(U > x, V > y)) \Bigg\} \\
&= \sum_{|T|\geq m} \sum_{|S|\geq k} \Bigg\{ \exp\Bigg(|T\cap S| \log\left(1 + \frac{P(U \leq x, V \leq y+\varepsilon) - P(U \leq x, V \leq y)}{P(U \leq x, V \leq y)}\right) \\
&\qquad + \left|T\cap S^{\complement}\right| \log\left(1 + \frac{P(U \leq x, V > y+\varepsilon) - P(U \leq x, V > y)}{P(U \leq x, V > y)}\right) \\
&\qquad + \left|T^{\complement}\cap S\right| \log\left(1 + \frac{P(U > x, V \leq y+\varepsilon) - P(U > x, V \leq y)}{P(U > x, V \leq y)}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\left|T^{\complement}\cap S^{\complement}\right|\log\left(1+\frac{P(U>x,V>y+\varepsilon)-P(U>x,V>y)}{P(U>x,V>y)}\right)\Bigg)\\
&\quad -1\Bigg\}\\
&\quad\times P(U\le x,V\le y)^{|T\cap S|}P(U\le x,V>y)^{\left|T\cap S^{\complement}\right|}\\
&\qquad\times P(U>x,V\le y)^{\left|T^{\complement}\cap S\right|}P(U>x,V>y)^{\left|T^{\complement}\cap S^{\complement}\right|}
\end{aligned}
$$

For $\varepsilon\downarrow 0$ we have the expansions

$$
\begin{aligned}
P(U\le x,V\le y+\varepsilon)-P(U\le x,V\le y)&=\frac{\partial}{\partial y}C(x,y)\varepsilon+o(\varepsilon),\\
P(U\le x,V> y+\varepsilon)-P(U\le x,V> y)&=-\frac{\partial}{\partial y}C(x,y)\varepsilon+o(\varepsilon),\\
P(U> x,V\le y+\varepsilon)-P(U> x,V\le y)&=\left(1-\frac{\partial}{\partial y}C(x,y)\right)\varepsilon+o(\varepsilon),\\
P(U> x,V> y+\varepsilon)-P(U> x,V> y)&=\left(\frac{\partial}{\partial y}C(x,y)-1\right)\varepsilon+o(\varepsilon).
\end{aligned}
$$

From the Taylor expansions $\log(1+x)=x+o(x)$ and $\exp(x)-1=x+o(x)$ as $x\to 0$ we, thus, obtain from the preceding equations

$$
\begin{aligned}
&\frac{P(U_{m:n}\le x,V_{k:n}\le y+\varepsilon)-P(U_{m:n}\le x,V_{k:n}\le y)}{\varepsilon}\\
&\to_{\varepsilon\downarrow 0}\sum_{|T|\ge m}\sum_{|S|\ge k}\Bigg\{|T\cap S|\,\frac{\frac{\partial}{\partial y}C(x,y)}{p_1}\\
&\qquad\qquad -\left|T\cap S^{\complement}\right|\frac{\frac{\partial}{\partial y}C(x,y)}{p_2}\\
&\qquad\qquad +\left|T^{\complement}\cap S\right|\frac{1-\frac{\partial}{\partial y}C(x,y)}{p_3}\\
&\qquad\qquad +\left|T^{\complement}\cap S^{\complement}\right|\frac{\frac{\partial}{\partial y}C(x,y)-1}{p_4}\Bigg\}\\
&\qquad\qquad \times p_1^{|T\cap S|}p_2^{\left|T\cap S^{\complement}\right|}p_3^{\left|T^{\complement}\cap S\right|}p_4^{\left|T^{\complement}\cap S^{\complement}\right|}\\
&=:f(x,y)
\end{aligned}
$$

with

$$\begin{aligned}
p_1 &:= P((U,V) \in A_1) := P(U \le x, V \le y),\\
p_2 &:= P((U,V) \in A_2) := P(U \le x, V > y),\\
p_3 &:= P((U,V) \in A_3) := P(U > x, V \le y),\\
p_4 &:= P((U,V) \in A_4) := P(U > x, V > y).
\end{aligned}$$

Note that $p_1 + p_2 + p_3 + p_4 = 1$. Set

$$n_j := \sum_{i=1}^{n} 1_{A_j}(U_i, V_i), \qquad 1 \le j \le 4.$$

Then we obtain

$$\begin{aligned}
f(x,y) = E\Bigg(\Bigg\{ &\frac{\frac{\partial}{\partial y}C(x,y)}{p_1} n_1 - \frac{\frac{\partial}{\partial y}C(x,y)}{p_2} n_2 \\
&+ \frac{1-\frac{\partial}{\partial y}C(x,y)}{p_3} n_3 + \frac{\frac{\partial}{\partial y}C(x,y) - 1}{p_4} n_4 \Bigg\} \\
&\times 1(n_1 + n_2 \ge m, n_1 + n_3 \ge k)\Bigg).
\end{aligned}$$

Put, for notational convenience, $\xi_j := (U_j, V_j)$, $1 \le j \le n$. We have

$$\begin{aligned}
&E(n_1 1(n_1 + n_2 \ge m, n_1 + n_3 \ge k)) \\
&= \sum_{j=1}^{n} P\left(\{\xi_j \in A_1\} \cap \left\{ \sum_{i=1}^{n} 1_{A_1 \cup A_2}(\xi_i) \ge m, \sum_{i=1}^{n} 1_{A_1 \cup A_3}(\xi_i) \ge k \right\} \right) \\
&= n p_1 P\left(\sum_{i=1}^{n-1} 1_{A_1 \cup A_2}(\xi_i) \ge m-1, \sum_{i=1}^{n-1} 1_{A_1 \cup A_3}(\xi_i) \ge k-1 \right) \\
&= n p_1 P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i) \ge m-1, \sum_{i=1}^{n-1} 1_{[0,y]}(V_i) \ge k-1 \right),
\end{aligned}$$

as well as

$$E(n_2 1(n_1 + n_2 \ge m, n_1 + n_3 \ge k))$$

$$= np_2 P\left(\sum_{i=1}^{n-1} 1_{A_1\cup A_2}(\xi_i) \ge m-1, \sum_{i=1}^{n-1} 1_{A_1\cup A_3}(\xi_i) \ge k\right)$$
$$= np_2 P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i) \ge m-1, \sum_{i=1}^{n-1} 1_{[0,y]}(V_i) \ge k\right),$$

$$E(n_3 1(n_1+n_2 \ge m, n_1+n_3 \ge k))$$
$$= np_3 P\left(\sum_{i=1}^{n-1} 1_{A_1\cup A_2}(\xi_i) \ge m, \sum_{i=1}^{n-1} 1_{A_1\cup A_3}(\xi_i) \ge k-1\right)$$
$$= np_3 P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i) \ge m, \sum_{i=1}^{n-1} 1_{[0,y]}(V_i) \ge k-1\right),$$

and, finally,

$$E(n_4 1(n_1+n_2 \ge m, n_1+n_3 \ge k))$$
$$= np_4 P\left(\sum_{i=1}^{n-1} 1_{A_1\cup A_2}(\xi_i) \ge m, \sum_{i=1}^{n-1} 1_{A_1\cup A_3}(\xi_i) \ge k\right)$$
$$= np_4 P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i) \ge m, \sum_{i=1}^{n-1} 1_{[0,y]}(V_i) \ge k\right).$$

Altogether we obtain from the preceding equations

$$\begin{aligned}
f(x,y) &= n\frac{\partial}{\partial y}C(x,y)P(U_{m-1:n-1} \le x, V_{k-1:n-1} \le y) \\
&\quad - n\frac{\partial}{\partial y}C(x,y)P(U_{m-1:n-1} \le x, V_{k:n-1} \le y) \\
&\quad + n\left(1 - \frac{\partial}{\partial y}C(x,y)\right)P(U_{m:n-1} \le x, V_{k-1:n-1} \le y) \\
&\quad - n\left(1 - \frac{\partial}{\partial y}C(x,y)\right)P(U_{m:n-1} \le x, V_{k:n-1} \le y) \\
&= n\frac{\partial}{\partial y}C(x,y)\Big(P(U_{m-1:n-1} \le x,\, V_{k-1:n-1} \le y) \\
&\qquad\qquad - P(U_{m-1:n-1} \le x,\, V_{k:n-1} \le y)\Big)
\end{aligned}$$

$$+ n\left(1-\frac{\partial}{\partial y}C(x,y)\right)\Big(P(U_{m:n-1}\le x,\, V_{k-1:n-1}\le y) \\ - P(U_{m:n-1}\le x,\, V_{k:n-1}\le y)\Big),$$

We, thus, have established so far

$$\begin{aligned} &P(U_{m:n}\le x \mid V_{k:n}=y) \\ &= \frac{n}{g_{k,n}(y)}\left\{\frac{\partial}{\partial y}C(x,y)P\left(U_{m-1:n-1}\le x, \sum_{i=1}^{n-1} 1_{(0,y]}(V_i)=k-1\right)\right. \\ &\qquad \left. + \left(1-\frac{\partial}{\partial y}C(x,y)\right)P\left(U_{m:n-1}\le x,\, \sum_{i=1}^{n-1} 1_{(0,y]}(V_i)=k-1\right)\right\} \\ &= \frac{n}{g_{k,n}(y)}\left\{P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i)\ge m, \sum_{i=1}^{n-1} 1_{(0,y]}(V_i)=k-1\right)\right. \\ &\qquad \left. + \frac{\partial}{\partial y}C(x,y)P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i)=m-1, \sum_{i=1}^{n-1} 1_{(0,y]}(V_i)=k-1\right)\right\}. \end{aligned}$$

From the fact that

$$P\left(\sum_{i=1}^{n-1} 1_{[0,y]}(V_i)=k-1\right)=\frac{g_{k,n}(y)}{n}$$

we, thus, obtain the representation

$$\begin{aligned} &P(U_{m:n}\le x \mid V_{k:n}=y) \\ &= P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i)\ge m \,\Big|\, \sum_{i=1}^{n-1} 1_{[0,y]}(V_i)=k-1\right) \\ &\qquad + \frac{\partial}{\partial y}C(x,y)P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i)=m-1 \,\Big|\, \sum_{i=1}^{n-1} 1_{[0,y]}(V_i)=k-1\right). \end{aligned}$$

We know from the theory of point processes (see, e.g. Reiss (1993, E.18)) that

$$P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i)\ge m \,\Big|\, \sum_{i=1}^{n-1} 1_{[0,y]}(V_i)=k-1\right)$$

$$= P\left(\sum_{i=1}^{n-1} 1_{[0,x]}(U_i) \geq m \;\middle|\; \sum_{i=1}^{n-1} 1_{[0,y]}(V_i) = k-1, \sum_{i=1}^{n-1} 1_{(y,1]}(V_i) = n-k\right)$$
$$= P\left(\sum_{i=1}^{k-1} 1_{[0,x]}\left(U_i^{(1)}\right) + \sum_{i=1}^{n-k} 1_{[0,x]}\left(U_i^{(2)}\right) \geq m\right),$$

where $U_1^{(1)}, \dots, U_{k-1}^{(1)}, U_1^{(2)}, \dots, U_{n-k}^{(2)}$ are independent rv with

$$P\left(U_i^{(1)} \leq u\right) = P(U \leq u \mid V \leq y) = \frac{C(u,y)}{y}$$

and

$$P\left(U_i^{(2)} \leq u\right) = P(U \leq u \mid V > y) = \frac{u - C(u,y)}{1-y}, \qquad 0 \leq u \leq 1,$$

which completes the proof of Lemma 2.2. □

2.3 Asymptotic Independence of Order Statistics

Throughout this section, $(U_{r:n}, V_{k:n})$ denotes a rv of componentwise os pertaining to independent copies $(U_1, V_1), \dots, (U_n, V_n)$ of a rv (U, V), which follows a copula C. Note that the dependence between the two components does not matter for the next theorem. By X, Y, η_j we denote independent rv, where X and Y are standard normal distributed and η_j has df $G_j(x) = \exp(x)\sum_{i=0}^{j-1}(-x)^i/i!$, $x \leq 0$. The following main result establishes asymptotic independence of $U_{r:n}$ and $V_{k:n}$ for various sequences $r = r(n)$, $k = k(n)$, $n \in \mathbb{N}$.

Theorem 2.3. *Let $k = k(n)$, $j = j(n) \in \{1, \dots, n\}$, $n \in \mathbb{N}$.*

(i) If k satisfies $k \to_{n\to\infty} \infty$, $k/n \to_{n\to\infty} 0$, then, for fixed $j \in \mathbb{N}$,

$$\left(\frac{n}{\sqrt{k}}\left(U_{n-k+1:n} - \frac{n-k+1}{n+1}\right), n(V_{n-j+1:n} - 1)\right) \to_D (X, \eta_j).$$

(ii) With k and j as in (i),

$$\left(\frac{n}{\sqrt{k}}\left(U_{k:n} - \frac{k}{n+1}\right), n(V_{n-j+1:n} - 1)\right) \to_D (X, \eta_j).$$

(iii) If k satisfies $k/n \to_{n\to\infty} \lambda \in (0,1)$ and $j \in \mathbb{N}$ is fixed, then

$$\left(\sqrt{n}\left(U_{k:n} - \frac{k}{n+1}\right), n(V_{n-j+1:n} - 1)\right) \to_D \left((\lambda(1-\lambda))^{1/2}X, \eta_j\right).$$

(iv) With k is chosen as in (iii) and $j \to_{n\to\infty} \infty$, $j/n \to_{n\to\infty} 0$

$$\left(\sqrt{n}\left(U_{k:n} - \frac{k}{n+1}\right), \frac{n}{\sqrt{j}}\left(V_{n-j+1:n} - \frac{n-j+1}{n+1}\right)\right)$$
$$\to_D \left((\lambda(1-\lambda))^{1/2}X, Y\right).$$

(v) With k as chosen in (i), j chosen as in (iv) and, in addition, $j/\sqrt{k} \to_{n\to\infty}$ 0,

$$\left(\frac{n}{\sqrt{k}}\left(U_{n-k+1:n} - \frac{n-k+1}{n+1}\right), \frac{n}{\sqrt{j}}\left(V_{n-j+1:n} - \frac{n-j+1}{n+1}\right)\right) \to_D (X, Y).$$

More results can immediately be deduced from the preceding result by noting that $(1-U_{r:n}, 1-V_{k:n}) = \left(\bar{U}_{n-r+1:n}, \bar{V}_{n-k+1:n}\right)$, which are os pertaining to the iid sequence $(\bar{U}_1, \bar{V}_1), \dots, (\bar{U}_n, \bar{V}_n) = (1-U_1, 1-V_1), \dots, (1-U_n, 1-V_n)$ with copula $\bar{C}(u,v) = P(1-U \le u, 1-V \le v)$.

Proof. We prove only assertion (i). The remaining parts can be shown in complete analogy. By $P * X$ we denote in what follows the distribution of a rv X, i.e., $(P*X)(B) = P(X \in B)$ for any B in the Borel-σ-field of $\mathbb{R}$. We have with $\mu_n := (n-k+1)/(n+1)$ and $x \in \mathbb{R}$, $y < 0$ by (2.3) the representation

$$P\left(\frac{n}{\sqrt{k}}(U_{n-k+1:n} - \mu_n) \le x, n(V_{n-j+1:n} - 1) \le y\right)$$
$$= \int_{-n}^{y} P\left(\frac{n}{\sqrt{k}}(U_{n-k+1:n} - \mu_n) \le x \mid n(V_{n-j+1:n} - 1) = z\right)$$
$$(P * n(V_{n-j+1:n} - 1))(dz)$$
$$= \int_{-n}^{y} P\left(U_{n-k+1:n} \le \frac{\sqrt{k}}{n}x + \mu_n \mid V_{n-j+1:n} = 1 + \frac{z}{n}\right)$$
$$(P * n(V_{n-j+1:n} - 1))(dz)$$
$$= \int_{-n}^{y} P\left(\sum_{i=1}^{n-j} 1_{\left[0, \frac{\sqrt{k}}{n}x + \mu_n\right]}\left(U_i^{(1)}\right) + \sum_{i=1}^{j-1} 1_{\left[0, \frac{\sqrt{k}}{n}x + \mu_n\right]}\left(U_i^{(2)}\right) \ge n-k+1\right)$$

$$+\frac{\partial}{\partial y}C(x,1+\frac{z}{n})$$
$$\times P\left(\sum_{i=1}^{n-j}1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(1)}\right)+\sum_{i=1}^{j-1}1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(2)}\right)=n-k\right)$$
$$(P*n(V_{n-j+1:n}-1))(dz). \tag{2.4}$$

It is well known that $n(V_{n-j+1:n}-1)\to_D G_j$, see, e.g. equation (5.1.28) in Reiss (1989).

We claim that

$$P\left(\sum_{i=1}^{n-j}1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(1)}\right)+\sum_{i=1}^{j-1}1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(2)}\right)\geq n-k+1\right)\to_{n\to\infty}\Phi(x),$$

where $\Phi(\cdot)$ denotes the df of the standard normal distribution.

Note that

$$p_n:=P\left(U_i^{(1)}\leq\frac{\sqrt{k}}{n}x+\mu_n\right)=\frac{C\left(\frac{\sqrt{k}}{n}x+\mu_n,1+\frac{z}{n}\right)}{1+\frac{z}{n}}\to_{n\to\infty}1$$

and

$$\begin{aligned}1-p_n&=\frac{1+\frac{z}{n}-C\left(\frac{\sqrt{k}}{n}x+\mu_n,1+\frac{z}{n}\right)}{1+\frac{z}{n}}\\&=\frac{1+\frac{z}{n}-\frac{\sqrt{k}}{n}x-\mu_n+\left(\frac{\sqrt{k}}{n}x+\mu_n-C\left(\frac{\sqrt{k}}{n}x+\mu_n,1+\frac{z}{n}\right)\right)}{1+\frac{z}{n}}\\&=\frac{\frac{z}{n}-\frac{\sqrt{k}}{n}x+\frac{k}{n+1}+\int_{1+z/n}^{1}\frac{\partial}{\partial v}C\left(\frac{\sqrt{k}}{n}x+\mu_n,v\right)\,dv}{1+\frac{z}{n}}\\&=\frac{-\frac{\sqrt{k}}{n}x+\frac{k}{n+1}+O\left(\frac{z}{n}\right)}{1+\frac{z}{n}}\end{aligned}$$

by Theorem 2.1. We obtain that $(n-j)p_n(1-p_n)$ is of order $k(n)$ as $n\to\infty$ and, thus, the central limit theorem for arrays of binomial distributions implies

$$\frac{\sum_{i=1}^{n-j}\left(1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(1)}\right)-p_n\right)}{\left((n-j)p_n(1-p_n)\right)^{1/2}}\to_D N(0,1).$$

As a consequence we obtain

$$
\begin{aligned}
&P\left(\sum_{i=1}^{n-j} 1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(1)}\right) + \sum_{i=1}^{j-1} 1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(2)}\right) \ge n-k+1\right)\\
&= P\left(\frac{\sum_{i=1}^{n-j}\left(1_{\left[0,\frac{\sqrt{k}}{n}x+\mu_n\right]}\left(U_i^{(1)}\right)-p_n\right)}{((n-j)p_n(1-p_n))^{1/2}} + o(1) \ge \frac{n-k+1-(n-j)p_n}{((n-j)p_n(1-p_n))^{1/2}}\right)\\
&\to_{n\to\infty} 1-\Phi(-x) = \Phi(x),
\end{aligned}
$$

since

$$
\frac{n-k+1-(n-j)p_n}{((n-j)p_n(1-p_n))^{1/2}} = \frac{n(1-p_n)-k+O(1)}{\sqrt{k}(1+o(1))} = -x+o(1).
$$

This implies that the integrand in representation (2.4) converges to $\Phi(x)$. The assertion now follows from the dominated convergence theorem. □

Chapter 3

Conditional Tail Independence

3.1 Introduction

Let $\boldsymbol{U} = (U_1, \dots, U_d)$ be a rv, whose df F is in the domain of attraction of a multivariate extreme value df G, denoted by $F \in \mathcal{D}(G)$, i.e., there are constants $\boldsymbol{a}_n = (a_{n1}, \dots, a_{nd}) > \boldsymbol{0} \in \mathbb{R}^d$, $\boldsymbol{b}_n = (b_{n1}, \dots, b_{nd}) \in \mathbb{R}^d$, $n \in \mathbb{N}$, such that for each $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$

$$F^n(\boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n) \to_{n\to\infty} G(\boldsymbol{x}).$$

The rv $\boldsymbol{U}$, or, equivalently, the df F, is said to have asymptotically independent (upper) tails, if

$$G(\boldsymbol{x}) = \prod_{i=1}^{d} G_i(x_i),$$

where G_i, $1 \le i \le d$, denote the univariate margins of G.

We require that the df F of $\boldsymbol{U}$ coincides in its upper tail with a *copula*, say C, i.e., there exists $\boldsymbol{u}_0 = (u_{01}, \dots, u_{0d}) \in (0,1)^d$ such that

$$F(\boldsymbol{u}) = C(\boldsymbol{u}), \qquad \boldsymbol{u} \in [\boldsymbol{u}_0, \boldsymbol{1}] \subset \mathbb{R}^d.$$

Each univariate margin of a copula is the uniform distribution $H(u) = u$ for $0 \le u \le 1$ and, thus, each univariate margin of F equals $H(u)$ for $u \in [v_0, 1]$, where $v_0 := \max_{1\le i\le d} u_{0i}$.

More specifically, we require in this chapter that the upper tail of C is that of an *Archimedean copula* C_φ that was introduced in Section 1.2. This means that there exists a convex and strictly decreasing function $\varphi : (0, 1] \to [0, \infty)$ with $\varphi(1) = 0$, such that

$$C_\varphi(\boldsymbol{u}) = \varphi^{-1}\left(\varphi(u_1) + \cdots + \varphi(u_d)\right)$$

for $\boldsymbol{u} \in [\boldsymbol{u}_0, \boldsymbol{1}] \subset \mathbb{R}^d$, where $\boldsymbol{u}_0 = (u_{01}, \dots, u_{0d}) \in (0, 1)^d$.

A prominent example is $\varphi_p(s) := (1 - s)^p$, $s \in [0, 1]$, where $p \geq 1$. In this case we obtain

$$C_{\varphi_p}(\boldsymbol{u}) = 1 - \left(\sum_{i=1}^d (1 - u_i)^p\right)^{1/p}, \qquad \boldsymbol{u} \in [\boldsymbol{u}_0, \boldsymbol{1}]. \tag{3.1}$$

Note that

$$C_{\varphi_p}(\boldsymbol{u}) := \max\left(0, 1 - \left(\sum_{i=1}^d (1 - u_i)^p\right)^{1/p}\right), \qquad \boldsymbol{u} \in [0, 1]^d,$$

defines a multivariate df only in dimension $d = 2$, see, e.g., McNeil and Nešlehová (2009, Examples 2.1, 2.2). But one can find for arbitrary dimension $d \geq 2$ a rv, whose df satisfies equation (3.1), see, e.g., Falk (2019, (2.15)). This is the reason, why we require the Archimedean structure of C_φ only on some upper interval $[\boldsymbol{u}_0, \boldsymbol{1}]$ and we do not speak of C_φ as a *copula*, but rather of a *distribution function*.

The behavior of $C_\varphi(\boldsymbol{u})$ for $\boldsymbol{u}$ close to $\boldsymbol{1} \in \mathbb{R}^d$ determines the upper tail behavior of the components of $\boldsymbol{U}$. Precisely, suppose that $C_\varphi \in \mathcal{D}(G)$, i.e.,

$$\left[C_\varphi\left(\boldsymbol{1} + \frac{\boldsymbol{x}}{n}\right)\right]^n \to_{n\to\infty} G(\boldsymbol{x}), \qquad \boldsymbol{x} \leq \boldsymbol{0} \in \mathbb{R}^d,$$

where the norming constants are prescribed by the univariate margins of C_φ, which is the df $H(u) = u$, $u \in [v_0, 1]$. We obviously have for arbitrary $x \leq 0$ and n large enough

$$\left[H\left(1 + \frac{x}{n}\right)\right]^n = \left(1 + \frac{x}{n}\right)^n \to \exp(x).$$

The multivariate max-stable df G, consequently, has standard negative exponential margins $G_i(x) = \exp(x)$, $x \leq 0$.

Moreover, there exists a norm $\|\cdot\|_D$ on $\mathbb{R}^d$, such that $G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_D)$ for $\boldsymbol{x} \leq \boldsymbol{0} \in \mathbb{R}^d$; see Section 1.2. This norm $\|\cdot\|_D$ describes the asymptotic tail dependence of the margins of C_φ. In particular $\|\cdot\|_D = \|\cdot\|_1$ is the case of (asymptotic) independence of the margins, whereas $\|\cdot\|_D = \|\cdot\|_\infty$ yields their total dependence. For the df C_{φ_p} in (3.1) we obtain, for example, for n large,

$$\left[C_{\varphi_p}\left(\mathbf{1}+\frac{\boldsymbol{x}}{n}\right)\right]^n = \left(1-\frac{1}{n}\left(\sum_{i=1}^d |x_i|^p\right)^{1/p}\right)^n$$
$$\to_{n\to\infty} \exp\left(-\|\boldsymbol{x}\|_p\right), \qquad \boldsymbol{x}=(x_1,\dots,x_d)\le \mathbf{0}\in\mathbb{R}^d,$$

where $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, $p \ge 1$, is the logistic norm on $\mathbb{R}^d$. In this case we have tail independence only for $p = 1$.

We will investigate the problem, if conditioning on a margin $U_j = u$ has an influence on the tail dependence of the left margins $U_1,\dots,U_{j-1},U_{j+1},\dots,U_d$. Actually, we will show that the rv $(U_1,\dots,U_{j-1},U_{j+1},\dots,U_d)$, conditional on $U_j = u$, has in general independent tails, for each choice of j, no matter what the unconditional tail behavior is; see Section 3.3. This is achieved under a mild condition on the generator function φ, which is introduced in Section 3.2.

3.2 Condition on the generator function

Our results are achieved under the following condition on the generator function φ. There exists a number $p \ge 1$ such that

$$\lim_{s\downarrow 0}\frac{\varphi(1-sx)}{\varphi(1-s)} = x^p, \qquad x>0. \tag{C0}$$

Remark 3.1. The exponent p in condition (C0) is necessarily greater than one by the convexity of φ, which can easily be seen as follows. We have for arbitrary $\lambda, x, y \in (0,1]$

$$\varphi(\lambda x + (1-\lambda)y) \le \lambda\varphi(x) + (1-\lambda)\varphi(y).$$

Setting $x = 1-s$ and $y = 1$, we obtain

$$\varphi(\lambda(1-s)+1-\lambda) = \varphi(1-\lambda s) \le \lambda\varphi(1-s)$$

and, thus,

$$\lim_{s\downarrow 0}\frac{\varphi(1-\lambda s)}{\varphi(1-s)} = \lambda^p \le \lambda.$$

But this requires $p \ge 1$.

A df C_φ, whose generator satisfies condition (C0), is in the domain of attraction of a multivariate extreme value distribution. Precisely, we have the following result.

Proposition 3.2. *Suppose that the generator φ satisfies condition* (C0)*. Then we have $C_\varphi \in \mathcal{D}(G)$, where $G(\boldsymbol{x}) = \exp\left(-\left\|\boldsymbol{x}\right\|_p\right)$, $\boldsymbol{x} \leq \mathbf{0} \in \mathbb{R}^d$.*

Proof. First we show that condition (C0) implies for $x > 0$

$$\lim_{s\downarrow 0} \frac{1-\varphi^{-1}(sx)}{1-\varphi^{-1}(s)} = x^{1/p}. \tag{3.2}$$

Choose $\delta_{sx}, \delta_s \in (0,1)$ such that

$$\varphi(1-\delta_{sx}) = sx, \quad \varphi(1-\delta_s) = s,$$

i.e.,

$$\varphi^{-1}(sx) = 1-\delta_{sx}, \quad \varphi^{-1}(s) = 1-\delta_s.$$

Condition (C0) implies for $s \downarrow 0$

$$x = \frac{\varphi(1-\delta_{sx})}{\varphi(1-\delta_s)} = \frac{\varphi\left(1-\delta_s\frac{\delta_{sx}}{\delta_s}\right)}{\varphi(1-\delta_s)} \sim \left(\frac{\delta_{sx}}{\delta_s}\right)^p,$$

where $\sim$ means that the ratio of the left hand side and the right hand side converges to one as s converges to zero. But this is

$$\lim_{s\downarrow 0} \frac{1-\varphi^{-1}(sx)}{1-\varphi^{-1}(s)} = x^{1/p}.$$

Next we show that for $\boldsymbol{x} = (x_1, \ldots, x_d) \leq \mathbf{0} \in \mathbb{R}^d$

$$\lim_{n\to\infty} C_\varphi^n\left(\mathbf{1}+\frac{\boldsymbol{x}}{n}\right) = \lim_{n\to\infty}\left[\varphi^{-1}\left(\sum_{i=1}^d \varphi\left(1+\frac{x_i}{n}\right)\right)\right]^n = \exp\left(-\left\|\boldsymbol{x}\right\|_p\right).$$

Taking logarithms on both sides, this is equivalent with

$$\lim_{n\to\infty} n\left[1-\varphi^{-1}\left(\sum_{i=1}^d \varphi\left(1+\frac{x_i}{n}\right)\right)\right] = \left\|\boldsymbol{x}\right\|_p.$$

Write

$$\frac{1}{n} = 1 - \varphi^{-1}\left(\varphi\left(1 - \frac{1}{n}\right)\right).$$

Then

$$\begin{aligned} n\left[1 - \varphi^{-1}\left(\sum_{i=1}^{d} \varphi\left(1 + \frac{x_i}{n}\right)\right)\right] &= \frac{1 - \varphi^{-1}\left(\sum_{i=1}^{d} \varphi\left(1 + \frac{x_i}{n}\right)\right)}{1 - \varphi^{-1}\left(\varphi\left(1 - \frac{1}{n}\right)\right)} \\ &= \frac{1 - \varphi^{-1}\left(\varphi\left(1 - \frac{1}{n}\right) \sum_{i=1}^{d} \frac{\varphi\left(1 + \frac{x_i}{n}\right)}{\varphi\left(1 - \frac{1}{n}\right)}\right)}{1 - \varphi^{-1}\left(\varphi\left(1 - \frac{1}{n}\right)\right)} \\ &\to_{n\to\infty} \left(\sum_{i=1}^{d} (-x_i)^p\right)^{1/p} \end{aligned}$$

by condition (C0) and equation (3.2), which is the assertion. □

Condition (C0) on φ is, for example, implied by the condition

$$\lim_{s\downarrow 0} \frac{\varphi(1-s)}{s^p} = A \tag{C1}$$

for some constant $A > 0$ and $p \geq 1$, which is obviously satisfied by the generator $\varphi_p(s) = (1-s)^p$.

Condition (C1) is by l'Hopital's rule implied by

$$-\lim_{s\downarrow 0} \frac{\varphi'(1-s)}{s^{p-1}} = pA. \tag{C2}$$

As a consequence, (C2) implies the condition

$$-\lim_{s\downarrow 0} \frac{s\varphi'(1-s)}{\varphi(1-s)} = p. \tag{C3}$$

Charpentier and Segers (2009, Theorem 4.1) showed, among others, that a copula C_φ, whose generator satisfies (C3), is in the domain of attraction of $G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_p)$, $\boldsymbol{x} \leq \boldsymbol{0} \in \mathbb{R}^d$; see also Falk (2019, Corollary 3.1.15). In this case we have tail independence only if $p = 1$.

The Clayton family, for instance, with generator $\varphi_\vartheta(t) := \left(t^{-\vartheta} - 1\right)/\vartheta$ and $\vartheta > 0$, satisfies condition (C2) with $p = 1$ and $A = 1$. As a consequence, we have independent tails for each $\vartheta > 0$.

The Frank family has the generator

$$\varphi_\vartheta(t) := -\log\left(\frac{e^{-\vartheta t}-1}{e^{-\vartheta}-1}\right), \qquad \vartheta > 0.$$

It satisfies condition (C0) with $p = 1$, i.e., we have again independent tails for each $\vartheta > 0$.

Consider, on the other hand, the generator $\varphi_\vartheta(t) := (-\log(t))^\vartheta$, $\vartheta \geq 1$, of the Gumbel-Hougaard family of Archimedean copulas. This generator satisfies condition (C0) with $p = \vartheta$ and, thus, we have tail independence only for $\vartheta = 1$.

3.3 Main Theorem

In this section we establish the conditional tail independence of the margins of C_φ, if the generator φ satisfies condition (C0). However, the following lemma does not require that condition: first we compute the conditional df of $(U_1, \dots, U_{j-1}, U_{j+1}, \dots, U_d)$, given that $U_j = u$.

Lemma 3.3. *For $j \in \{1, \dots, d\}$ and $\boldsymbol{u} = (u_1, \dots, u_{j-1}, u, u_{j+1}, \dots, u_d) \in [\boldsymbol{u}_0, \boldsymbol{1})$ we have*

$$\begin{aligned} H_{j,u}(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_d) &:= P(U_i \leq u_i,\, 1 \leq i \leq d,\, i \neq j \mid U_j = u) \\ &= \frac{\varphi'(u)}{\varphi'(C(\boldsymbol{u}))} \\ &= \frac{\varphi'(u)}{\varphi'\left(\varphi^{-1}\left(\varphi(u) + \sum_{1\leq i\leq d,\, i\neq j} \varphi(u_i)\right)\right)}, \end{aligned}$$

provided the derivative $\varphi'(v)$ exists in a neighborhood of u, that φ' is continuous at u with $\varphi'(u) \neq 0$, and that $C(\boldsymbol{u}) \neq 0$ as well.

Proof. For notational simplicity we establish the result for the choice $j = d$. We have for for $\boldsymbol{u} = (u_1, \dots, u_d) \in [\boldsymbol{u}_0, \boldsymbol{1})$

$$\begin{aligned} &P(U_i \leq u_i,\, 1 \leq i \leq d-1 \mid U_d = u_d) \\ &= \lim_{\varepsilon \downarrow 0} \frac{P(U_i \leq u_i,\, 1 \leq i \leq d-1,\, U_d \in [u_d, u_d + \varepsilon])}{P(U_d \in [u_d, u_d+\varepsilon])} \\ &= \lim_{\varepsilon \downarrow 0} \left(\frac{P(U_i \leq u_i,\, 1 \leq i \leq d-1,\, U_d \leq u_d + \varepsilon)}{\varepsilon} \right. \end{aligned}$$

$$
\begin{aligned}
&\qquad - \frac{P(U_i \le u_i,\, 1 \le i \le d-1,\, U_d \le u_d)}{\varepsilon}\Bigg) \\
&= \lim_{\varepsilon \downarrow 0} \frac{\varphi^{-1}\left(\sum_{i=1}^{d-1} \varphi(u_i) + \varphi(u_d+\varepsilon)\right) - \varphi^{-1}\left(\sum_{i=1}^{d} \varphi(u_i)\right)}{\varepsilon} \\
&= \left(\varphi^{-1}\right)' \left(\sum_{i=1}^{d} \varphi(u_i)\right) \varphi'(u_d) \\
&= \frac{\varphi'(u_d)}{\varphi'\left(\varphi^{-1}\left(\sum_{i=1}^{d} \varphi(u_i)\right)\right)} \\
&= \frac{\varphi'(u_d)}{\varphi'(C_\varphi(\boldsymbol{u}))},
\end{aligned}
$$

which is the assertion. □

Note that the univariate margins of the df $H_{j,u}$, $1 \le j \le d$, coincide in their upper tails, where they are equal to

$$
H_u(v) := \frac{\varphi'(u)}{\varphi'\Big(\varphi^{-1}\left(\varphi(u)+\varphi(v)\right)\Big)}, \qquad v_0 \le v \le 1,
$$

with $v_0 = \max_{1\le i\le d} u_{0i}$.

The upper endpoint of H_u is one. Therefore, if the df H_u is in the domain of attraction of a univariate extreme value df G, then the family of negative Weibull distributions $G_\alpha(x) := \exp\left(-|x|^\alpha\right)$, $x \le 0$, with $\alpha > 0$, is the first choice, see, e.g., the Gnedenko–de Haan Theorem Falk et al. (2011, Theorem 2.1.1) which says that this is the family with an upper endpoint. Note that $\alpha = 1$ yields the standard negative exponential distribution.

The univariate df H_u is in the domain of attraction of G_α for some $\alpha > 0$ if and only

$$
\lim_{s\downarrow 0} \frac{1 - H_u(1-sx)}{1-H_u(1-s)} = x^\alpha, \qquad x > 0,
$$

see, e.g., Galambos (1987, Theorem 2.1.2).

Lemma 3.4. *Suppose that the second derivative of φ exists in a neighborhood of $u > v_0$, and that it is continuous in u with $\varphi''(u) \ne 0 \ne \varphi'(u)$. The univariate df H_u satisfies $H_u \in \mathcal{D}(G_p)$ for some $p \ge 1$ iff φ satisfies condition* (C0).

Proof. Applying Taylor's formula twice shows that

$$\begin{aligned} 1 - H_u(1-s) &= \frac{\varphi'\Big(\varphi^{-1}\big(\varphi(u)+\varphi(1-s)\big)\Big) - \varphi'(u)}{\varphi'\Big(\varphi^{-1}\big(\varphi(u)+\varphi(1-s)\big)\Big)} \\ &\sim \frac{\varphi''(u)}{\varphi'(u)^2}\varphi(1-s) \end{aligned}$$

as $s \downarrow 0$, which is the assertion. □

The next result is this chapter's main theorem:

Theorem 3.5. *Suppose the generator φ of C_φ satisfies condition* (C0)*. Then, if $u > u_{0j}$, and φ satisfies the differentiability conditions in Lemma 3.4, we obtain for $\boldsymbol{x} = (x_1, \dots, x_d) \le \mathbf{0} \in \mathbb{R}^{d-1}$*

$$[H_{j,u}(\mathbf{1} + ca_n\boldsymbol{x})]^n \to_{n\to\infty} \exp\left(-\sum_{i=1}^{d-1}(-x_i)^p\right),$$

with $c := \big(\varphi'(u)^2/\varphi''(u)\big)^{1/p}$ and $a_n := 1 - \varphi^{-1}(1/n)$, $n \ge n_0$.

Note that the convexity of φ implies that $\varphi''(u) \ge 0$.

Remark 3.6. The preceding result shows tail independence of $H_{j,u}$, as the limiting df is the product of its margins.

Lemma 3.4 implies, moreover, that also the reverse implication in the previous result holds, i.e., if $H_{j,u}$ is in the domain of attraction of a multivariate max-stable df G with negative Weibull margins having parameter at least one, then condition (C0) is satisfied by Lemma 3.4, and G has by the preceding result identical independent margins.

Finally, by the preceding arguments, we have $H_{j,u} \in \mathcal{D}(G)$, where G has negative Weibull margins, iff just one univariate margin of $H_{j,u}$ is in the domain of attraction of a univariate extreme value distribution, and in this case G has identical and independent margins.

Proof. For notational simplicity we establish this result for $j = d$. It is sufficient to establish for $\boldsymbol{x} = (x_1, \dots, x_d) \le \mathbf{0} \in \mathbb{R}^{d-1}$

$$n\big(1 - H_{d,u}(\mathbf{1} + ca_n\boldsymbol{x})\big) \to_{n\to\infty} \sum_{i=1}^{d-1}(-x_i)^p. \tag{3.3}$$

We know from Lemma 3.3 that for $(u_1, \ldots, u_{d-1}, u) \in [\boldsymbol{u}_0, \mathbf{1}]$,

$$H_{d,u}(u_1, \ldots, u_{d-1}) = \frac{\varphi'(u)}{\varphi'\left(\varphi^{-1}\left(\varphi(u) + \sum_{i=1}^{d-1} \varphi(u_i)\right)\right)}. \tag{3.4}$$

As a consequence we obtain, with $(u_1, \ldots, u_{d-1}) = \mathbf{1} + ca_n\boldsymbol{x}$,

$$\begin{aligned} &n\big(1 - H_{d,u}\left(\mathbf{1} + ca_n\boldsymbol{x}\right)\big) \\ &= n\left(1 - \frac{\varphi'(u)}{\varphi'\left(\varphi^{-1}\left(\varphi(u) + \sum_{i=1}^{d-1} \varphi\left(1 + ca_n x_i\right)\right)\right)}\right) \\ &= n\frac{\varphi'\left(\varphi^{-1}\left(\varphi(u) + \sum_{i=1}^{d-1} \varphi\left(1 + ca_n x_i\right)\right)\right) - \varphi'(u)}{\varphi'\left(\varphi^{-1}\left(\varphi(u) + \sum_{i=1}^{d-1} \varphi\left(1 + ca_n x_i\right)\right)\right)}, \end{aligned}$$

where the denominator converges to $\varphi'(u)$ as n increases, because $a_n \downarrow 0$.

Taylor's formula yields that the nominator equals

$$\varphi''(\vartheta_n)\left(\varphi^{-1}\left(\varphi(u) + \sum_{i=1}^{d-1} \varphi\left(1 + ca_n x_i\right)\right) - u\right),$$

where $\varphi''(\vartheta_n)$ converges to $\varphi''(u)$ as n increases. Applying Taylor's formula again yields

$$\varphi^{-1}\left(\varphi(u) + \sum_{i=1}^{d-1} \varphi\left(1 + ca_n x_i\right)\right) - u = \frac{1}{\varphi'\left(\varphi^{-1}(\xi_n)\right)} \sum_{i=1}^{d-1} \varphi\left(1 + ca_n x_i\right),$$

where ξ_n converges to $\varphi(u)$ as n increases. But

$$n\sum_{i=1}^{d-1} \varphi\left(1 + ca_n x_i\right) = \sum_{i=1}^{d-1} \frac{\varphi\left(1 + ca_n x_i\right)}{\varphi(1 - a_n)} \to_{n\to\infty} \sum_{i=1}^{d-1} (-cx_i)^p.$$

by condition (C0). This yields the assertion. □

Remark 3.7. The preceding result shows tail independence of $H_{j,u}$, as the limiting df is the product of its margins.

Lemma 3.4 implies, moreover, that also the reverse implication in the previous result holds, i.e., if $H_{j,u}$ is in the domain of attraction of a multivariate max-stable df G with negative Weibull margins having parameter at least one, then condition (C0) is satisfied by Lemma 3.4, and G has by the preceding result identical independent margins.

Finally, by the preceding arguments, we have $H_{j,u} \in \mathcal{D}(G)$, where G has negative Weibull margins, iff just one univariate margin of $H_{j,u}$ is in the domain of attraction of a univariate extreme value distribution, and in this case G has identical and independent margins.

Remark 3.8. Theorem 3.5 enables the simulation of an Archimedean copula from an extreme area. In particular from equation (3.3) we have the approximation

$$\left(1 - H_{d,u}\left(\mathbf{1} + c a_n \boldsymbol{x}\right)\right) \approx \sum_{i=1}^{d-1} (-x_i/n^{1/p})^p, \quad n \to \infty.$$

A random vector of dimension $d-1$ that has the survival probability on the right hand side above, together with an independent and on $(0,1)$ uniformly distributed random variable, then provides the simulation of an Archimedean copula in its extreme region. Whether this is an *efficient* way of simulation requires, however, further work. A first and quite promising attempt was made by Kloss (2020).

3.4 Archimax Copulas

Let $\varphi : (0,1] \to [0,\infty)$ be the generator of an Archimedean copula $C_\varphi(\boldsymbol{u}) = \varphi^{-1}\left(\sum_{i=1}^d \varphi(u_i)\right)$, $\boldsymbol{u} = (u_1,\dots,u_d) \in (0,1]^d$, and let $\|\cdot\|_D$ be an arbitrary D-norm. Put

$$C(\boldsymbol{u}) := \varphi^{-1}\left(\|(\varphi(u_1),\dots,\varphi(u_d))\|_D\right), \qquad \boldsymbol{u} \in (0,1]^d. \tag{3.5}$$

It was established by Charpentier et al. (2014) that C actually defines a copula on $\mathbb{R}^d$, called *Archimax copula*. Choosing $\|\cdot\|_D = \|\cdot\|_1$ yields $C(\boldsymbol{u}) = C_\varphi(\boldsymbol{u})$ and, therefore, the concept of Archimax copulas generalizes that of Archimedean copulas considerably.

To include also the generator family $\varphi_p(s) = (1-s)^p$, $s \in [0,1]$, $p \geq 1$, we require the representation of C in equation (3.5) only for $\boldsymbol{u} \in [\boldsymbol{u}_0, \mathbf{1}] \subset (0,1]^d$. There actually exists a rv, whose copula satisfies

$$C(\boldsymbol{u}) = \varphi^{-1}\left(\|(\varphi(u_1),\dots,\varphi(u_d))\|_p\right), \qquad \boldsymbol{u} \in [\boldsymbol{u}_0, \mathbf{1}]$$

with some $\boldsymbol{u}_0 \in (0,1)^d$. This follows from the fact that $\|(|x_1|^p, \dots, |x_d|^p)\|_D^{1/p}$ is again a D-norm, with an arbitrary D-norm $\|\cdot\|_D$ and $p \ge 1$, see Proposition 2.6.1 and equations (2.14), (2.15) in Falk (2019).

An Archimax copula is in the domain of attraction of a multivariate extreme value distribution, if the generator satisfies condition (C0). Precisely, we have the following result.

Proposition 3.9. *Suppose the generator φ satisfies condition* (C0)*. Then the corresponding Archimax copula C, with arbitrary D-norm $\|\cdot\|_D$, satisfies $C \in \mathcal{D}(G)$, where $G(\boldsymbol{x}) = \exp\left(-\|(|x_1|^p, \dots, |x_d|^p)\|_D^{1/p}\right)$, $\boldsymbol{x} \le \mathbf{0} \in \mathbb{R}^d$.*

Proof. We have for $\boldsymbol{x} = (x_1, \dots, x_d) \le \mathbf{0} \in \mathbb{R}^d$

$$
\begin{aligned}
&n\left[1 - \varphi^{-1}\left(\left\|\left(\varphi\left(1 + \frac{x_1}{n}\right), \dots, \varphi\left(1 + \frac{x_d}{n}\right)\right)\right\|_D\right)\right] \\
&= \frac{1 - \varphi^{-1}\left(\varphi\left(1 - \frac{1}{n}\right)\left\|\left(\frac{\varphi\left(1+\frac{x_1}{n}\right)}{\varphi\left(1-\frac{1}{n}\right)}, \dots, \frac{\varphi\left(1+\frac{x_d}{n}\right)}{\varphi\left(1-\frac{1}{n}\right)}\right)\right\|_D\right)}{1 - \varphi^{-1}\left(\varphi\left(1 - \frac{1}{n}\right)\right)} \\
&\to_{n\to\infty} \|(|x_1|^p, \dots, |x_d|^p)\|_D^{1/p}
\end{aligned}
$$

by condition (C0) and equation (3.2). Repeating the arguments in the proof of Proposition 3.2 yields the assertion. □

Let the rv $\boldsymbol{U} = (U_1, \dots, U_d)$ follow an Archimax copula with generator function φ and D-norm $\|\cdot\|_D$. Does it also have independent tails, conditional on one of its components? We give a partial answer to this question.

Suppose the underlying $\|\cdot\|_D$ is a logistic one $\|\cdot\|_q$, with $q \ge 1$. Then

$$
\begin{aligned}
\varphi^{-1}\left(\|(\varphi(u_1), \dots, \varphi(u_d))\|_q\right) &= \varphi^{-1}\left(\left(\sum_{i=1}^d \varphi(u_i)^q\right)^{1/q}\right) \\
&= \psi^{-1}\left(\sum_{i=1}^d \psi(u_i)\right),
\end{aligned}
$$

where

$$
\psi(s) := \varphi(s)^q, \qquad s \in [0,1].
$$

If the generator φ satisfies condition (C0), then the generator ψ clearly satisfies condition (C0) as well:

$$\lim_{s\downarrow 0}\frac{\psi(1-sx)}{\psi(1-s)} = x^{pq}, \qquad x>0.$$

If φ satisfies the differentiability conditions in Lemma 3.4, then the conclusion of Theorem 3.5 applies, i.e., with the choice $\|\cdot\|_D = \|\cdot\|_q$, $q \geq 1$, the rv $\boldsymbol{U}$ has again independent tails, conditional on one of its components.

Set, on the other hand $\boldsymbol{U} = (U,\dots,U)$, where U is a rv that follows the uniform distribution on $(0,1)$. Choose $\|\cdot\|_D = \|\cdot\|_\infty$ with $\|\boldsymbol{x}\|_\infty = \max_{1\leq i\leq d}(|x_i|)$. Then we have for every function $\varphi : (0,1] \to [0,\infty)$, which is continuous and strictly decreasing,

$$\begin{aligned} C(\boldsymbol{u}) &= P(U \leq u_1,\dots,U\leq u_d)\\ &= \min_{1\leq i\leq d} u_i\\ &= \varphi^{-1}\left(\|(\varphi(u_1),\dots,\varphi(u_d))\|_\infty\right), \qquad \boldsymbol{u}\in(0,1]^d. \end{aligned}$$

The copula C is, therefore, an Archimax copula, but it has completely dependent conditional margins.

3.5 Simulation Study

We conducted a simulation study to illustrate our findings on the conditional tail independence of the Archimedean Gumbel-Hougaard copula family with dimension $d > 2$ and dependence parameter $\vartheta > 1$. The condition on ϑ implies that copula's tails are asymptotically dependent. There are several statistical tests to verify whether the tails of a multivariate distribution are asymptotically independent, provided that the latter is in the domain of attraction of a multivariate extreme value df. In the bivariate case, some tests have been suggested by Draisma et al. (2004), Hüsler and Li (2009), Chapter 6.5 in Falk et al. (2011). However, to extend them into higher dimensions than two is not straightforward. Therefore, we rely on the hypothesis testing proposed by Guillou et al. (2018), which is based on the componentwise maximum approach and is suitable for an arbitrary dimension $d \geq 2$. Such a test is based on a system of hypotheses where under the null hypothesis it is assumed that $A(\boldsymbol{t}) = 1$ for all $\boldsymbol{t} \in \mathcal{S}_d$, i.e., the tails are asymptotically independent, while under the alternative hypothesis it is assumed that $A(\boldsymbol{t}) < 1$ for at least one $\boldsymbol{t} \in \mathcal{S}_d$, i.e., some tails are asymptotically dependent. Here, A is the Pickands dependence function and $\mathcal{S}_d$ is d-dimensional unit simplex (e.g., Falk et al., 2011, Ch. 4). In Guillou et al. (2018) the authors proposed to use the test statistic $\widehat{S}_n = \sup_{\boldsymbol{t}\in\mathcal{S}_d} \sqrt{n}|\widehat{A}_n(\boldsymbol{t}) - 1|$ to decide whether or not to reject the null hypothesis, where $\widehat{A}_n$ is an appropriate estimator of the Pickands dependence function and n is the sample size of the componentwise maxima. Under the null hypothesis, the asymptotic behavior of the test statistic is known. The corresponding quantiles can be used for rejection of the null hypothesis and while there is no closed form available to compute them, they can be approximated by Monte Carlo methods. The 1% and 5% quantiles can be found in Table 1 of Guillou et al. (2018) together with some investigation into the finite-sample power for the test.

We performed the following simulation study. In the first step we simulated a sample of size $n = 110$K of independent observations from a Gumbel-Hougaard copula with $d = 3$ and $\vartheta = 3$. Then, we computed the vector of normalized componentwise maxima $m_{n,j} = \max_{i=1,\dots,n}(u_{i,j} - b_{n,j})/a_{n,j}$ with $a_{n,j} = n$, $b_{n,j} = 1$ and $j = 1,\dots,d$. In the second step, for $u = 0.99$ and $\varepsilon = 0.0005$ we selected the observations $(u_{i,1},\dots,u_{i,j-1},u_{i,j+1},\dots,u_{i,d})$ such that $u_{i,j} \in [u-\varepsilon, u+\varepsilon]$, $i = 1,\dots,n$. To work with a sample with fixed size we considered only $k = 1000$ of such observations. Then, we computed the vector of normalized componentwise maxima $m^*_{k,s} = \max_{i=1,\dots,k} u_{i,s}/(ca_{k,s})$, where $c = \left(\varphi'(u)^2/\varphi''(u)\right)^{1/\vartheta}$ and $a_{k,s} := 1 - \varphi^{-1}(1/k)$ with $\varphi(t) := (-\log(t))^{\vartheta}$ and $s = 1,\dots,j-1,j-1,\dots,d$. We repeated the first and second steps $N = 100$ times obtaining two samples of componentwise maxima, one from the d-dimensional copula and one from the corresponding $d-1$ conditional distribution. The top-left and top-right panel of Figure 3.1 display an example of maxima obtained from the Gumbel-Hougaard and the associated estimate of the Pickands dependence function, respectively. A strong dependence among the variables is evident. To see this better in the middle panels the maxima of a pair of variables and the relative estimate of the Pickands dependence function are reported. Indeed, the latter is close to lower bound $\max(1-t,t)$, i.e. the case of complete dependence. The bottom panels of Figure 3.1 display the maxima obtained with the second step of the simulation experiment and the associated estimate of the Pickands dependence function. These maxima, in contrast to the previous ones, seem to be independent and indeed the estimated Pickands dependence function is close to the upper bound (i.e. the

Table 3.1: Rejection rate (in percentage) of the null hypothesis (asymptotic independent tails) based on $M = 1000$ simulations.

Dimension		Dependence parameter				
d	ϑ :	2	3	4	5	6
3		5.414	4.877	5.438	5.352	5.725
4		5.216	5.783	5.491	4.841	4.591
5		5.353	4.396	5.791	4.685	4.454

case of independence). Then, we applied the hypothesis test with the sample of maxima obtained in the first and second step of the simulation experiment, leading to the observed values of test statistic of 3.843 and 0.348, respectively. Since the 0.95-quantiles of the distribution of S are 1.300 and 0.960 for $d = 3$ and $d = 2$, respectively (Guillou et al., 2018), we conclude that we reject the hypothesis of tails independence with the first sample of maxima whereas we do not reject it with the second sample. These results are consistent with our theoretical finding.

We repeated this simulation experiment $M = 1000$ times and with the maxima obtained with the second step of the simulation experiment we computed the rejection rate of the null hypothesis. Since we simulated data under the null hypothesis we expect that the rejection rate is close the nominal value of the first type error, i.e. 5%. We did this for different dimension d and values of the parameter ϑ. The results are collected in Table 3.1. Again the simulation results support our theoretical findings.

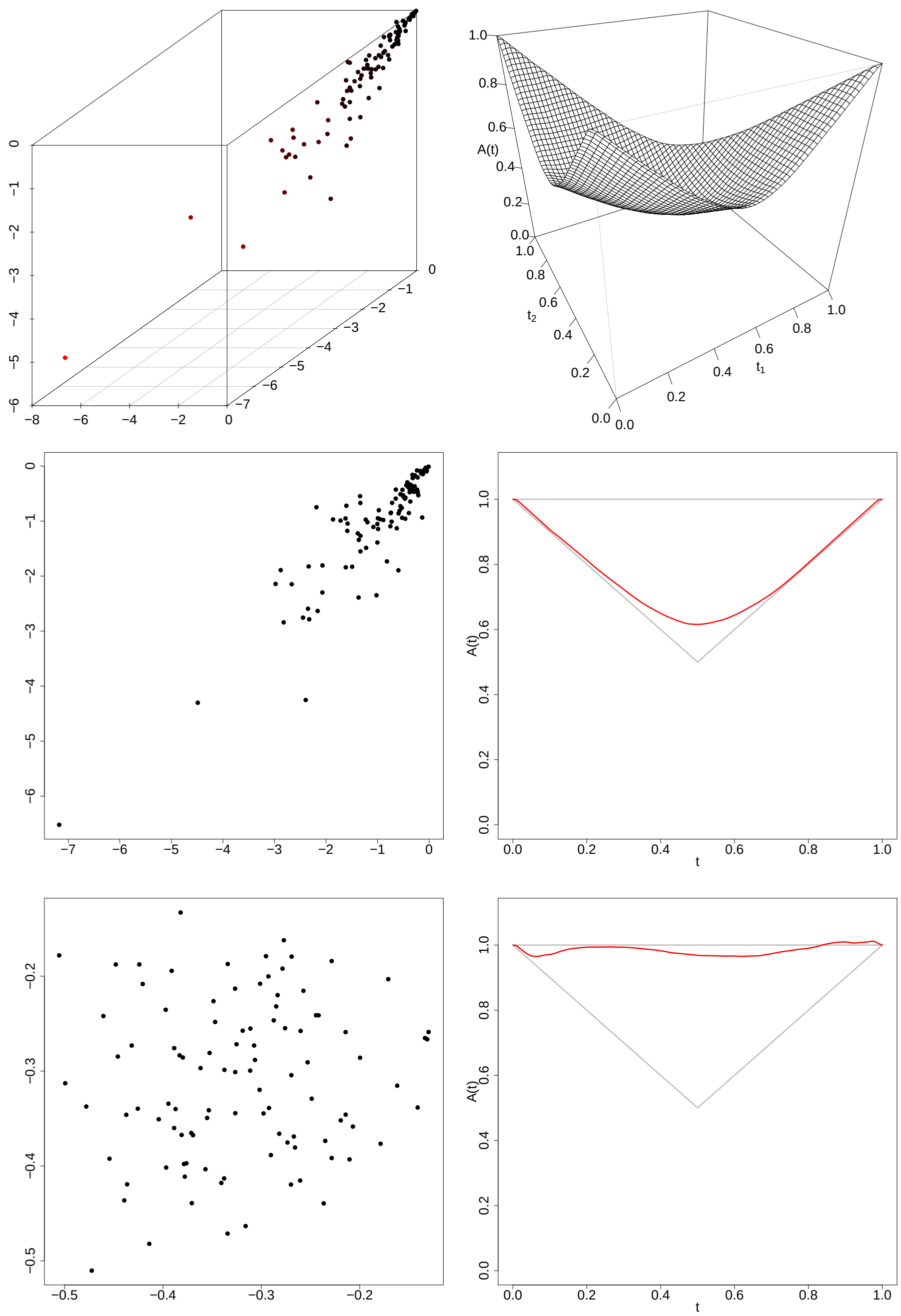

Figure 3.1: Top-left panel displays the maxima obtained with the data simulated from a trivariate Gumbel-Hougaard copula with $\vartheta = 4$. The middle one shows the maxima corresponding to two components. Finally, the one below shows the maxima obtained with the simulated data where one component is set to be a high value. The right-column reports the respective estimated Pickands dependence functions.

Chapter 4

Generalized Pareto Distributions

This chapter is organized as follows: using D-norms, we investigate the generalized Pareto copulas from Section 1.2 in more detail. The characteristic property of a GPC is its excursion or exceedance stability which is established in Theorem 4.1. The family of GPC together with the well-known set of univariate generalized Pareto distributions enables the definition of multivariate GPD in Section 4.2. As the set of univariate GPD equals the set of univariate non-degenerate exceedance stable distributions, its extension to higher dimensions via a GPC and GPD margins is an obvious idea. δ-neighborhoods of a GPC are introduced in Section 4.3. The normal copula is a prominent example. Among others we show how to simulate data, which follow a copula from such a δ-neighborhood. In Section 4.4 we show how our findings on GPC can be used to estimate exceedance probabilities above high thresholds, including confidence intervals. Finally, we conduct a case study in Section 4.5 on joint exceedance probabilities for air pollutants such as ozone, nitrogen dioxide, nitrogen oxide, sulphur dioxide and particulate matter.

4.1 Characterization of a GPC

Building on the introduction of Generalized Pareto copulas in the first chapter, we derive a characteristic property of a GPC. Suppose the rv $\boldsymbol{U}$ follows a GPC C. Then its survival function equals

$$P(\boldsymbol{U} \geq \boldsymbol{u}) = \wr\wr\, \mathbf{1} - \boldsymbol{u} \,\wr\wr_D, \qquad \boldsymbol{u} \in [\boldsymbol{u}_0, \mathbf{1}] \subset \mathbb{R}^d,$$

where

$$\wr \boldsymbol{x} \wr_D := E\left(\min_{1\le i\le d}(|x_i|\, Z_i)\right), \qquad \boldsymbol{x}\in\mathbb{R}^d,$$

is called the *dual D-norm function* pertaining to $\|\cdot\|_D$ with generator $\boldsymbol{Z} = (Z_1,\dots,Z_d)$, see the proof of Theorem 4.1. Using the equations (4.2) below it is straightforward to prove that $\wr \cdot \wr_D$ does not depend on the particular choice of the generator $\boldsymbol{Z}$ of $\|\cdot\|_D$. We have, for example,

$$\wr \boldsymbol{x} \wr_1 = 0, \quad \wr \boldsymbol{x} \wr_\infty = \min_{1\le i\le d}|x_i|, \qquad \boldsymbol{x} = (x_1,\dots,x_d)\in\mathbb{R}^d.$$

Note that the mapping $\|\cdot\|_D \mapsto \wr \cdot \wr_D$ is not one-to-one, i.e., two different D-norms can have identical dual D-norm functions.

The function $\wr \cdot \wr_D$ is obviously homogeneous of order one:

$$\wr t\boldsymbol{x} \wr_D = t \wr \boldsymbol{x} \wr_D, \qquad t\ge 0.$$

As a consequence, a GPC is *excursion stable:*

$$P\left(\boldsymbol{U}\ge \boldsymbol{1}-t\boldsymbol{u}\mid \boldsymbol{U}\ge\boldsymbol{1}-\boldsymbol{u}\right) = \frac{\wr t\boldsymbol{u} \wr_D}{\wr \boldsymbol{u} \wr_D} = t, \quad t\in[0,1],$$

for $\boldsymbol{u}$ close to $\boldsymbol{0}\in\mathbb{R}^d$, provided $\wr \boldsymbol{u} \wr_D > 0$.

Note that each marginal distribution of a GPC C is a lower dimensional GPC as well: If the rv $\boldsymbol{U} = (U_1,\dots,U_d)$ follows the GPC C on $\mathbb{R}^d$, then the rv $\boldsymbol{U}_T := (U_{i_1},\dots,U_{i_m})$ follows a GPC on $\mathbb{R}^m$, for each nonempty subset $T=\{i_1,\dots,i_m\}\subset\{1,\dots,d\}$. We have

$$P\left((U_{i_1},\dots,U_{i_m})\le \boldsymbol{v}\right) = 1-\left\|\sum_{j=1}^m (1-v_j)\boldsymbol{e}_{i_j}\right\|_D,$$

for $\boldsymbol{v}=(v_1,\dots,v_m)\in[0,1]^m$ close to $\boldsymbol{1}\in\mathbb{R}^m$. Recall that $\boldsymbol{e}_i$ denotes the i-th unit vector in $\mathbb{R}^d$, $1\le i\le d$.

The characteristic property of a GPC is its excursion stability, as formulated in the next result.

Theorem 4.1. *Let the rv $\boldsymbol{U} = (U_1,\dots,U_d)$ follow a copula C. Then C is a GPC iff for each nonempty subset $T=\{i_1,\dots,i_m\}$ of $\{1,\dots,d\}$ the rv $\boldsymbol{U}_T = (U_{i_1},\dots,U_{i_m})$ is exceedance stable, i.e.,*

$$P\left(\boldsymbol{U}_T\ge\boldsymbol{1}-t\boldsymbol{u}\right) = tP(\boldsymbol{U}_T\ge\boldsymbol{1}-\boldsymbol{u}), \qquad t\in[0,1], \tag{4.1}$$

for $\boldsymbol{u}$ close to $\boldsymbol{0}\in\mathbb{R}^m$.

Proof. The implication "⇐" in the preceding result is just a reformulation of Falk and Guillou (2008, Proposition 6). The conclusion "⇒" can be seen as follows. We can assume without loss of generality that $T = \{1, \dots, d\}$.

Using induction, it is easy to see that arbitrary numbers $a_1, \dots, a_d \in \mathbb{R}$ satisfy the equations

$$\begin{aligned}\max(a_1, \dots, a_d) &= \sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} \min_{i \in T} a_i,\\ \min(a_1, \dots, a_d) &= \sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} \max_{i \in T} a_i.\end{aligned} \tag{4.2}$$

By choosing $a_1 = \dots = a_d = 1$, the preceding equations imply in particular

$$1 = \sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1}. \tag{4.3}$$

The inclusion-exclusion principle implies for $\boldsymbol{v} \in [0,1]^d$ close to $\boldsymbol{0} \in \mathbb{R}^d$

$$\begin{aligned}P(\boldsymbol{U} \geq 1 - \boldsymbol{v}) &= 1 - P\left(\bigcup_{i=1}^{d} \{U_i \leq 1 - v_i\}\right)\\ &= 1 - \sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} P(U_i \leq 1 - v_i,\, i \in T)\\ &= 1 - \sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} \left(1 - \left\| \sum_{i \in T} v_i \boldsymbol{e}_i \right\|_D\right)\\ &= \sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} \left\| \sum_{i \in T} v_i \boldsymbol{e}_i \right\|_D.\end{aligned}$$

Choose a generator $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ of $\|\cdot\|_D$. From equation (4.2) we obtain

$$\begin{aligned}&\sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} \left\| \sum_{i \in T} v_i \boldsymbol{e}_i \right\|_D\\ &= \sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} E\left(\max_{i \in T}(v_i Z_i)\right)\end{aligned}$$

$$= E\left(\sum_{\emptyset \neq T \subset \{1,\dots,d\}} (-1)^{|T|-1} \max_{i \in T}(v_i Z_i)\right)$$
$$= E\left(\min_{1 \le i \le d}(v_i Z_i)\right) = \wr\, \boldsymbol{v} \,\wr_D.$$

Replacing $\boldsymbol{v}$ by $t\boldsymbol{u}$ yields the assertion. □

If $P(\boldsymbol{U}_T \ge \boldsymbol{1} - \boldsymbol{u}) > 0$, then (4.1) clearly becomes

$$P\left(\boldsymbol{U}_T \ge \boldsymbol{1} - t\boldsymbol{u} \mid \boldsymbol{U}_T \ge \boldsymbol{1} - \boldsymbol{u}\right) = t, \qquad t \in [0,1].$$

But $P(\boldsymbol{U}_T \ge \boldsymbol{1} - \boldsymbol{u})$ can be equal to zero for all $\boldsymbol{u}$ close to $\boldsymbol{1} \in \mathbb{R}^m$. This is for example the case, when the underlying D-norm $\|\cdot\|_D$ is $\|\cdot\|_1$. Then $\wr \cdot \wr_D = 0$, and, thus, $P(\boldsymbol{U}_T \ge \boldsymbol{1} - \boldsymbol{u}) = 0$ for all $\boldsymbol{u}$ close to $\boldsymbol{0} \in \mathbb{R}^m$, unless $m = 1$.

While the characteristic property of a GPC is its excursion stability, an extreme value copula $C_G(\boldsymbol{u}) = G\left(G_1^{-1}(u_1), \dots, G_d^{-1}(u_d)\right)$, $\boldsymbol{u} \in (0,1)^d$, which corresponds to a max-stable df G, has its max-stability as the characteristic property, which is defined below. By transforming the univariate margins to the standard negative distribution, we can assume without loss of generality that G is a SMS df. In this case we have $G_i^{-1}(u) = \log(u)$, $u \in (0,1]$, and, thus, we obtain the representation of the copula of an *arbitrary* max-stable df

$$C_G(\boldsymbol{u}) = \exp\left(-\|(\log(u_1), \dots, \log(u_d))\|_D\right), \qquad \boldsymbol{u} \in (0,1]^d, \tag{4.4}$$

with some D-norm $\|\cdot\|_D$. For a discussion of parametric families of extreme value copulas and their statistical analysis we refer to Genest and Nešlehová (2012).

Equation (4.4) obviously implies the *max-stability* of an extreme value copula C_G:

$$C_G^n\left(\boldsymbol{u}^{1/n}\right) = C_G(\boldsymbol{u}), \qquad \boldsymbol{u} \in (0,1]^d,\ n \in \mathbb{N}. \tag{4.5}$$

If, on the other hand, an *arbitrary* copula C satisfies equation (4.5), then it is clearly the copula C_G of a SMS df G. As a consequence, we have two stabilities of copulas: max-stability and exceedance stability.

Let C be an arbitrary copula on $\mathbb{R}^d$. The considerations in this section show that the copula C_{C^n} of C^n converges point-wise to a max-stable copula if, and only if, C is in its upper tail close to that of an excursion stable copula, i.e., to that of a GPC.

The message of the considerations in this section is: If one wants to model the *copula* of multivariate exceedances above high thresholds, then a GPC is a first option.

4.2 Multivariate Generalized Pareto Distributions

Let $\{G_\alpha : \alpha \in \mathbb{R}\}$ be the set of univariate max-stable df as defined by the equations above and in (1.2). The family of univariate *generalized Pareto distributions* (GPD) is the family of univariate excursion stable distributions:

$$\begin{aligned} H_\alpha(x) &:= 1 + \log(G_\alpha(x)), \qquad G_\alpha(x) > \exp(-1), \\ &= \begin{cases} 1-(-x)^\alpha, & -1 \le x \le 0, \quad \text{if } \alpha > 0, \\ 1 - x^\alpha, & x \ge 1, \quad \text{if } \alpha < 0, \\ 1 - \exp(-x), & x \ge 0, \quad \text{if } \alpha = 0. \end{cases} \end{aligned}$$

Suppose the rv V follows the df H_α. Then

$$P(V > tx \mid V > x) = t^\alpha \quad \text{for } \begin{cases} t \in [0,1], & -1 \le x < 0, \text{ if } \alpha > 0, \\ t \ge 1, & x \ge 1, \text{ if } \alpha < 0, \end{cases}$$

$$P(V > x + t \mid V > x) = \exp(-t), \quad \text{for } t \ge 0, \quad x \ge 0, \text{ if } \alpha = 0.$$

For a threshold s and an $x > s$, the univariate GPD takes the form of the following scale and shape family of distributions

$$H_{1/\xi}((x-s)/\sigma) = 1 - (1 + \xi(x-s)/\sigma)^{-1/\xi}, \tag{4.6}$$

where $\xi = 1/\alpha$ and $\sigma > 0$ (e.g. Falk et al., 2011, page 35).

The definition of a *multivariate* GPD is, however, not unique in the literature. There are different approaches (Rootzén and Tajvidi (2006), Falk et al. (2011)), each one trying to catch the excursion stability of a multivariate rv. The following suggestion might conclude this debate. Clearly, the excursion stability of a rv $\boldsymbol{X}$ should be satisfied by its margins *and* its copula. This is reflected in the following definition.

Definition 4.2. A rv $\boldsymbol{X} = (X_1, \dots, X_d)$ follows a multivariate GPD, if each component X_i follows a univariate GPD (at least in its upper tail), and if the copula C corresponding to $\boldsymbol{X}$ is a GPC, i.e., there exists a D-norm $\|\cdot\|_D$ on $\mathbb{R}^d$ and $\boldsymbol{u}_0 \in [0,1)^d$ such that

$$C(\boldsymbol{u}) = 1 - \|\boldsymbol{1} - \boldsymbol{u}\|_D, \qquad \boldsymbol{u} \in [\boldsymbol{u}_0, \boldsymbol{1}].$$

As a consequence, each such rv $\boldsymbol{X}$, which follows a multivariate GPD, is exceedance stable and vice versa.

Example 4.3. The following construction extends the bivariate approach proposed by Buishand et al. (2008) to arbitrary dimension. It provides a rv, which follows an arbitrary multivariate GPD as in Definition 4.2. Let $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ be the generator of a D-norm $\|\cdot\|_D$, with the additional property that each $Z_i \le c$, for some $c \ge 1$. Note that such a generator exists for an arbitrary D-norm according to the *normed generators theorem* for D-norms (Falk (2019)). Let the rv U be uniformly on $(0,1)$ distributed and independent of $\boldsymbol{Z}$. Put

$$\boldsymbol{V} = (V_1, \dots, V_d) := \frac{1}{U}(Z_1, \dots, Z_d) := \frac{1}{U}\boldsymbol{Z}. \tag{4.7}$$

Then, for each $i \in \{1, \dots, d\}$,

$$P\left(\frac{1}{U} Z_i \le x\right) = 1 - \frac{1}{x}, \qquad x \text{ large},$$

i.e., V_i follows in its upper tail a univariate standard Pareto distribution, and, by elementary computation, we have

$$P\left(\boldsymbol{V} \le \boldsymbol{x}\right) = 1 - \left\|\frac{\boldsymbol{1}}{\boldsymbol{x}}\right\|_D, \qquad \boldsymbol{x} \text{ large}.$$

The preceding equation implies that the copula of $\boldsymbol{V}$ is a GPC with corresponding D-norm $\|\cdot\|_D$. The rv $\boldsymbol{V}$ can be seen as a prototype of a rv, which follows a multivariate GPD. This GPD is commonly called *simple*.

Choose $\boldsymbol{V} = (V_1, \dots, V_d)$ as in equation (4.7) and numbers $\alpha_1, \dots, \alpha_d \in \mathbb{R}$. Then

$$\begin{aligned} \boldsymbol{Y} &:= (Y_1, \dots, Y_d) \\ &:= \left(H_{\alpha_1}^{-1}\left(1 - \frac{1}{V_1}\right), \dots, H_{\alpha_d}^{-1}\left(1 - \frac{1}{V_d}\right)\right) \\ &= \left(H_{\alpha_1}^{-1}\left(1 - \frac{U}{Z_1}\right), \dots, H_{\alpha_d}^{-1}\left(1 - \frac{U}{Z_d}\right)\right) \end{aligned} \tag{4.8}$$

follows a *general* multivariate GPD with margins $H_{\alpha_1}, \dots, H_{\alpha_d}$ in its univariate upper tails.

With the particular choice $\alpha_1 = \cdots = \alpha_d = 1$ we obtain a *standard* multivariate GPD

$$\boldsymbol{Y} = -U\left(\frac{1}{Z_1}, \dots, \frac{1}{Z_d}\right)$$

or

$$\boldsymbol{Y} = \left(\max\left(-\frac{U}{Z_1}, K\right), \dots, \max\left(-\frac{U}{Z_d}, K\right)\right)$$

where $K < 0$ is an arbitrary number to avoid division by zero. Its df is

$$P(\boldsymbol{Y} \le \boldsymbol{x}) = 1 - \|\boldsymbol{x}\|_D$$

for $\boldsymbol{x} \le \boldsymbol{0} \in \mathbb{R}^d$ close enough to zero.

With the particular choice $\alpha_1 = \cdots = \alpha_d = 0$, we obtain a multivariate GPD with Gumbel margins in the upper tails

$$\boldsymbol{Y} = (\log(Z_1) - \log(U), \dots, \log(Z_d) - \log(U)),$$

where $-\log(U)$ follows the standard exponential distribution on $(0, \infty)$, or, to avoid the logarithm of zero,

$$\boldsymbol{Y} = \left(\max\left(\log\left(\frac{U}{Z_1}\right), 0\right), \dots, \max\left(\log\left(\frac{U}{Z_d}\right), 0\right)\right).$$

Up to a possible location and scale shift, *each* rv $\boldsymbol{X} = (X_1, \dots, X_d)$, which follows a multivariate GPD as defined in Definition 4.2, can in its upper tail be modelled by the rv $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ in equation (4.8). This makes such rv $\boldsymbol{Y}$ in particular natural candidates for simulations of multivariate exceedances above high thresholds.

4.3 δ-Neighborhoods of GPC

A major problem with the construction in (4.7) is the additional boundedness condition on the generator $\boldsymbol{Z}$. This is, for example, not given in case of the logistic D-norm $\|\cdot\|_p$ with $p \in (1, \infty)$ or the Hüsler-Reiss D-norm. From the normed generators theorem in Falk (2019) we know that bounded generators exist, but, to the best of our knowledge, they are unknown in both cases.

In this section we drop this boundedness condition and show that the construction (4.7) provides a copula, which is in a particular neighborhood of a GPC, called δ-neighborhood. We are going to define this neighborhood next.

Denote by $R := \left\{\boldsymbol{t} \in [0,1]^d : \|\boldsymbol{t}\|_1 = 1\right\}$ the unit sphere in $[0,\infty)^d$ with respect to the norm $\|\boldsymbol{x}\|_1 = \sum_{i=1}^d |x_i| = 1$, $\boldsymbol{x} \in \mathbb{R}^d$. Choose an arbitrary copula C on $\mathbb{R}^d$ and put for $\boldsymbol{t} \in R$

$$C_{\boldsymbol{t}} := C(\mathbf{1} + s\boldsymbol{t}), \qquad s \le 0.$$

Then $C_{\boldsymbol{t}}$ is a univariate df on $(-\infty, 0]$, and the copula C is obviously determined by the family

$$\mathcal{P}(C) := \{C_{\boldsymbol{t}} : \boldsymbol{t} \in R\}$$

of univariate *spectral* df $C_{\boldsymbol{t}}$. The family $\mathcal{P}(C)$ is the *spectral decomposition* of C; cf Falk et al. (2011, Section 5.4). A copula C is, consequently, in $\mathcal{D}(G)$ iff its spectral decomposition satisfies

$$C_{\boldsymbol{t}}(s) = 1 + s\,\|\boldsymbol{t}\|_D + o(s), \qquad \boldsymbol{t} \in R,$$

as $s \uparrow 0$. The copula C is by definition in the δ*-neighborhood* of the GPC C_D with D-norm $\|\cdot\|_D$ if their upper tails are close to one another, precisely, if

$$\begin{aligned} 1 - C_{\boldsymbol{t}}(s) &= (1 - C_{D,\boldsymbol{t}}(s))\left(1 + O\left(|s|^\delta\right)\right) \\ &= |s|\,\|\boldsymbol{t}\|_D\left(1 + O\left(|s|^\delta\right)\right) \end{aligned} \tag{4.9}$$

as $s \uparrow 0$, uniformly for $\boldsymbol{t} \in R$. In this case we know from Falk et al. (2011, Theorem 5.5.5) that

$$\sup_{\boldsymbol{x}\in(-\infty,0]^d}\left|C^n\left(\mathbf{1} + \frac{1}{n}\boldsymbol{x}\right) - \exp(-\|\boldsymbol{x}\|_D)\right| = O\left(n^{-\delta}\right). \tag{4.10}$$

Under additional differentiability conditions on $C_{\boldsymbol{t}}(s)$ with respect to s, also the reverse implication (4.10) $\implies$ (4.9) holds; cf. Falk et al. (2011, Theorem 5.5.5). Therefore, the δ-neighborhood of a GPC, roughly, collects those copula with a polynomial rate of convergence for maxima.

Condition (4.9) can also be formulated in the following way:

$$\begin{aligned} 1 - C(\boldsymbol{u}) &= (1 - C_D(\boldsymbol{u}))\left(1 + O\left(\|\mathbf{1} - \boldsymbol{u}\|^\delta\right)\right) \\ &= \|\mathbf{1} - \boldsymbol{u}\|_D\left(1 + O\left(\|\mathbf{1} - \boldsymbol{u}\|^\delta\right)\right) \end{aligned}$$

as $\boldsymbol{u} \to \mathbf{1} \in \mathbb{R}^d$, uniformly for $\boldsymbol{u} \in [0,1]^d$, where $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^d$.

Example 4.4. Choose $\boldsymbol{u} \in (0,1)^d$ and put for $t \in [0,1]$

$$\mathrm{FI}\,(t,\boldsymbol{u}) := E\left(\sum_{i=1}^d 1_{(U_i > 1-tu_i)} \,\Big|\, \sum_{i=1}^d 1_{(U_i>1-u_i)} > 0\right).$$

With $t = 1$, this is the *fragility index*, introduced by Geluk et al. (2007) to measure the stability of the stochastic system $U_1,\dots,U_d$. The system is called *stable* if $\mathrm{FI}\,(1,\boldsymbol{u})$ is close to one, otherwise it is called *fragile*. The asymptotic distribution of $N_{\boldsymbol{u}} = \sum_{i=1}^d 1_{(U_i>1-tu_i)}$, given $N_{\boldsymbol{u}} > 0$, was investigated in Falk and Tichy (2011, 2012).

If $\boldsymbol{U}$ follows a GPC with corresponding D-norm $\|\cdot\|_D$, we obtain for $\boldsymbol{u}$ close enough to zero

$$\begin{aligned}\mathrm{FI}\,(t,\boldsymbol{u}) &= \sum_{i=1}^d \frac{P(U_i > 1-tu_i)}{P\left(\sum_{j=1}^d 1_{(U_j>1-u_j)} > 0\right)}\\ &= \sum_{i=1}^d \frac{tu_i}{1-P(\boldsymbol{U} \le 1-\boldsymbol{u})}\\ &= t\frac{\|\boldsymbol{u}\|_1}{\|\boldsymbol{u}\|_D}.\end{aligned}$$

Writing

$$\frac{\|\boldsymbol{u}\|_1}{\|\boldsymbol{u}\|_D} = \frac{1}{\left\|\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_1}\right\|_D}$$

implies that there is a least favourable direction $\boldsymbol{r}_0 \in R$ with

$$\|\boldsymbol{r}_0\|_D = \min_{\boldsymbol{r}\in R} \|\boldsymbol{r}\|_D\,.$$

A vector $\boldsymbol{u}$ with $\boldsymbol{u} = s\boldsymbol{r}_0$, $s > 0$, maximizes the fragility index. For arbitrary $d \ge 2$ and $\|\cdot\|_D = \|\cdot\|_p$, $p \in (1,\infty)$, one obtains for example $\boldsymbol{r}_0$ with constant entry $1/d$ and

$$\mathrm{FI}\,(t,\boldsymbol{u}) = t\frac{d}{d^{1/p}}.$$

If $\boldsymbol{U}$ follows a copula, which is in a δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$, then we obtain the representation

$$\mathrm{FI}\,(t,\boldsymbol{u}) = t\frac{\|\boldsymbol{u}_1\|}{\|\boldsymbol{u}\|_D}\left(1 + O\left(\|\boldsymbol{u}\|^\delta\right)\right), \qquad \text{for } \boldsymbol{u} \to \boldsymbol{0} \in \mathbb{R}^d.$$

If we replace $\boldsymbol{U}$ for example by $\boldsymbol{X} = (F^{-1}(U_1), \dots, F^{-1}(U_d))$, where $F(x) = 1 - 1/x$, $x \geq 1$, is the standard Pareto df, then we obtain for the fragility index

$$\mathrm{FI}\,(t, \boldsymbol{x}) = E\left(\sum_{i=1}^{d} 1_{(X_i > t x_i)} \,\middle|\, \sum_{i=1}^{d} 1_{(X_i > x_i)} > 0\right), \qquad \boldsymbol{x} \geq \mathbf{1} \in \mathbb{R}^d,\ t \geq 1,$$

the equality

$$\mathrm{FI}\,(t, \boldsymbol{x}) = \frac{1}{t} \frac{\|\mathbf{1}/\boldsymbol{x}\|_1}{\|\mathbf{1}/\boldsymbol{x}\|_D} \left(1 + O\left(\|\mathbf{1}/\boldsymbol{x}\|^{\delta}\right)\right) \qquad \text{for } x_i \to \infty,\ 1 \leq i \leq d.$$

Let $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ be a generator of the D-norm $\|\cdot\|_D$ and let U be a rv, which is independent of $\boldsymbol{Z}$ and which follows the uniform distribution on $(0, 1)$. If $\boldsymbol{Z}$ is bounded, then the copula of $\boldsymbol{Z}/U$ is a GPC C_D as established in Section 4.2. If we drop the boundedness of $\boldsymbol{Z}$ and require that $E(Z_i^2) < \infty$, then, roughly, the copula of $\boldsymbol{Z}/U$ is in a δ-neighborhood of C_D with $\delta = 1$. This is the content of our next result.

Theorem 4.5. *Let $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ generate the D-norm $\|\cdot\|_D$. Suppose that $E(Z_i^2) < \infty$ and that the df of Z_i is continuous, $1 \leq i \leq d$. Then the copula $C_{\boldsymbol{V}}$ of*

$$\boldsymbol{V} := \frac{1}{U}\boldsymbol{Z} = \frac{1}{U}(Z_1, \dots, Z_d)$$

is in the δ-neighborhood of the GPC C_D with $\delta = 1$.

Proof. The df F_i of Z_i/U satisfies for large x

$$\begin{aligned}
F_i(x) &= P(Z_i/x \leq U) \\
&= \int_0^x P(U \geq z/x)\,(P * Z_i)(dz) \\
&= \int_0^x 1 - \frac{z}{x}\,(P * Z_i)(dz) \\
&= P(Z_i \leq x) - \frac{1}{x} E\left(Z_i 1_{(Z_i \leq x)}\right) \\
&= 1 - P(Z_i > x) - \frac{1}{x}\left(1 - E\left(Z_i 1_{(Z_i > x)}\right)\right) \\
&= 1 - \frac{1}{x} - \left(P(Z_i > x) - \frac{1}{x} E\left(Z_i 1_{(Z_i > x)}\right)\right)
\end{aligned}$$

$$= \left(1 - \frac{1}{x}\right)\left(1 - \frac{P(Z_i > x) - \frac{1}{x}E\left(Z_i 1_{(Z_i > x)}\right)}{1 - \frac{1}{x}}\right),$$

where by Markov's inequality

$$P(Z_i > x) \le \frac{1}{x^2} E(Z_i^2)$$

and, using also Hölder's inequality

$$E\left(Z_i 1_{(Z_i > x)}\right) \le E(Z_i^2)^{1/2} P(Z_i > x)^{1/2} \le E(Z_i^2)^{1/2} \frac{E(Z_i^2)^{1/2}}{x} = \frac{1}{x} E(Z_i^2).$$

As a consequence we obtain

$$F_i(x) = \left(1 - \frac{1}{x}\right)\left(1 + O\left(\frac{1}{x^2}\right)\right) \qquad \text{as } x \to \infty$$

and, thus,

$$1 - F_i(x) = \frac{1}{x}\left(1 + O\left(\frac{1}{x}\right)\right) \qquad \text{as } x \to \infty.$$

Therefore, the df F_i of Z_i/U is in the δ-neighborhood of the standard Pareto distribution with $\delta = 1$.

From Falk et al. (2011, Proposition 2.2.1) we obtain as a consequence

$$F_i^{-1}(1 - q) = \frac{1}{q}(1 + O(q))$$

for $q \in (0, 1)$ as $q \to 0$.

Note that each df F_i is continuous, $1 \le i \le d$. Choose $\boldsymbol{t} = (t_1, \ldots, t_d) \in R$. We have for $s < 0$ close enough to zero

$$\begin{aligned}
&C_{\boldsymbol{t}}(s)\\
&= P(F_i(Z_i/U) \le 1 + st_i,\ 1 \le i \le d)\\
&= P(Z_i/U \le F_i^{-1}(1 + st_i),\ 1 \le i \le d)\\
&= P\left(\frac{Z_i}{U} \le \frac{1}{|s|\, t_i}(1 + O(s)),\ 1 \le i \le d\right)\\
&= P(U \ge |s|\, t_i(1 + O(s))Z_i,\ 1 \le i \le d)\\
&= P\left(U \ge |s| \max_{1 \le i \le d}(t_i(1 + O(s))Z_i)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\left\{\max_{1\le i\le d}(t_i(1+O(s))z_i)\le 1/|s|\right\}} P\left(U \ge |s| \max_{1\le i\le d}(t_i(1+O(s))z_i)\right) (P*\boldsymbol{Z})(d\boldsymbol{z})\\
&= \int_{\left\{\max_{1\le i\le d}(t_i(1+O(s))z_i)\le 1/|s|\right\}} 1-|s| \max_{1\le i\le d}(t_i(1+O(s))z_i)\,(P*\boldsymbol{Z})(d\boldsymbol{z})\\
&= P\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i) \le \frac{1}{|s|}\right)\\
&\qquad - |s|\, E\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)\, 1_{\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)\le \frac{1}{|s|}\right)}\right)\\
&= 1 - P\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i) > \frac{1}{|s|}\right) - |s|\, E\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)\right)\\
&\qquad + |s|\, E\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)\, 1_{\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)> \frac{1}{|s|}\right)}\right).
\end{aligned}
$$

We have

$$
E\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)\right) = E\left(\max_{1\le i\le d}(t_iZ_i)\right)(1+O(s)) = \|\boldsymbol{t}\|_D\,(1+O(s))
$$

and, thus, applying Markov's inequality and Hölder's inequality again,

$$
\begin{aligned}
&1 - C_{\boldsymbol{t}}(s)\\
&= P\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i) > \frac{1}{|s|}\right) + |s|\,\|\boldsymbol{t}\|_D\,(1+O(s))\\
&\qquad - |s|\, E\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)\, 1_{\left(\max_{1\le i\le d}(t_i(1+O(s))Z_i)> \frac{1}{|s|}\right)}\right)\\
&= |s|\,\|\boldsymbol{t}\|_D\,(1+O(s))\\
&= (1 - C_{D,\boldsymbol{t}}(s))\,(1+O(s))
\end{aligned}
$$

as $s \uparrow 0$, uniformly for $\boldsymbol{t} \in R$. Note that there exist constants $K_1, K_2 > 0$ such that $K_1 \le \|\boldsymbol{t}\|_D \le K_2$ for each $\boldsymbol{t} \in R$. This completes the proof of Theorem 4.5. □

An obvious example is the generator of a Hüsler-Reiss D-norm

$$
\boldsymbol{Z}^{(1)} = \left(\exp\left(X_1 - \frac{\sigma_{11}}{2}\right), \dots, \exp\left(X_d - \frac{\sigma_{dd}}{2}\right)\right),
$$

where $\boldsymbol{X} = (X_1, \dots, X_d)$ is multivariate normal $N(\boldsymbol{0}, \Sigma)$, $\Sigma = (\sigma_{ij})_{1\le i,j\le d}$.

Another example is the generator of the logistic norm $\|\cdot\|_p$, $p \in (2, \infty)$,

$$\boldsymbol{Z}^{(2)} = (Y_1, \dots, Y_d)/\Gamma(1 - 1/p),$$

where $Y_1, \dots, Y_p$ are iid Fréchet distributed with df $F(x) = \exp(x^{-p})$, $x > 0$, with parameter $p > 2$.

Both generators are unbounded, but they have square integrable components with continuous df. It is known that *bounded* generators actually exist in both cases, but to the best of our knowledge, they are unknown.

Aulbach et al. (2018) propose and extensively discuss a χ^2-goodness-of-fit test for testing, whether the underlying copula of iid rv in arbitrary dimension is in the δ-neighborhood of a GPC with an arbitrary $\delta > 0$. This test might also used to test for a GPC.

4.4 Estimation of Exceedance Probability

In this section we apply the preceding results to derive estimates of the probability that a rv $\boldsymbol{U} = (U_1, \dots, U_d)$, which follows a copula, realizes in an interval $[\boldsymbol{x}_0, \boldsymbol{1}] \subset [0, 1]^d$, where $\boldsymbol{x}_0$ is close to $\boldsymbol{1} \in \mathbb{R}^d$ and, thus, there are typically no observations available to estimate this probability by its empirical counterpart. This is a typical applied problem in extreme value analysis.

Suppose that the copula of $\boldsymbol{U}$, say C, is in the domain of attraction of a max-stable df. In this case, its upper tail is by Corollary 1.3 close to that of a GPC.

We assume that the copula C is a GPC (or very close to one in its upper tail). Being a GPC is by Theorem 4.1 characterized by the equation

$$P(\boldsymbol{U} \geq \boldsymbol{1} - t\boldsymbol{u}) = tP(\boldsymbol{U} \geq \boldsymbol{1} - \boldsymbol{u}), \tag{4.11}$$

$t \in [0, 1]$, for $\boldsymbol{u} \geq \boldsymbol{0} \in \mathbb{R}^d$ close enough to zero.

We want to estimate

$$q := P(\boldsymbol{U} \geq \boldsymbol{x}_0)$$

for some $\boldsymbol{x}_0$ close to one, based on independent copies $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$ of $\boldsymbol{U}$. Even more, we want to derive confidence interval pertaining to our estimators of q.

Choose $\boldsymbol{u}_0$ close to zero, such that equation (4.11) is satisfied for each $t \in [0, 1]$, and put

$$\boldsymbol{x}_0 = 1 - t_0\boldsymbol{u}_0$$

with some $t_0 \in (0,1)$. Then the unknown probability q satisfies the equation

$$q = P(\boldsymbol{U} \geq \mathbf{1} - t_0\boldsymbol{u}_0) = t_0 P(\boldsymbol{U} \geq \mathbf{1} - \boldsymbol{u}_0) =: t_0 p. \tag{4.12}$$

The threshold $\mathbf{1} - \boldsymbol{u}_0$ should be much smaller than the initial threshold $\boldsymbol{x}_0 = \mathbf{1} - t_0\boldsymbol{u}_0$, in which case the the unknown probability p can be estimated from the data by

$$\hat{p}_n := \frac{1}{n}\sum_{i=1}^{n} 1\left(\boldsymbol{U}^{(i)} \geq \mathbf{1} - \boldsymbol{u}_0\right).$$

Note that $n\hat{p}_n$ is binomially distributed $B(n,p)$; a confidence interval for p can be obtained by Clopper-Pearson, for example. A popular approach is due to Agresti and Coull (1998); see also Brown et al. (2001).

A confidence interval for p, say $I = (a,b)$, can by equation (4.12) be turned into a confidence interval I^* for q (with the same confidence level) by putting

$$I^* := t_0 I = (t_0 a, t_0 b).$$

4.4.1 Determination of u_0

It is clear that one would like to choose $\boldsymbol{u}_0$ as large as possible, so that one has more observations in $[\mathbf{1} - \boldsymbol{u}_0, \mathbf{1}]$. But, on the other hand, the GPC property equation (4.11) needs to be satisfied as well. In what follows we describe a proper way how to choose $\boldsymbol{u}_0$.

A possible solution to check whether our condition (4.11) is satisfied for $\boldsymbol{u}_0 = (u_{01}, \dots, u_{0d})$ is as follows: if the condition is satisfied, then we obtain for the conditional distribution

$$P(\boldsymbol{U} \geq \mathbf{1} - t\boldsymbol{u}_0 \mid \boldsymbol{U} \geq \mathbf{1} - \boldsymbol{u}_0) = t, \qquad t \in [0,1],$$

or

$$P\left(\max_{1\leq j\leq d}\left(\frac{1-U_j}{u_{0j}}\right) \leq t \,\Big|\, \max_{1\leq j\leq d}\left(\frac{1-U_j}{u_{0j}}\right) \leq 1\right) = t, \qquad t \in [0,1].$$

This means that those observations in the data $\max_{1\leq j\leq d}\left(\left(1 - U_j^{(i)}\right)/u_{0j}\right)$, $1 \leq i \leq n$, which are not greater than one, actually follow the uniform distribution on $(0,1)$, no matter what the underlying D-norm is. We denote these by $M_1, \dots, M_m$, where their number m is a random variable:

$$m = \sum_{i=1}^{n} 1\left(\max_{1\leq j\leq d}\left(\left(1 - U_j^{(i)}\right)/u_{0j}\right) \leq 1\right).$$

It is easy to check, if $M_1, \dots, M_m$ are independent and on $(0,1)$ uniformly distributed random variables, conditional on m. Standard goodness-of-fit tests like the Kolmogorov–Smirnov test or the Cramér–von Mises test can be applied. Alternatively, $M_1, \dots, M_m$ can be transformed to independent standard normal random variables by considering $\Phi^{-1}(M_i)$, and standard tests for normality such as the Shapiro-Wilk test can be applied. The preceding problem was already discussed in Falk et al. (2011, Section 5.8).

Put for $t \in [0,1]$

$$\boldsymbol{u}(t) := \frac{\boldsymbol{1} - \boldsymbol{x}_0}{t}.$$

Then, clearly,

$$\boldsymbol{x}_0 = (x_{01}, \dots, x_{0d}) = \boldsymbol{1} - t\boldsymbol{u}(t), \qquad t \in [0,1].$$

But as $\boldsymbol{u}(t)$ needs to be in $[0,1]^d$, we obtain the restriction for $i \in \{1, \dots, d\}$

$$0 \le \frac{1 - x_{0j}}{t} \le 1,$$

or

$$1 - x_{0j} \le t \le 1,$$

i.e.,

$$t_{\text{low}} := \max_{1 \le j \le d} (1 - x_{0j}) \le t \le 1.$$

Choosing $\boldsymbol{u}_0$ as large as possible now becomes choosing $t \ge t_{\text{low}}$ as small as possible.

Put for $t \in [t_{\text{low}}, 1]$

$$\hat{q}_n(t) := t\hat{p}_n(t) := \frac{t}{n} \sum_{i=1}^{n} 1\left(\boldsymbol{U}^{(i)} \ge \boldsymbol{1} - \boldsymbol{u}(t)\right). \tag{4.13}$$

For each $t \in [t_{\text{low}}, 1]$ we obtain observations $M_1(t), \dots, M_{m(t)}(t)$ in the data $\max_{1 \le j \le d}\left(\left(1 - U_j^{(i)}\right)/u_j(t)\right)$, $1 \le i \le n$, which are not greater than one. We check for each t, whether $M_1(t), \dots, M_{m(t)}(t)$ follow the uniform distribution on $(0,1)$ by plotting corresponding p-value functions:

$$(t, p_1(t)), \quad (t, p_2(t)), \qquad t \in [t_l, 1].$$

Precisely, we plot the minimum of $p_1(t)$ and $p_2(t)$ that are obtained from the Kolmogorov–Smirnov test and the Cramer-Von Mises test.

A candidate for t_0 is the lowest possible value that leads to a minimum p-value of at least 50%. This is done in Figure 4.1. Note that the preceding approach for determining a proper threshold t should be viewed as a sensitivity analysis; it is not meant as a proper test, where a p-value less than 5% would usually lead to a rejection of the null hypothesis that the sample follows a uniform distribution.

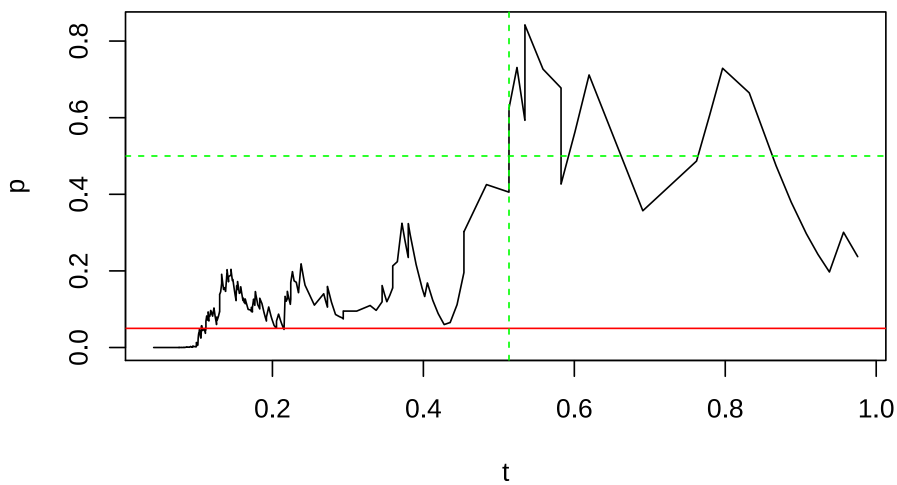

Figure 4.1: Plot of p-value of the test for uniform distribution on $(0, 1)$ depending on the factor t. The red line marks the 5% rejection level, the horizontal dashed green line is at the 50% level and the vertical one marks our selected value t_0.

4.4.2 Confidence Interval

Now that we have chosen $\boldsymbol{u}_0$, we can estimate $p = P(\boldsymbol{U} \geq \boldsymbol{u}_0)$ as described before by

$$\hat{p}_n := \frac{1}{n} \sum_{i=1}^{n} 1\left(\boldsymbol{U}^{(i)} \geq \mathbf{1} - \boldsymbol{u}_0\right).$$

Under our model assumptions, the random variable $n\hat{p}_n$ is binomial distributed $B(n, p)$ and a confidence interval for p can be obtained, for example,

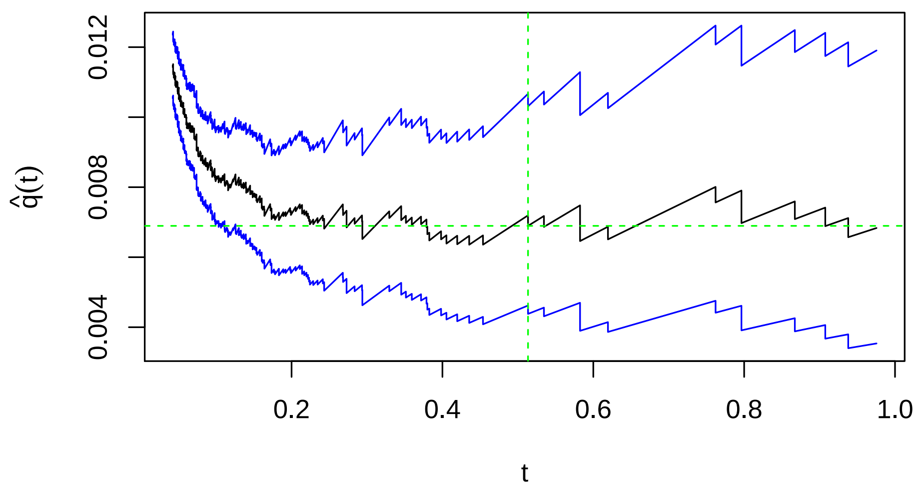

Figure 4.2: Plot of the function $\hat{q}_n(t) := t\hat{p}_n(t)$, i.e, of the estimated exceedance probability depending on the factor t. The blue lines are the corresponding confidence limits. The dashed green lines mark our selected value of t_0 and the corresponding estimated exceedance probability $\hat{q} = \hat{q}_n(t_0)$.

by Clopper–Pearson or the Agresti and Coull (1998) approach.

Figure 4.2 shows $\hat{q}_n(t)$ together with the upper and lower limits of the corresponding confidence interval at the 95% level.

4.4.3 Asymmetric Case

While in theory it does not matter whether all components of $\boldsymbol{x}_0$ are the same, there is a big advantage in that case: it permits the choice of $\boldsymbol{u}(t) = \mathbf{1}$ for the smallest value of t. That value most likely is not the correct one, meaning that our model assumptions do not hold for that particular value of t. But it leads to $m = n$ in the sample $M_1, \ldots, M_m$ for the value of $t = t_{\text{low}}$. More to the point, it guarantees $m > 0$ for at least some permissible values of t. This is important as we need the sample of M_i to be able to even check our model assumptions and determine our final value of t.

However, if $\boldsymbol{x}_0$ is too asymmetric in the sense that $\max_{1 \le j \le d} x_{0j}$ differs too much from $\min_{1\le j\le d} x_{0j}$, then m might be zero for all possible choices of t. In that case, zero is still the obvious point estimate for our probability q, but we cannot derive a confidence interval as above because there is no data in the M-sample we could check for uniform distribution to verify our model and pick t.

What we can do now is consider the symmetrized version with $\tilde{\boldsymbol{x}}_0 := (\min_{1\le j\le d} x_{0j})\mathbf{1}$. As explained, in this case there will be non-empty samples $M(t)_1, \dots, M(t)_{m(t)}$ and we can check them for their uniform distribution to support our model assumption of a GPC. If we find a suitable value t_0 following our procedure, we can be confident that the GPC model holds for $\boldsymbol{U} \ge \mathbf{1} - \tilde{\boldsymbol{u}}(t_0) = \mathbf{1} - (\mathbf{1} - \tilde{\boldsymbol{x}}_0)/t_0$. But considering the definition of $\tilde{\boldsymbol{x}}_0$, the GPC model holds for $\boldsymbol{U} \ge \mathbf{1} - \boldsymbol{u}(t_0) = \mathbf{1} - (\mathbf{1} - \boldsymbol{x}_0)/t_0$ with the same t_0 because $\tilde{\boldsymbol{x}}_0 \le \boldsymbol{x}_0$ and, therefore, $\mathbf{1} - \boldsymbol{u}(t_0) \ge \mathbf{1} - \tilde{\boldsymbol{u}}(t_0)$.

So the detour through the symmetrized version can justify the use of $\hat{p}_n = 0$ in the asymmetric case (recall that we only do this if $m = 0$ for all possible t). Using the same value of t_0, we arrive at $\hat{q}_n = 0$ and a corresponding confidence interval, stemming from a Binomial distribution with zero successes out of the n trials.

4.4.4 Simulation Study

We illustrate the performance of our nonparametric way to estimate tail probabilities through a simulation study.

First experiment: Initially, we show that our proposal provides accurate estimates of the joint tail probabilities and that there is not a sparsity issue as the dimension of the rv $\boldsymbol{U}$ increases. We consider the Gumbel-Hougaard family of copulas with dependence parameter $1 \le \tau < \infty$ that was introduced in Section 3.2 and which we already used for the simulation in Section 3.5. Recall that $\tau = 1$ represents independence, for $\tau > 1$ there is dependence and the dependence increases for increasing value of τ. We simulated a sample of $n = 1500$ and $n = 2000$ observations from a Gumbel-Hougaard copula with dimension d from 2 to 5 and a certain dependence level. Then, for a given large threshold we apply our estimation method to infer the joint tail (or exceedance) probability. We repeat this task $m = 1000$ times and we compute a Monte Carlo approximation of the bias term and the standard deviation of our proposed estimator. We repeat this experiment for different dependence levels and thresholds. Table 4.1 collects the results. The first column reports the value of the extremal coefficient $1 \le \theta \le d$, which summarizes the strength

of the tail dependence. The upper and lower bounds represent the cases of asymptotic independence and complete dependence, respectively. In particular, the extremal coefficient for a d-dimensional Gumbel-Hougaard is $\theta = d^{1/\tau}$. Hence, the first four, second four and third four raws of the table report the results concerning weak, mild and strong dependence levels, respectively, for such aforementioned copula. From the second to the fifth column the dimension of the copula, the high threshold, the True Exceedance Probability (TEP) and the sample size for each data generation, are reported. The sixth, seventh and eighth column show the Average of the Exceedance Probability Estimates (AEPE), the Standard Deviation (SD) and the bias terms (in absolute value) computed through the $m = 1000$ simulations. The AEPE is close to the TEP for the different dependence levels, thresholds and copula dimensions and the SD and Bias terms are relatively small. In conclusion this study highlights the good performance of our nonparametric estimator for estimating tail probabilities for small and moderately large dimensions.
Second experiment: We compare the performance of our nonparametric estimator with other competitors. There are several results available in the literature for estimating the joint tail dependence in arbitrary dimension Klüppelberg et al. (2008); Einmahl et al. (2012); Beirlant et al. (2016); Einmahl et al. (2018), to name a few, but then their use to estimating the joint tail (or exceedance) probabilities is not straightforward. Recently, in the article Krupskii and Joe (2019) three estimation methods to infer the joint tail probabilities have been proposed. For simplicity, we focus on these proposals as they are simple solutions which are easy to work with. Once again we simulate a sample of n observations from a five-dimensional Gumbel-Hougaard with a certain dependence level. For a large threshold, we use our proposal and those introduced in Krupskii and Joe (2019) to estimate the exceedance (joint tail) probability. We focus only on dimension $d = 5$ to show that our estimator and those proposed in Krupskii and Joe (2019) provide accurate estimates with relatively high dimensions. We repeat this task $m = 1000$ times and we compute a Monte Carlo approximation of the bias term and the standard deviation for the four estimators.

We repeat this experiment for different dependence levels and thresholds and summarize the results in Table 4.2. Its format is similar to that of Table 4.1: TEP is the true exceedance probability over the given threshold. EM is the estimation method where KJ1, KJ2 and KJ3 are methods 1, 2 and 3 from Krupskii–Joe and GPC is the method presented in this paper. AEPE, SD and Bias are the average over the exceedance probability estimates together with standard deviation and bias. Results (third, fifth–seventh column) are given in percentage format. The values $\theta = 4.92, 3.00, 1.11$ represent the cases of weak, mild and strong dependence, respectively, among the 5 variables. To be concise, Table 4.2 reports only the three best estimation results for each simulation setting. The lines highlighted in gray concern the best estimation results, i.e. the value of AEPE closest to EM and with the smallest SD and Bias. For most of the cases our nonparametric estimator outperforms the estimators introduced in Krupskii and Joe (2019). We point out that the second estimation method proposed in Krupskii and Joe (2019) outperforms our method in some cases of strong dependence. The preceding simulation study shows that there are situations where our method performs better than competing methods; a further investigation would be desirable, but this is outside the scope of this thesis.

4.5 A Case Study

Air pollution is an important social issue. It is well-recognized that high emissions of air pollutants have a negative impact on the environment, climate and living beings, e.g., Rossi et al. (1999); Brunekreef and Holgate (2002); World Health Organization (2006); Guerreiro et al. (2016, and the references therein). According to Guerreiro et al. (2016), over a number of decades the European policy on air-quality standards has assisted in reducing emissions of air pollutants. The European air pollution directives regulate emissions of certain pollutants as ozone (O_3), nitrogen dioxide (NO_2), nitrogen oxide (NO), sulphur dioxide (SO_2) and particulate matter (PM_{10}), with the aim of reducing the risk of the negative effects on human health and the environment that these might cause. The last three pollutants are mainly produced by fuel motor vehicles, industry and house-heating, while the first two are produced by some reactions in the atmosphere. On the basis of the World Health Organization (WHO) guidelines World Health Organization (2006), the European emission regulation for air quality standard provides some pollutant concentrations that should not be exceeded. The WHO survey refers how air pollution is linked to adverse health effects by examining appropriate literature. Table 4.3 reports the short-term guideline values, (see World Health Organization (2006, Chapters 10–13) and Guerreiro et al. (2016, Chapters 4–6,8)). For NO the same thresholds than those for NO_2 can be considered. Meeting the short-term concentrations protects against air pollution peaks which can be dangerous to health. The Limit threshold is a high percentile of the pollutant concentration (e.g. hourly, daily mean) in a year. It is recommended not to exceed this threshold with the objective to minimize health effects. Similarly, Target thresholds are proposed for the reduction of air pollution when the pollutant concentrations are still considered very high. Finally, in a country where the Information threshold is exceeded the authorities need to notify their citizens by a public information notice. When even the Alert threshold is exceeded for three consecutive hours, the authorities need to draw up a shortterm action plan in accordance with specific provisions established in European Directive. The threshold values are set for each individual pollutant without taking into account the dependence among pollutants. However, it is well understood that certain pollutants can be dependent on each other; see, e.g., Dahlhaus (2000); Clapp and Jenkin (2001); Heffernan and Tawn (2004); World Health Organization (2006).

Here, we investigate which combinations of thresholds in Table 4.3 are likely to be jointly exceeded and which ones are not. Exceedances of individual thresholds are scarce when these are indeed high pollutant concentrations. This implies in this case that joint exceedances are even more rare. Although the latter event may be very rare by the same token it is a very severe pollution episode. Therefore, accurate estimation of joint exceedance probabilities is an important task. We show how to perform this ambitious mission using the method described in the previous section. We do so analyzing the concentration of O_3, NO_2, NO, SO_2 and PM_{10}, measured at the ground level in $\mu g/m^3$ in the Milan city center, Italy, during the years 2002–2017. Data are collected and made available by the Italian government agency *Agenzia Regionale per la Protezione dell'Ambiente (ARPA)*, see `http://www.arpalombardia.it/sites/QAria`. The first four pollutants are recorded in the average hourly format while the fifth in the daily average. To reveal the dependence among the pollutants we focus on two seasons: summer (May–August) and winter (November–February) (Heffernan and Tawn (2004)). Since the thresholds in Table 4.3 are designed for different averaging periods, for comparison purposes we focus on the daily maximum (of hourly averages) for all the pollutants except for PM_{10} where we are forced to consider the daily average. Figure 4.3 displays in the left and right parts the pairwise scatter plot for the summer and winter datasets, respectively, together with histograms of the individual pollutants levels. SO_2 and O_3 have been removed from the summer and winter datasets, respectively, because they seem independent from the other pollutants. In each dataset the pollutants seem to

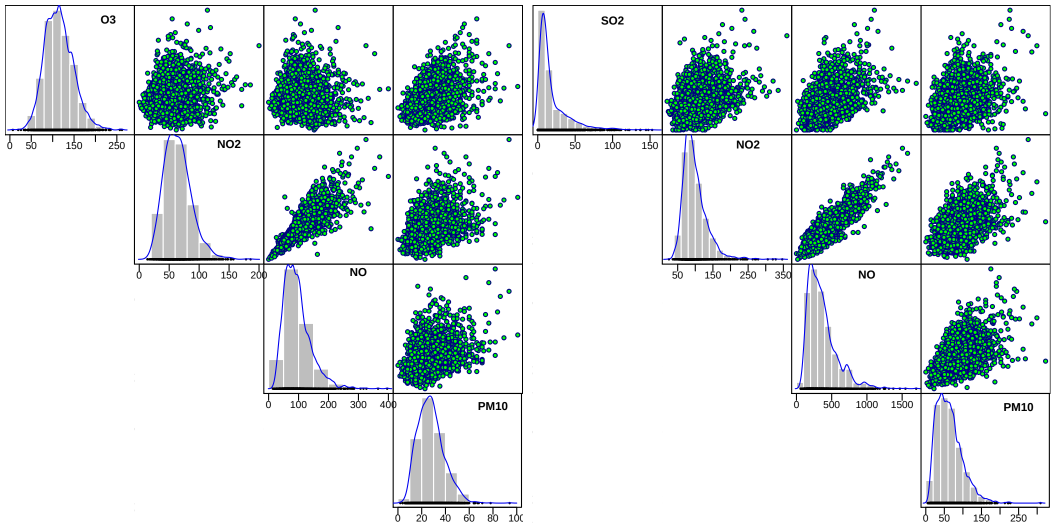

Figure 4.3: Histograms and pairwise plots of pollutants levels in μg/m^3. Left and right panels concern the summer and winter data, respectively.

be highly dependent and this is especially true for NO_2, NO and PM_{10}. In summer, O_3 is moderately dependent to NO_2 and PM_{10}. Finally, we found that in the winter season NO_2, NO and PM_{10} reach much higher pollution concentrations than in summer.

Before applying our approach to estimate the probabilities of joint exceedances (based on the exceedance stability property), we verify whether the assumption of tail dependence is supported by the data. To check whether the tails of a multivariate distribution are asymptotically independent one can use the hypothesis testing proposed by Guillou et al. (2018). Such a test is based on the componentwise maximum approach and works in arbitrary dimensions (greater than or equal to two). We refer to Draisma et al. (2004); Hüsler and Li (2009) and (Falk et al., 2011, Chapter 6.5) for alternative tests that work in the bivariate case. Guillou et al. (2018) proposed using the test statistic $\widehat{S}_n = \sup_{\boldsymbol{t}\in\mathcal{S}_d} \sqrt{n}|\widehat{A}_n(\boldsymbol{t}) - 1|$ to determine whether or not to reject the null hypothesis $A(\boldsymbol{t}) = 1$ for all $\boldsymbol{t} \in \mathcal{S}_d$, i.e., the tails are asymptotically independent, against the alternative hypothesis $A(\boldsymbol{t}) < 1$ for at least one $\boldsymbol{t} \in \mathcal{S}_d$, i.e., some tails are asymptotically dependent (see Guillou et al. (2018) for details). Specifically, A is the Pickands dependence function and $\mathcal{S}_d$ is the d-dimensional unit simplex (see, e.g., (Falk et al., 2011, Ch. 4)). Under the null hypothesis, the asymptotic distribution of th test statistic is known and can be used to provide evidence against the null hypothesis. The quantiles of this distribution are reported in Table 1 of Guillou et al. (2018). We have applied this hypothesis test to the annual summer maxima of (O_3, NO_2, NO, PM_{10}) and annual winter maxima of (SO_2, NO_2, NO, PM_{10}) and we have obtained the following observed values of the test statistics 1.881 and 2.034, respectively. Because the 0.95- and 0.99-quantiles of the distribution of S are 1.480 and 1.740, respectively (see dimension four of Table 1 in Guillou et al. (2018)) we conclude that we reject the hypothesis of the tails' independence and then the assumption of tail dependence is reasonable and we can carry on with our approach for estimating tail probabilities.

Table 4.4 reports 7 possible combinations of the thresholds listed in Table 4.3. In summer, with O_3 approximately 40% of observations exceed the Limit threshold (120 μg/m^3). Similarly, in winter, with NO approximately 78% and 32% of the observations exceed the Limit and Alert thresholds (200 and 400 μg/m^3). Also, with PM approximately 59% of the observations exceed the Limit threshold (50 μg/m^3). Therefore, these thresholds can not be considered extreme values. On the other hand, all the remaining thresholds can be considered extreme values, since only a few observations exceed such pollutants concentrations. In particular, with NO we found that an extreme concentration is 800 μg/m^3, i.e. 2 times the Alert threshold. We estimate the probability of joint exceedances.

We use our approach to estimate the probabilities of joint exceedances that are concerning extreme thresholds. For this purpose, first we estimate for each pollutant the probability, say p_0, of being below an extreme threshold, say y. We do this using the piecing together approach (Falk et al., 2011, Chapter 2.7). In short, we find a high-threshold, say s, with which we can use the survival function of the univariate GPD to approximate the exceeding probability of y, given that the latter is greater than s. We multiply an estimate of such a probability for the probability of exceeding s (which we estimate by the empirical survival function) obtaining an estimate for the unconditional probability of exceeding y (which allows to estimate the unconditional probability of being below than y). The threshold s is selected through the commonly used exploratory graphical methods that are described in (Coles, 2001, Chapters 4.3.1, 4.3.4). The parameters of the generalized Pareto distribution are estimated using the maximum likelihood method (Coles, 2001, Chapter 4.3.2).

Estimates of the variances for the parameters estimates are obtained using the asymptotic variance, see Smith (1984). An estimate of the variance for the estimate of the probability x_0 is obtained using the delta method (Van der

Vaart, 2000, Chapter 3).

Table 4.5 shows the estimation results. Specifically, the column named Threshold reports the extreme thresholds of the scenarios in Table 4.4 with small percentages of exceedances. s indicates the threshold used for estimating the univariate GPD parameters. NE is the number of exceedances of s and EEP is the relative empirical exceedance probability (in percentage format). The values $\hat{\sigma}$ and $\hat{\xi}$ are the estimates of the scale and shape parameters of the univariate GPD, see equation (4.6). The value $\hat{p}_0$ is an estimate of the unconditional probability (in percentage) to be below the extreme threshold reported in the third column (from the left). The standard errors are reported in parentheses. The variance of EEP is obtained using the fact that NE follows a binomial distribution with unknown exceedance probability (estimated by EEP) and sample size n (see Table 4.4).

Once the extreme thresholds were transformed to values in (0,1), we apply the estimation method introduced in Section 4.4 for estimating the probabilities of their exceedances on the copula level, using the empirical copula of the data. Estimation of joint exceedance probabilities on the copula level can be based on on the transformation of the margins if their df are known. It was, however, shown in Bücher (2012) that it is more efficient if the additional knowledge of the margins is ignored and estimators are based on ranks, i.e., if the empirical copula of the initial data is used.

Table 4.6 reports, in the column labeled by $\hat{q}_n$, the estimates of exceedances probabilities (in percentage) for the scenarios listed in Table 4.4. The lower and upper bounds of their 95% confidence interval are reported in the columns LB-CI and UB-CI, respectively. The factors t_0 are given in percentage format as well. Furthermore, estimates for some combinations of three and two extreme thresholds are also reported. The lines highlighted in grey concern the higher estimated probabilities. Scenarios 1 and 4 are not considered because the thresholds for O_3 (in summer) and NO_2 and NO (in winter) are not extreme. However, upper bounds for those probabilities are given by the results listed in the second and twentieth line. Some interpretations are as follows. In summer, we expect that the Information and Limit thresholds for O_3 and PM_{10}, respectively, are simultaneously exceeded on average approximately between two and four times every three years (with the latter that also means once per year). In winter, we expect that the Limit, 2 times the Alert and the Alert thresholds for NO_2, NO and PM_{10}, respectively, are simultaneously exceeded on average approximately between once every two years and once per year. Finally, we expect that 2 times the Alert and the Alert thresholds for NO_2 and NO, respectively, are simultaneously exceeded on average approximately between once and twice per year. Although joint thresholds exceedances do not happen very often, they should not happen at all since the involved thresholds mean indeed very extreme pollution concentrations.

To conclude this chapter, we want to briefly recall what we did: in the setup of a Generalized Pareto model, we were able to derive an estimator for the exceedance probability over some threshold and to justify that the model assumptions hold. Comparing our new approach to some existing ones, we found that our estimator outperforms the competitors when the dependence in the data is not too high and still does well when the dependence increases. Finally, we applied our procedure to air pollution data in Milan, Italy. As the pollution levels were higher in winter than in the summer, we analyzed the two seasons separately. We found that some of the supposedly problematic levels, specifically those for ozone in the summer and NO in the winter, are not actually extreme thresholds, i.e., are exceeded quite often. Even joint exceedances of multiple pollutants at the same time are not quite as uncommon as one might hope.

θ	d	Threshold	TEP	n	AEPE	SD	Bias
1.80	2	$(0.99, 0.99)$	0.208	1500	0.250	0.072	0.042
				2000	0.242	0.067	0.035
2.54	3	$(0.99, 0.99, 0.995)$	0.106	1500	0.126	0.038	0.019
				2000	0.127	0.034	0.021
3.24	4	$(0.99, 0.99, 0.995, 0.999)$	0.033	1500	0.037	0.013	0.004
				2000	0.037	0.012	0.004
3.91	5	$(0.99, 0.99, 0.995, 0.999, 0.9995)$	0.012	1500	0.016	0.009	0.003
				2000	0.015	0.007	0.002
1.50	2	$(0.99, 0.99)$	0.502	1500	0.550	0.084	0.048
				2000	0.541	0.084	0.039
1.90	3	$(0.99, 0.99, 0.995)$	0.293	1500	0.316	0.054	0.023
				2000	0.324	0.050	0.031
2.25	4	$(0.99, 0.99, 0.995, 0.999)$	0.080	1500	0.083	0.019	0.003
				2000	0.082	0.017	0.002
2.57	5	$(0.99, 0.99, 0.995, 0.999, 0.9995)$	0.033	1500	0.035	0.012	0.002
				2000	0.034	0.008	0.001
1.20	2	$(0.99, 0.99)$	0.801	1500	0.838	0.073	0.037
				2000	0.830	0.070	0.029
1.34	3	$(0.99, 0.99, 0.995)$	0.475	1500	0.477	0.041	0.003
				2000	0.490	0.030	0.016
1.44	4	$(0.99, 0.99, 0.995, 0.999)$	0.100	1500	0.101	0.014	0.001
				2000	0.100	0.013	0.000
1.53	5	$(0.99, 0.99, 0.995, 0.999, 0.9995)$	0.048	1500	0.048	0.010	0.000
				2000	0.049	0.010	0.001

Table 4.1: Performance of the nonparametric estimator (4.13) for the estimation of the joint tail probabilities based on Gumbel-Hougaard family of copulas. Results (fourth, sixth–eighth column) are given in percentage format. TEP is the true exceedance probability over the given threshold. AEPE, SD and Bias are the average over the exceedance probability estimates together with standard deviation and bias over 1000 iterations.

θ	Threshold	TEP	EM	AEPE	SD	Bias
4.92	(0.95,0.95,0.95,0.95,0.95)	0.031	KJ2	0.331	0.195	0.300
			KJ3	0.349	0.188	0.318
			GPC	0.072	0.046	0.041
3.00		1.594	KJ2	2.398	0.573	0.803
			KJ3	2.443	0.584	0.849
			GPC	1.790	0.434	0.196
1.11		3.662	KJ2	3.967	0.671	0.305
			KJ3	4.032	0.650	0.371
			GPC	3.880	0.497	0.218
4.92	(0.99,0.99,0.99,0.99,0.99)	0.006	KJ2	0.028	0.035	0.022
			KJ3	0.032	0.035	0.027
			GPC	0.014	0.009	0.008
3.00		0.310	KJ1	0.474	0.162	0.164
			KJ2	0.435	0.152	0.125
			GPC	0.359	0.096	0.048
1.11		0.727	KJ1	0.789	0.174	0.062
			KJ2	0.751	0.172	0.024
			GPC	0.770	0.111	0.044
4.92	(0.995,0.995,0.995,0.995,0.995)	0.003	KJ2	0.012	0.018	0.009
			KJ3	0.014	0.017	0.011
			GPC	0.007	0.007	0.004
3.00		0.155	KJ1	0.235	0.086	0.080
			KJ2	0.214	0.081	0.059
			GPC	0.183	0.044	0.028
1.11		0.363	KJ1	0.398	0.094	0.035
			KJ2	0.379	0.093	0.016
			GPC	0.393	0.063	0.030
4.92	(0.999,0.999,0.999,0.999,0.999)	0.001	KJ2	0.002	0.003	0.001
			KJ3	0.002	0.004	0.002
			GPC	0.001	0.001	0.001
3.00		0.031	KJ1	0.045	0.019	0.014
			KJ2	0.041	0.019	0.010
			GPC	0.036	0.010	0.005
1.11		0.073	KJ1	0.077	0.022	0.004
			KJ2	0.073	0.021	0.001
			GPC	0.078	0.011	0.005

Table 4.2: Comparison of our nonparametric estimator with other competitors.

Pollutant	Threshold	Period	Value in μg/m^3	Recommendation
O_3	Limit	Daily max	120	no more than 25 exceedances per year
	Information		180	
	Alert		240	
NO_2	Limit	1-hour mean	200	no more than 18 exceedances per year
	Alert		400	
SO_2	Limit	24-hour mean	125	no more than 3 exceedances per year
PM_{10}	Limit	24-hour mean	50	no more than 35 exceedances per year
	Target		150	

Table 4.3: Pollutant concentrations (thresholds) that should not be exceeded according the European emission regulation for air quality standard.

Scenario		Pollutant				JEEP
Summer ($n = 1655$)		O_3	NO_2	NO	PM_{10}	
1	Threshold	120	200	200	50	0
	MEEP	40.181	0	2.961	3.444	
2	Threshold	180	200	200	50	0
	MEEP	3.263	0	2.961	3.444	
3	Threshold	240	400	400	150	0
	MEEP	0	0	0	0	
		Pollutant				
Winter ($n = 1713$)		SO_2	NO_2	NO	PM_{10}	
4	Threshold	125	200	200	50	0.0584
	MEEP	0.350	1.459	78.167	58.785	
5	Threshold	125	200	400	150	0.0584
	MEEP	0.350	1.459	32.399	1.926	
6	Threshold	125	200	800	150	0.0584
	MEEP	0.350	1.459	3.853	1.926	
7	Threshold	125	400	800	150	0
	MEEP	0.350	0	3.853	1.926	

Table 4.4: Marginal and joint empirical probability of threshold exceedances for different combinations of thresholds. The first 3 scenarios concern the summer season (with $n = 1655$ observations) and the last four (with $n = 1713$ observations) the winter season. For each scenario the Joint Empirical Exceedance Probability (JEEP) and for each individual pollutant the Marginal Empirical Exceedance Probability (MEEP) over the given threshold is reported (in percentage format).

	Pollutant	Threshold	GPD Estimates					
			s	NE	EEP	$\hat{\sigma}$	$\hat{\xi}$	$\hat{p}_0$
Summer	O_3	180	150	226	13.656 (0.844)	21.860 (1.936)	-0.114 (0.058)	96.93 0 (0.804)
		240						99.947 (0.091)
	NO_2	200	96	136	8.218 (0.675)	14.870 (1.862)	0.067 (0.091)	99.973 (0.048)
		400						99.999 (0.0001)
	NO	200	150	176	10.634 (0.758)	36.401 (3.970)	0.047 (0.079)	97.191 (0.822)
		400						99.972 (0.048)
	PM_{10}	50	47	89	5.378 (0.554)	6.387 (0.977)	0.041 (0.110)	96.623 (2.391)
		150						99.999 (0.0002)
Winter	SO_2	125	40	233	13.602 (0.843)	24.131 (2.206)	-0.026 (0.064)	99.662 (0.258)
	NO_2	200	130	240	14.011	24.376 (0.853)	0.192 (0.077)	98.576 (0.517)
		400						99.963 (0.046)
	NO	800	600	206	12.026 (0.800)	195.61 (18.991)	-0.029 (0.068)	95.741 (0.978)
	PM_{10}	150	100	238	13.894 (0.850)	28.222 (2.558)	-0.023 (0.063)	97.722 (0.684)

Table 4.5: Estimate of the GPD parameters and the unconditional probability that the amount of the pollutant is below the individual extreme threshold. s is the threshold used for estimating the univariate GPD parameters, NE the number of exceedances of s and EEP is the relative empirical exceedance probability (in percentage format). $\hat{\sigma}$ and $\hat{\xi}$ are the estimates of the scale and shape parameters of the univariate GPD, see equation (4.6). $\hat{p}_0$ is an estimate of the unconditional probability (in percentage) to be below the respective extreme threshold reported. The standard errors are reported in parentheses.

Scn.	(O_3, NO_2, NO, PM_{10})	t_0	$\hat{p}_n$	$\hat{q}_n$	LB-CI	UB-CI
2	(180, 200, 200, 50)	4.6606	0.6042	0.0282	0.0135	0.0517
	(, 200, 200, 50)	3.3894	0.8459	0.0287	0.0157	0.0480
	(180, , 200, 50)	44.561	0.4834	0.2154	0.0931	0.4234
	(180, 200, , 50)	4.7789	0.5438	0.0260	0.0119	0.0492
	(180, 200, 200,)	3.3901	0.7855	0.0266	0.0142	0.0454
	(180, 200, ,)	4.0065	0.6647	0.0266	0.0133	0.0475
	(180, , 200,)	39.4044	0.7251	0.2857	0.1478	0.4977
	(180, , , 50)	26.1881	3.6858	0.9652	0.7413	1.2334
	(, 200, 200,)	2.9383	0.9063	0.0266	0.0149	0.0438
	(, 200, , 50)	3.3901	0.7855	0.0266	0.0142	0.0454
	(, , 200, 50)	37.2618	1.5710	0.5854	0.3833	0.8546
3	(240, 400, 400, 150)	0.3435	0	0	0	0.0008
Scn.	(SO_2, NO_2, NO, PM_{10})	t_0	$\hat{p}_n$	$\hat{q}_n$	LB-CI	UB-CI
5	(125, 200, 400, 150)	76.2093	0.1751	0.1335	0.0275	0.3894
6	(125, 200, 800, 150)	7.6186	1.5178	0.1156	0.0757	0.1688
	(, 200, 800, 150)	51.3463	1.3427	0.6894	0.4380	1.0310
	(125, , 800, 150)	24.0850	0.7589	0.1828	0.0975	0.3117
	(125, 200, , 150)	22.0570	0.5838	0.1288	0.0618	0.2362
	(125, 200, 800,)	38.6023	0.3503	0.1352	0.0497	0.2937
	(125, 200, ,)	7.3745	1.9848	0.1464	0.1016	0.2037
	(125, , 800,)	4.2888	7.8809	0.3380	0.2852	0.3971
	(125, , , 150)	7.0823	2.7437	0.1943	0.1433	0.2572
	(, 200, 800,)	4.2635	33.3917	1.4236	1.3285	1.5213
	(, 200, , 150)	14.5064	5.5458	0.8045	0.6542	0.9773
	(, , 800, 150)	22.4274	5.6042	1.2569	1.0233	1.5253
7	(125, 400, 800, 150)	4.8869	0.1751	0.0086	0.0018	0.0250

Table 4.6: The probability estimates of joint exceedances of extreme thresholds for the scenarios listed in Table 4.4 are in the column labeled $\hat{q}_n$. Estimates for some combinations of three and two extreme thresholds are also reported. The columns LB-CI and UB-CI are the lower and upper bounds of their 95% confidence intervals, respectively. The selected factors t_0 and the intermediate probabilities p_0 are included as well. Results are given in percentage format.

Chapter 5

Miscellaneous results

5.1 Derivatives of D-norms

An elementary concept of calculus is the derivative of a univariate function $f : \mathbb{R} \mapsto \mathbb{R}$ at some point x, defined by the limit $f'(x) := \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ if it exists. One way to generalize this to scalar functions $f : \mathbb{R}^d \mapsto \mathbb{R}$ is to consider the derivative in some direction $\boldsymbol{y} \in \mathbb{R}^d$ as

$$\frac{\partial}{\partial \boldsymbol{y}} f(\boldsymbol{x}) := \lim_{\varepsilon\to 0} \frac{f(\boldsymbol{x}+\varepsilon\boldsymbol{y}) - f(\boldsymbol{x})}{\varepsilon}$$

if the limit exits at the point $\boldsymbol{x}$. A special case of these directional derivatives are the partial derivatives where $\boldsymbol{y} = \boldsymbol{e}_i$ is a unit vector and we use the shorthand $\frac{\partial}{\partial_i} f(\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{e}_i} f(\boldsymbol{x})$ for that. Note that $\boldsymbol{y} = \boldsymbol{0}$ is not excluded in the definition, but in this case we have trivially $\frac{\partial}{\partial \boldsymbol{0}} f(\boldsymbol{x}) = 0$.

Sometimes, the limit does not exist but its one-sided counterpart does. We denote these one-sided directional derivatves as

$$\frac{\partial^+}{\partial \boldsymbol{y}} f(\boldsymbol{x}) := \lim_{\varepsilon\downarrow 0} \frac{f(\boldsymbol{x}+\varepsilon\boldsymbol{y}) - f(\boldsymbol{x})}{\varepsilon}$$

or

$$\frac{\partial^-}{\partial \boldsymbol{y}} f(\boldsymbol{x}) := \lim_{\varepsilon\uparrow 0} \frac{f(\boldsymbol{x}+\varepsilon\boldsymbol{y}) - f(\boldsymbol{x})}{\varepsilon} = \lim_{\varepsilon\downarrow 0} \frac{f(\boldsymbol{x}) - f(\boldsymbol{x}-\varepsilon\boldsymbol{y})}{\varepsilon}$$

respectively. Obviously, the directional derivative exists if both one-sided derivatives exist and have the same value.

Remark 5.1. *The derivatives of D-norms have already been considered in the literature. As an example, Einmahl et al. (2012) uses them in terms of the stable tail dependence function and noted that the one-sided partial derivatives always exist. Aulbach et al. (2015) addressed the topic for functional D-norms. However, they all just use the fact that a norm is convex, hence differentiable almost everywhere and formulate their results for the points where differentiability holds. We will give a condition that ensures the existence of the derivative for a specific point and direction and work out a representation from there.*

We define the (random) set

$$\mathcal{M}_{\boldsymbol{X}} := \left\{ i \,\middle|\, X_i = \max_{1\le j\le d} X_j \right\}$$

of indices of a maximal component of some d-variate rv $\boldsymbol{X}$ and put

$$M_{\boldsymbol{X}} := \min \mathcal{M}_{\boldsymbol{X}}.$$

So $M_{\boldsymbol{X}}$ is the (random) minimal element of $\mathcal{M}_{\boldsymbol{X}}$; note that $\mathcal{M}_{\boldsymbol{X}}$ cannot be empty because there always is a maximal component. In general, we have $\{M_{\boldsymbol{X}} = i\} \subseteq \{\max_{1\le j\le d} X_j = X_i\}$ for all $i = 1,\dots,d$. Equality holds if the maximal components are unique and equality holds almost surely if the uniqueness holds almost surely.

While being somewhat more complicated, this construction has an advantage: the sets $\{M_{\boldsymbol{X}} = i\}$ for $i = 1,\dots,d$ are a partition of Ω. The sets $\{\max_{1\le j\le d} X_j = X_i\}$, $i = 1,\dots,d$, on the other hand are not disjoint in general. This partitioning is used in the following lemma:

Lemma 5.2. *Consider $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d$ and a d-variate rv $\boldsymbol{Z}$ with finite expectation. Then we have for $\varepsilon > 0$*

$$E\left(\max_{1\le j\le d}(\boldsymbol{x}+\varepsilon\boldsymbol{y})_j Z_j\right) - E\left(\max_{1\le j\le d}(x_j Z_j)\right) = \varepsilon \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right) + o(\varepsilon)$$

if $P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right) \to_{\varepsilon\downarrow 0} 0$ for $i,k = 1,\dots,d$ with $k \ne i$.

Proof. We have for $\varepsilon > 0$

$$E\left(\max_{1\le j\le d}(\boldsymbol{x}+\varepsilon\boldsymbol{y})_j Z_j\right) - E\left(\max_{1\le j\le d}(x_j Z_j)\right)$$

$$
\begin{aligned}
&= E\left(\max_{1\le j\le d}(\boldsymbol{x}+\varepsilon\boldsymbol{y})_j Z_j - \max_{1\le j\le d}(x_j Z_j)\right)\\
&= E\left(\left(\max_{1\le j\le d}(\boldsymbol{x}+\varepsilon\boldsymbol{y})_j Z_j - \max_{1\le j\le d}(x_j Z_j)\right)\sum_{i=1}^d 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\sum_{k=1}^d 1_{\left\{M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right)\\
&= \sum_{i=1}^d\sum_{k=1}^d E\left(\left(\max_{1\le j\le d}(\boldsymbol{x}+\varepsilon\boldsymbol{y})_j Z_j - \max_{1\le j\le d}(x_j Z_j)\right)1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right)\\
&= \sum_{i=1}^d\sum_{k=1}^d E\left(\left((x_k+\varepsilon y_k)Z_k - x_i Z_i\right)1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right)\\
&= \varepsilon\sum_{i=1}^d E\left(y_i Z_i 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=i\right\}}\right)\\
&\quad + \varepsilon\sum_{i=1}^d\sum_{\substack{k=1\\k\ne i}}^d E\left(y_k Z_k 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right)\\
&\quad + \sum_{i=1}^d\sum_{\substack{k=1\\k\ne i}}^d E\left(\left(x_k Z_k - x_i Z_i\right)1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right)\\
&= \varepsilon\sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right) + o(\varepsilon)
\end{aligned}
$$

under the condition on the sets $\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}$. To see this, we compute

$$
\begin{aligned}
&\varepsilon\sum_{i=1}^d E\left(y_i Z_i 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=i\right\}}\right)\\
&\quad = \varepsilon\sum_{i=1}^d E\left(y_i Z_i\left(1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}} - \sum_{\substack{k=1\\k\ne i}}^d 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right)\right)\\
&\quad = \varepsilon\sum_{i=1}^d E\left(y_i Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right)\\
&\qquad - \varepsilon\sum_{i=1}^d\sum_{\substack{k=1\\k\ne i}}^d E\left(y_i Z_i 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\,M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right).
\end{aligned}
$$

and observe that

$$\begin{aligned}&\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\\&\quad\subseteq \{x_iZ_i \ge x_kZ_k,\, (x_k+\varepsilon y_k)Z_k \ge (x_i+\varepsilon y_i)Z_i\}\\&\quad= \{0 \ge x_kZ_k - x_iZ_i \ge \varepsilon y_iZ_i - \varepsilon y_kZ_k\}.\end{aligned}$$

Therefore, we have for the error term

$$2\varepsilon \sum_{i=1}^{d} \sum_{\substack{k=1\\k\neq i}}^{d} E\left((y_kZ_k - y_iZ_i)\, 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}\right) = o(\varepsilon)$$

by the dominated convergence theorem because $\boldsymbol{Z}$ has finite expectation and the condition

$$P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right) \to_{\varepsilon\downarrow 0} 0$$

implies that $((y_kZ_k - y_iZ_i)\, 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}}=i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}}=k\right\}}$ converges to 0 in probability as ε converges to zero. Note that it already is equal to 0 if both y_i and y_k are 0. □

Remark 5.3. *If we have $P(x_iZ_i = x_kZ_k) = 0$ for fixed $\boldsymbol{x} \in \mathbb{R}^d$ and some $i, k \in \{1, \dots, d\}$ with $k \neq i$, then the condition*

$$P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right) \to_{\varepsilon\to 0} 0$$

is fulfilled for all $\boldsymbol{y} \in \mathbb{R}^d$. This can be seen as follows:

$$\begin{aligned}&P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right)\\&\quad\le P\left(\bigcap_{j=1}^{d}\{x_iZ_i \ge x_jZ_j\} \cap \bigcap_{j=1}^{d}\{x_kZ_k + \varepsilon y_kZ_k \ge x_jZ_j + \varepsilon y_jZ_j\}\right)\\&\quad\le P\left(\{x_iZ_i \ge x_kZ_k\} \cap \{x_kZ_k + \varepsilon y_kZ_k \ge x_iZ_i + \varepsilon y_iZ_i\}\right)\\&\quad= P\left(\{\varepsilon(y_kZ_k - y_iZ_i) \ge x_iZ_i - x_kZ_k \ge 0\}\right)\\&\quad= P\left(\left\{\varepsilon \ge \frac{x_iZ_i - x_kZ_k}{y_kZ_k - y_iZ_i} \ge 0,\, y_kZ_k \neq y_iZ_i\right\}\right)\\&\qquad+ P\left(\varepsilon\cdot 0 \ge x_iZ_i - x_kZ_k \ge 0,\, y_kZ_k = y_iZ_i\right)\\&\quad\to_{\varepsilon\to 0} P\left(\left\{0 \ge \frac{x_iZ_i - x_kZ_k}{y_kZ_k - y_iZ_i} \ge 0, y_kZ_k \neq y_iZ_i\right\}\right)\end{aligned}$$

$$+ P\left(x_iZ_i - x_kZ_k = 0,\, y_kZ_k = y_iZ_i\right)$$
$$\le P\left(x_iZ_i = x_kZ_k\right) = 0$$

In particular, if $Z_1, \dots, Z_d$ are pairwise independent rvs with continuous distributions, this is true for arbitrary $\boldsymbol{x} > \boldsymbol{0}$ or even $\boldsymbol{x} \in \mathbb{R}^d$ where at most one component x_i is equal to zero.

Examples where the condition does not hold will be considered after we apply the result to D-norms:

Lemma 5.4. *Let $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ be a generator of the (d-variate) D-norm $\|\cdot\|_D$. The one-sided directional derivatives are for $\boldsymbol{x} > \boldsymbol{0} \in \mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^d$*

$$\frac{\partial^+}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_D = \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}} = i\}}\right),$$

if $P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right) \to_{\varepsilon\downarrow 0} 0$ for $i, k = 1, \dots, d$ with $k \ne i$.

Proof. We have $\|\boldsymbol{x}\|_D = E\left(\max_{1\le i\le d} |x_i|\, Z_i\right) = E\left(\max_{1\le i\le d} x_iZ_i\right)$ because $\boldsymbol{x} > \boldsymbol{0} \in \mathbb{R}^d$ and, for small enough $\varepsilon > 0$,

$$\|\boldsymbol{x} + \varepsilon\boldsymbol{y}\|_D = E\left(\max_{1\le i\le d} |x_i + \varepsilon y_i|\, Z_i\right) = E\left(\max_{1\le i\le d} (x_i + \varepsilon y_i) Z_i\right)$$

holds as well. As a consequence, we can apply the previous lemma to obtain for small enough $\varepsilon > 0$

$$\frac{\|\boldsymbol{x} + \varepsilon\boldsymbol{y}\|_D - \|\boldsymbol{x}\|_D}{\varepsilon} = \frac{E\left(\max_{1\le j\le d}(x_j + \varepsilon y_j)Z_j\right) - E\left(\max_{1\le j\le d}(x_jZ_j)\right)}{\varepsilon}$$
$$\to_{\varepsilon\downarrow 0} \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}} = i\}}\right)$$

as the one-sided limit. □

If we make the condition a bit stronger by requiring the limit of zero not only for $\varepsilon \downarrow 0$ but for $\varepsilon \to 0$, we get the result for directional derivatives:

Lemma 5.5. *Let $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ be a generator of the (d-variate) D-norm $\|\cdot\|_D$. Then the directional derivatives are for $\boldsymbol{x} > \boldsymbol{0} \in \mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^d$*

$$\frac{\partial}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_D = \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}} = i\}}\right),$$

if $P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right) \to_{\varepsilon\to 0} 0$ for $i, k = 1, \dots, d$ with $k \ne i$.

Proof. We already know that $\frac{\partial^+}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_D = \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right)$. For $\varepsilon < 0$, we can switch to $\tilde{\varepsilon} := -\varepsilon$ and $\tilde{\boldsymbol{y}} := -\boldsymbol{y}$ to compute

$$\frac{\|\boldsymbol{x}+\varepsilon\boldsymbol{y}\|_D - \|\boldsymbol{x}\|_D}{\varepsilon} = -\frac{\|\boldsymbol{x}+\tilde{\varepsilon}\tilde{\boldsymbol{y}}\|_D - \|\boldsymbol{x}\|_D}{\tilde{\varepsilon}}$$
$$\to_{\varepsilon\downarrow 0} -\sum_{i=1}^d \tilde{y}_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right) = \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right).$$

So both one-sided limits are the same, hence this is the directional derivative at $\boldsymbol{x}$. □

Remark 5.6. *Note that the derivatives do not depend on the scale of $\boldsymbol{x}$ as we have $M_{c\boldsymbol{x}\boldsymbol{Z}} = M_{\boldsymbol{x}\boldsymbol{Z}}$ for $c > 0$. Therefore, the expectation does not change. We do not have a problem with the condition either because*

$$P\left(\left\{M_{c\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(c\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right) = P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{c(\boldsymbol{x}+\frac{\varepsilon}{c}\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right)$$
$$= P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\tilde{\varepsilon}\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right)$$

and $\tilde{\varepsilon} := \varepsilon/c \to 0$ iff $\varepsilon \to 0$. The same considerations apply to the one-sided derivatives.

In the next examples we use our result to compute some derivatives of D-norms, but also take a closer look at the meaning of the additional condition and its relation to the differentiability of the norm.

Example 5.7. We first look at $\|\cdot\|_\infty$ with constant generator $\boldsymbol{Z} = \boldsymbol{1} \in \mathbb{R}^d$ and some $\boldsymbol{x} > \boldsymbol{0}$. Recall that $\mathcal{M}_{\boldsymbol{x}}$ is the (deterministic) index-set of the maximal components of $\boldsymbol{x}$ and $m := M_{\boldsymbol{x}}$ is the minimum of those indices. Let $y_{m^+} := \max_{i\in\mathcal{M}_{\boldsymbol{x}}} y_i$ and $y_{m^-} := \min_{i\in\mathcal{M}_{\boldsymbol{x}}} y_i$ be the maximal and minimal of those components of $\boldsymbol{y}$ that correspond to the maximal entries of $\boldsymbol{x}$.

Now there are two cases to consider: if $y_{m^-} = y_{m^+}$, then $y_m = y_i$ for all $i \in \mathcal{M}_{\boldsymbol{x}}$. We have $M_{\boldsymbol{x}\boldsymbol{Z}} = M_{\boldsymbol{x}} = m$ and, for $|\varepsilon|$ small enough, $M_{\boldsymbol{x}+\varepsilon\boldsymbol{y}} = m$ as well. Therefore $P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = j\right\}\right) = 0$ for $i \neq j$, the condition of Lemma 5.5 is fulfilled and it turns out that

$$\frac{\partial}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_\infty = y_m$$

as $E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right) = 0$ for $i \neq m$ and only $E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=m\}}\right) = E(1\cdot 1) = 1$. Note that we are trivially in this case if the maximal component of $\boldsymbol{x}$ is unique.

On the other hand, we could have $y_{m^-} < y_{m^+}$. As before, $M_{\boldsymbol{x}} = m$, yet $M_{\boldsymbol{x}+\varepsilon \boldsymbol{y}} = m^+$ for $\varepsilon > 0$ and $M_{\boldsymbol{x}+\varepsilon \boldsymbol{y}} = m^-$ for $\varepsilon < 0$ with small enough $|\varepsilon|$. But this means that the condition of Lemma 5.5 is violated for at least one of the pairs $i = m, j = m^+$ or $i = m, j = m^-$ as $m^+ \neq m^-$. We can check directly that $\|\cdot\|_\infty$ is not differentiable in direction $\boldsymbol{y}$ in this case:

$$\lim_{\varepsilon \downarrow 0} \frac{\|\boldsymbol{x} + \varepsilon \boldsymbol{y}\|_\infty - \|\boldsymbol{x}\|_\infty}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{(x_{m^+} + \varepsilon y_{m^+}) - x_m}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{\varepsilon y_{m^+}}{\varepsilon} = y_{m^+}$$

and

$$\lim_{\varepsilon \downarrow 0} \frac{\|\boldsymbol{x}\|_\infty - \|\boldsymbol{x} - \varepsilon \boldsymbol{y}\|_\infty}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{x_m - (x_{m^-} - \varepsilon y_{m^-})}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{\varepsilon y_{m^-}}{\varepsilon} = y_{m^-}$$

but $y_{m^+} \neq y_{m^-}$. So the points of non-differentiability correspond perfectly to the violation of our condition.

This is no longer true if we consider the one-sided directional derivatives: by the considerations above, they exist in all points $\boldsymbol{x} > \boldsymbol{0}$ and directions $\boldsymbol{y}$, but if $y_{m^+} > y_m$, the one-sided condition is still violated. Even more, $\frac{\partial^+}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_\infty = y_{m^+}$ and if Lemma 5.4 were applicable, its result would be $\frac{\partial^+}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_\infty = y_m \neq y_{m^+}$. Therefore it is not just a case where the condition is unnecessarily strong, but the result is slightly different as well.

For our specific example, an idea to fix that might be to factor in the direction $\boldsymbol{y}$ to pick $M_{\boldsymbol{x}} = M_{\boldsymbol{x}\boldsymbol{Z}}$ instead of just picking the minimum entry of $\mathcal{M}_{\boldsymbol{x}}$. If we had $M_{\boldsymbol{x}} = m^+$, everything would work out. But it is highly unclear, how exactly and whether at all this would work for non-trival cases of other D-norms with non-constant generators.

Example 5.8. Now we want to investigate an instance where the condition of Lemma 5.5 is violated by a non-constant generator. Consider a rv $\boldsymbol{Z} = (Z_1, Z_2)$ with $P(Z_1 = \frac{1}{2}) = \frac{1}{2} = P(Z_1 = \frac{3}{2})$ and $Z_2 = 2 - Z_1$. Then $\boldsymbol{Z}$ generates the D-norm where we have for $\boldsymbol{x} > \boldsymbol{0} \in \mathbb{R}^2$

$$\begin{aligned}
\|\boldsymbol{x}\|_D &= E\left(\max\{x_1 Z_1, x_2 Z_2\}\right) \\
&= \frac{1}{2}\left(\max\left\{\frac{x_1}{2}, \frac{3x_2}{2}\right\} + \max\left\{\frac{3x_1}{2}, \frac{x_2}{2}\right\}\right) \\
&= \frac{1}{8}\left(x_1 + 3x_2 + |x_1 - 3x_2| + 3x_1 + x_2 + |3x_1 - x_2|\right) \\
&= \frac{x_1 + x_2}{2} + \frac{|x_1 - 3x_2| + |3x_1 - x_2|}{8}.
\end{aligned}$$

This is not differentiable for all $\boldsymbol{x} > \mathbf{0}$. Specifically, if $x_1 = 3x_2$, then the part

$$\frac{|(x_1 + \varepsilon y_1) - 3(x_2 + \varepsilon y_2)|}{\varepsilon} = \frac{|\varepsilon y_1 - 3\varepsilon y_2|}{\varepsilon} = \frac{|\varepsilon|}{\varepsilon}|y_1 - 3y_2|$$

leads to different limits for $\varepsilon \downarrow 0$ and $\varepsilon \uparrow 0$ unless $y_1 = 3y_2$. For the remaining parts, left- and right-sided limits coincide. Hence, at these points $\boldsymbol{x}$, the norm is only differentiable in direction $\boldsymbol{y} = (3,1)^T c$ for any factor $c \in \mathbb{R}$.
The same arguments applied to $|3x_1 - x_2|$ lead to the conclusion that for $3x_1 = x_2$, it is only differentiable in direction $\boldsymbol{y} = (1,3)^T c$ with $c \in \mathbb{R}$. But in all other points $\boldsymbol{x}$, it is differentiable in arbitrary direction $\boldsymbol{y}$.
These critical combinations are precisely the ones where the condition of Lemma 5.5 is violated: for $x_1 = 3x_2$ and $\varepsilon > 0$ we have

$$\begin{aligned}
&P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = 1,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = 2\right\}\right) \\
&= P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = 1,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = 2,\, Z_1 = \frac{1}{2}\right\}\right) \\
&\quad + P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = 1,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = 2,\, Z_1 = \frac{3}{2}\right\}\right) \\
&= P\left(\left\{\frac{3x_2}{2} + \frac{3}{2}\varepsilon y_2 > \frac{x_1}{2} + \frac{1}{2}\varepsilon y_1,\, Z_1 = \frac{1}{2}\right\}\right) \\
&\quad + P\left(\left\{\frac{x_2}{2} + \frac{1}{2}\varepsilon y_2 > \frac{3x_1}{2} + \frac{3}{2}\varepsilon y_1,\, Z_1 = \frac{3}{2}\right\}\right) \\
&= P\left(\left\{3y_2 > y_1,\, Z_1 = \frac{1}{2}\right\}\right) = \begin{cases} \frac{1}{2} & \text{if } 3y_2 > y_1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

and similarly

$$P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = 1,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = 2\right\}\right) = \begin{cases} \frac{1}{2} & \text{if } 3y_2 < y_1 \\ 0 & \text{otherwise} \end{cases}$$

for $\varepsilon < 0$. Therefore, the lemma is applicable only for $y_1 = 3y_2$ and this is the only direction of differentiability. The analogous result can be deduced for $3x_1 = x_2$. For all other $\boldsymbol{x} > \mathbf{0}$, we have differentiability in all directions according to remark 5.3 as $P(x_1 Z_1 = x_2 Z_2) = P\left(x_1 = 3x_2,\, Z_1 = \frac{1}{2}\right) + P\left(3x_1 = x_2,\, Z_1 = \frac{3}{2}\right) = 0$ in those cases.

So the points where our condition is violated correspond exactly to the points of non-differentiability again. The question whether this is always the case, however, requires further work.

Finally we come to the other p-norms, where the condition is always fulfilled:

Example 5.9. The case $p = 1$ is very easy: we know that a generator $\boldsymbol{Z}$ of $\|\cdot\|_1$ is given by the random permutation of $(d, 0, \dots, 0)^T \in \mathbb{R}^d$. In this case we have $\{M_{\boldsymbol{x}\boldsymbol{Z}} = i\} = \{Z_i = d, Z_j = 0, j \neq i\}$ for $\boldsymbol{x} > \boldsymbol{0}$. Therefore, $P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}} = k\right\}\right) = 0$ for $i, k = 1, \dots, d$ with $k \neq i$, arbitrary $\boldsymbol{y} \in \mathbb{R}^d$ and $|\varepsilon|$ small enough and we can apply our lemma to derive

$$\frac{\partial}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_1 = \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=i\}}\right) = \sum_{i=1}^d y_i E\left(Z_i 1_{\{Z_i=d\}}\right) = \sum_{i=1}^d y_i.$$

For $p \in (1, \infty)$ we have seen in Section 1.2 that each logistic norm $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, can be obtained from the generator $\boldsymbol{Z} = (Z_1, \dots, Z_d) = (Y_1, \dots, Y_d)/\Gamma(1 - 1/p)$, where $Y_1^{1/p}, \dots, Y_d^{1/p}$ are iid Fréchet-distributed rvs. Remark 5.3 shows that the condition of Lemma 5.5 is always fulfilled, hence the directional derivatives exist for all points $\boldsymbol{x} > \boldsymbol{0}$ and arbitrary direction.

For the actual computation, we want to evaluate the expectation using the joint density of $\boldsymbol{Y}$. As the components Y_i are independent, this is the product density

$$f_Y(\boldsymbol{y}) = \prod_{i=1}^d f_{Y_i}(y_i) = \prod_{i=1}^d \left[\frac{1}{y_i^2} \exp\left(-\frac{1}{y_i}\right)\right], \quad \boldsymbol{y} > \boldsymbol{0} \in \mathbb{R}^d.$$

We have $\{M_{\boldsymbol{x}\boldsymbol{Z}} = 1\} = \{x_1 Z_1 \geq x_j Z_j,\, 2 \leq j \leq d\}$ almost surely because the distributions are continuous, too. Specifically, the df of each Y_i is $F_{Y_i}(y_i) = \exp(-1/y_i)$ for $y_i > 0$, therefore $\int_0^z f_{Y_i}(y_i)\, dy_i = \exp(-1/z)$ for $z > 0$. These observations allow us to compute

$$\begin{aligned}
&E\left(Z_1\, 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}}=1\}}\right) \cdot \Gamma(1 - 1/p)\\
&\quad = E\left(Z_1\, 1_{\{x_1 Z_1 \geq x_j Z_j,\, 2 \leq j \leq d\}}\right) \cdot \Gamma(1 - 1/p)\\
&\quad = E\left(Y_1^{1/p}\, 1_{\left\{x_1 Y_1^{1/p} \geq x_j Y_j^{1/p},\, 2 \leq j \leq d\right\}}\right)\\
&\quad = E\left(Y_1^{1/p}\, 1_{\left\{\left(\frac{x_1}{x_j}\right)^p Y_i \geq Y_j,\, 2 \leq j \leq d\right\}}\right)\\
&\quad = \int_0^\infty \int_0^{\left(\frac{x_1}{x_2}\right)^p y_1} \cdots \int_0^{\left(\frac{x_1}{x_d}\right)^p y_1} y_1^{1/p} \prod_{i=1}^d \left[\frac{1}{y_i^2} \exp\left(-\frac{1}{y_i}\right)\right] dy_d \dots dy_1
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty y_1^{1/p} \frac{1}{y_1^2} \exp\left(-\frac{1}{y_1}\right) \prod_{i=2}^d \left[\int_0^{\left(\frac{x_1}{x_j}\right)^p y_1} \frac{1}{y_i^2} \exp\left(-\frac{1}{y_i}\right) dy_i \right] dy_1 \\
&= \int_0^\infty y_1^{1/p-2} \exp\left(-\frac{1}{y_1}\right) \prod_{i=2}^d \exp\left(-\frac{x_j^p}{x_1^p y_1}\right) dy_1 \\
&= \int_0^\infty y_1^{1/p-2} \exp\left(-\sum_{j=1}^d \frac{x_j^p}{x_1^p y_1}\right) dy_1 \\
&= \int_0^\infty \left(\sum_{j=1}^d \frac{x_j^p}{x_1^p} \frac{1}{y_1}\right)^{\frac{1}{p}-2} \exp\left(-y_1\right) \left(\sum_{j=1}^d \frac{x_j^p}{x_1^p}\right) \cdot \frac{1}{y_1^2}\, dy_1 \\
&= \left(\sum_{j=1}^d \frac{x_j^p}{x_1^p}\right)^{\frac{1}{p}-1} \int_0^\infty y_1^{-1/p} \exp\left(-y_1\right) dy_1
\end{aligned}
$$

where we used the substitution $y_1 \mapsto \sum_{j=1}^d \frac{x_j^p}{x_1^p} \frac{1}{y_1}$ in the second-to-last step. Obviously, $-1/p = (1-1/p) - 1$, so the integral evaluates to $\Gamma(1-1/p)$ and our expectation is equal to the factor in front of the integral. That quantity can be written a bit nicer as

$$
\left(\frac{x_1^p}{\sum_{j=1}^d x_j^p} \right)^{1-\frac{1}{p}} = \frac{x_1^{p-1}}{\left(\left\| \boldsymbol{x} \right\|_p^p \right)^{1-1/p}} = \left(\frac{x_1}{\left\| \boldsymbol{x} \right\|_p} \right)^{p-1}.
$$

Of course, the same holds for all the other expectations and the final answer for our derivative at $\boldsymbol{x} > \boldsymbol{0}$ in direction $\boldsymbol{y} \in \mathbb{R}^d$ is

$$
\frac{\partial}{\partial \boldsymbol{y}} \left\| \boldsymbol{x} \right\|_p = \sum_{i=1}^d y_i E\left(Z_i 1_{\{M_{\boldsymbol{x}\boldsymbol{Z}} = i\}} \right) = \sum_{i=1}^d y_i \left(\frac{x_i}{\left\| \boldsymbol{x} \right\|_p} \right)^{p-1}.
$$

Example 5.10. The last thing we want to look at in this section is the differentiation of convex combinations of D-norms. We know that that for two D-norms $\left\| \cdot \right\|_{D_1}$ and $\left\| \cdot \right\|_{D_2}$ with generators $\boldsymbol{Z}^{(1)}$ and $\boldsymbol{Z}^{(2)}$ their convex combination

$$
\left\| \boldsymbol{x} \right\|_{\lambda D_1 + (1-\lambda) D_2} := \left\| \boldsymbol{x} \right\|_{D_1} + (1-\lambda) \left\| \boldsymbol{x} \right\|_{D_2}, \quad \lambda \in [0,1],
$$

is a D-norm itself and can be generated, e.g., by $\boldsymbol{Z}^{(\xi)}$ where ξ is a rv with $P(\xi = 1) = \lambda = 1 - P(\xi = 2)$ that randomly picks between the generators $\boldsymbol{Z}^{(1)}$ and $\boldsymbol{Z}^{(2)}$ and is independent of the two, see Falk (2019, Section 1.4).

The condition of Lemma 5.5 is fulfilled for $\boldsymbol{x} > \boldsymbol{0}$ and $\boldsymbol{y} \in \mathbb{R}^d$, if it is fulfilled for both D-norms separately because we have

$$\begin{aligned}
&P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}^{(\xi)}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}^{(\xi)}} = k\right\}\right)\\
&= \sum_{j=1}^{2} P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}^{(j)}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}^{(j)}} = k,\, \xi = j\right\}\right)\\
&= \sum_{j=1}^{2} P\left(\left\{M_{\boldsymbol{x}\boldsymbol{Z}^{(j)}} = i,\, M_{(\boldsymbol{x}+\varepsilon\boldsymbol{y})\boldsymbol{Z}^{(j)}} = k\right\}\right) P\left(\{\xi = j\}\right)
\end{aligned}$$

which converges to zero for $\varepsilon \to 0$ if both summands converge to zero.

If that is the case, then the directional derivative of the convex combination turns out to be

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_{\lambda D_1+(1-\lambda)D_2} &= \sum_{i=1}^{d} y_i E\left(Z_i 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}^{(\xi)}}=i\right\}}\right)\\
&= \sum_{i=1}^{d} y_i E\left(Z_i 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}^{(\xi)}}=i\right\}}\left(1_{(\xi=1)} + 1_{(\xi=2)}\right)\right)\\
&= \sum_{i=1}^{d}\sum_{j=1}^{2} y_i E\left(Z_i 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}^{(j)}}=i\right\}} 1_{(\xi=j)}\right)\\
&= \sum_{j=1}^{2} E\left(1_{(\xi=j)}\right) \sum_{i=1}^{d} y_i E\left(Z_i 1_{\left\{M_{\boldsymbol{x}\boldsymbol{Z}^{(j)}}=i\right\}}\right)\\
&= \sum_{j=1}^{2} P\left(\xi = j\right) \frac{\partial}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_{D_j}\\
&= \lambda \frac{\partial}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_{D_1} + (1-\lambda)\frac{\partial}{\partial \boldsymbol{y}} \|\boldsymbol{x}\|_{D_2}
\end{aligned}$$

where we used the independence of $\boldsymbol{Z}^{(1)}$, $\boldsymbol{Z}^{(2)}$ and ξ. The result is exactly the convex combination of the two derivatives and the same holds for the one-sided derivatives in Lemma 5.4.

5.2 Multivariate Spacings

By the inclusion-exclusion-formula (see Billingsley (1995), p. 24, equation 2.9) we have for arbitrary sets A, $B = \bigcap_{i=1}^d B_i$ from the underlying sigma-algebra

$$\begin{aligned} P(A \cap B) &= P(A) - P\left(A \cap B^{\complement}\right) = P(A) - P\left(\bigcup_{i=1}^{d} A \cap B_i^{\complement}\right) \\ &= P(A) - \sum_{\emptyset \neq I \subseteq \{1,\dots,d\}} (-1)^{|I|-1} P\left(A \cap \bigcap_{i\in I} B_i^{\complement}\right) \\ &= \sum_{I \subseteq \{1,\dots,d\}} (-1)^{|I|} P\left(A \cap \bigcap_{i\in I} B_i^{\complement}\right). \end{aligned}$$

This allows us to express the (joint) survival function $S(\boldsymbol{x}) = P(\boldsymbol{X} > \boldsymbol{x})$ in terms of the distribution function $F(\boldsymbol{x}) = P(\boldsymbol{X} \leq \boldsymbol{x})$ for a d-variate random vector $\boldsymbol{X}$ and $\boldsymbol{x} \in \mathbb{R}^d$ according to the following formula:

$$\begin{aligned} P(\boldsymbol{X} > \boldsymbol{x}) &= 1 - P\left(\bigcup_{i=1}^{d} \{X_i \leq x_i\}\right) \\ &= 1 - \sum_{\emptyset \neq I \subseteq \{1,\dots,d\}} (-1)^{|I|-1} P\left(\bigcap_{i\in I} \{X_i \leq x_i\}\right) \\ &= \sum_{I \subseteq \{1,\dots,d\}} (-1)^{|I|} P\left(\bigcap_{i\in I} \{X_i \leq x_i\}\right) \\ &= \sum_{I \subseteq \{1,\dots,d\}} (-1)^{|I|} P\left(\boldsymbol{e}_I \boldsymbol{X} \leq \boldsymbol{e}_I \boldsymbol{x}\right) \end{aligned}$$

where the last equality holds by putting $\boldsymbol{e}_I$ as the vector that has ones in the entries specified by the set I, i.e., $\boldsymbol{e}_I = \sum_{i\in I} \boldsymbol{e}_i$, and entry-wise multiplication of vectors. If the underlying distribution is continuous at all points $\boldsymbol{x}_I := \boldsymbol{e}_I \boldsymbol{x}$, then it does not matter whether we put the inequalities with or without the equals.

In the univariate case, the stochastic behavior of the spacings $U_{i:n} - U_{i-1:n}$, $1 \leq i \leq n$, with $U_{0:n} := 0$, $U_{n+1:n} := 1$, follows from the representation

$$(U_{i:n})_{i=1}^n =_D \left(\frac{\sum_{j=1}^{i} \eta_j}{\sum_{i=1}^{n+1} \eta_j} \right)_{i=1}^{n},$$

where the $\eta_1, \dots, \eta_{n+1}$ are independent copies of a rv η that follows the standard exponential distribution, see Reiss (1989, Corollary 1.6.9). This gives us

$$(U_{i:n} - U_{i-1:n})_{i=1}^{n+1} =_D \left(\frac{\eta_i}{\sum_{i=1}^{n+1} \eta_j} \right)_{i=1}^{n+1}$$

and multiplication with $n+1$ yields

$$(n+1)(U_{i:n} - U_{i-1:n})_{i=1}^{n+1} =_D \left(\frac{\eta_i}{\sum_{i=1}^{n+1} \eta_j/(n+1)} \right)_{i=1}^{n} \approx (-\eta_i)_{i=1}^{n+1}$$

by the law of large numbers. But what happens in the multivariate case?

For simplicity, we suppose that $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \dots$ are independent copies of a random vector $\boldsymbol{U}$ that follows a d-variate GPC, i.e., there exists a D-norm such that

$$P(\boldsymbol{U} \le \boldsymbol{u}) = 1 - \|\boldsymbol{1} - \boldsymbol{u}\|_D$$

for all $\boldsymbol{u} \in [u_0, 1]^d$ with some lower bound $u_0 < 1$. As $U_{m:n}$ is defined as one for $m > n$ (in general: as the upper endpoint of the underlying distribution), the final spacing $\boldsymbol{U}_{n+1:n} - \boldsymbol{U}_{n:n} = \boldsymbol{1} - \boldsymbol{U}_{n:n}$ is no problem: for $\boldsymbol{x} \ge \boldsymbol{0}$ and large enough n such that we can use the representation for $P(\boldsymbol{U} \le \boldsymbol{u})$ from above, we have

$$\begin{aligned}
P\left((n+1)\left(\boldsymbol{1} - \boldsymbol{U}_{n:n}\right) \le \boldsymbol{x}\right) &= P\left(\boldsymbol{U}_{n:n} \ge \boldsymbol{1} - \frac{\boldsymbol{x}}{n+1}\right) \\
&= P\left(\boldsymbol{U}_{n:n} > \boldsymbol{1} - \frac{\boldsymbol{x}}{n+1}\right) \\
&= \sum_{I \subseteq \{1,\dots,d\}} (-1)^{|I|} P\left(\bigcap_{i \in I} \left\{U_{n:n,i} \le 1 - \frac{x_i}{n+1}\right\}\right) \\
&= \sum_{I \subseteq \{1,\dots,d\}} (-1)^{|I|} P\left(\bigcap_{i \in I} \left\{U_{n:n,i} \le 1 - \frac{x_i}{n+1}\right\} \cap \bigcap_{i \notin I} \{U_{n:n,i} \le 1\}\right) \\
&= \sum_{I \subseteq \{1,\dots,d\}} (-1)^{|I|} P\left(\boldsymbol{U}_{n:n} \le \boldsymbol{1} - \sum_{i \in I} \frac{x_i}{n+1} \boldsymbol{e}_i\right)
\end{aligned}$$

$$= \sum_{I\subseteq\{1,\dots,d\}} (-1)^{|I|} P\left(\boldsymbol{U}_{n:n} \le \boldsymbol{1} - \frac{\boldsymbol{x}_I}{n+1}\right)$$
$$= \sum_{I\subseteq\{1,\dots,d\}} (-1)^{|I|} P\left(\boldsymbol{U} \le \boldsymbol{1} - \frac{\boldsymbol{x}_I}{n+1}\right)^n$$
$$= \sum_{I\subseteq\{1,\dots,d\}} (-1)^{|I|} \left(1 - \frac{\|\boldsymbol{x}_I\|_D}{n+1}\right)^n$$
$$\to_{n\to\infty} \sum_{I\subseteq\{1,\dots,d\}} (-1)^{|I|} \exp\left(-\|\boldsymbol{x}_I\|_D\right). \tag{5.1}$$

The first spacing $\boldsymbol{U}_{1:n} - \boldsymbol{U}_{0:n} = \boldsymbol{U}_{1:n}$ is, in general, not related to the D-norm. To see this, we compute in a similar way

$$P\left((n+1)\boldsymbol{U}_{1:n} \le \boldsymbol{x}\right) = P\left(\boldsymbol{U}_{1:n} \le \frac{\boldsymbol{x}}{n+1}\right) = P\left(\bigcap_{i=1}^{d}\left\{\boldsymbol{U}_{1:n,i} \le \frac{\boldsymbol{x}}{n+1}\right\}\right)$$
$$= \sum_{I\subseteq\{1,\dots,d\}} (-1)^{|I|} P\left(\bigcap_{i\in I}\left\{\boldsymbol{U}_{1:n,i} > \frac{\boldsymbol{x}}{n+1}\right\}\right)$$
$$= \sum_{I\subseteq\{1,\dots,d\}} (-1)^{|I|} P\left(\boldsymbol{U} > \frac{\boldsymbol{x}_I}{n+1}\right)^n$$

and as $\frac{\boldsymbol{x}_I}{n+1} \not\ge \boldsymbol{u}_0$ for sufficiently large n (unless $(\boldsymbol{u}_0)_I = \boldsymbol{0}$), it is clear that we leave the upper tail where the dependence of our GPC is determined by the D-norm.

To deduce a result for the second-to-last spacing, we unfortunately need to limit ourselves to the bivariate case as the higher dimensions come with a few problems. Therefore, we consider the sequence (U_1, V_1), $(U_1, V_1), \dots$ of iid copies of (U, V) that follows a bivariate GPC as in the general case above and have for $\boldsymbol{x} > \boldsymbol{0} \in \mathbb{R}^2$

$$\bar{H}_n(x_1, x_2) = P\left(U_{n:n} - \frac{x_1}{n} > U_{n-1:n},\, V_{n:n} - \frac{x_2}{n} > V_{n-1:n}\right)$$
$$= \sum_{1\le i,j\le n} P\left(U_{n:n} - \frac{x_1}{n} > U_{n-1:n},\, V_{n:n} - \frac{x_2}{n} > V_{n-1:n},\, U_i = U_{n:n},\, V_j = V_{n:n}\right)$$
$$= \sum_{1\le i,j\le n} P\left(U_i - \frac{x_1}{n} > U_{n-1:n},\, V_j - \frac{x_2}{n} > V_{n-1:n},\, U_i = U_{n:n},\, V_j = V_{n:n}\right)$$

$$
\begin{aligned}
&= \sum_{1\le i\le n} P\left(U_i - \frac{x_1}{n} > U_{n-1:n},\, V_i - \frac{x_2}{n} > V_{n-1:n}\right) \\
&\qquad + \sum_{\substack{1\le i,j\le n\\ i\ne j}} P\left(U_i - \frac{x_1}{n} > U_{n-1:n},\, V_j - \frac{x_2}{n} > V_{n-1:n}\right) \\
&= \sum_{1\le i\le n} P\left(U_1 - \frac{x_1}{n} > U_{n-1:n},\, V_1 - \frac{x_2}{n} > V_{n-1:n}\right) \\
&\qquad + \sum_{\substack{1\le i,j\le n\\ i\ne j}} P\left(U_1 - \frac{x_1}{n} > U_{n-1:n},\, V_2 - \frac{x_2}{n} > V_{n-1:n}\right) \\
&= nP\left(U_1 - \frac{x_1}{n} > \max_{1<i\le n} U_i,\, V_1 - \frac{x_2}{n} > \max_{1<i\le n} V_i\right) \\
&\qquad + n(n-1)P\Bigg(U_1 - \frac{x_1}{n} > U_2,\, U_1 - \frac{x_1}{n} > \max_{2<i\le n} U_i, \\
&\qquad\qquad\qquad V_2 - \frac{x_2}{n} > V_1,\, V_2 - \frac{x_2}{n} > \max_{2<i\le n} V_i\Bigg) \\
&=: nA_n + n(n-1)B_n
\end{aligned}
$$

Our next steps are to figure out the behavior of A_n and B_n separately. For both, we will write the probability as an integral, and split that up into a part where our GPC model holds and a remainder term that turns out to be asymptotically negligible.

Specifically, put $\boldsymbol{M}_{n,k} = n\left(\max_{1\le j\le n-k} U_j - 1,\, \max_{1\le j\le n-k} V_j - 1\right)$. Independent of the value for k, we know that $\boldsymbol{M}_{n,k} \to_D \boldsymbol{\eta}$ where $\boldsymbol{\eta}$ has the df $G(\tilde{\boldsymbol{x}}) = \exp(-\|\tilde{\boldsymbol{x}}\|_D)$ for $\tilde{\boldsymbol{x}} \le \boldsymbol{0}$. The D-norm is the one from the GPC. As $\max_{i<j\le n} U_j$ and $\max_{1\le j\le n-i} U_j$ are equal in distribution, conditioning on $\boldsymbol{M}_{n,1} = \boldsymbol{z}$ yields

$$
\begin{aligned}
A_n &= P\left(U_1 > \left(\max_{1<i\le n} U_i - 1\right) + 1 + \frac{x_1}{n},\, V_1 > \left(\max_{1<i\le n} V_i - 1\right) + 1 + \frac{x_2}{n}\right) \\
&= \int_{(-\infty,0]\times(-\infty,0]} P\left(U > \frac{z_1}{n} + 1 + \frac{x_1}{n},\, V > \frac{z_2}{n} + 1 + \frac{x_2}{n}\right)(P * \boldsymbol{M}_{n,1})(d\boldsymbol{z}) \\
&= \int_{[-n,-\boldsymbol{x}]} P\left(U > \frac{z_1 + x_1}{n} + 1,\, V > \frac{z_2 + x_2}{n} + 1\right)(P * \boldsymbol{M}_{n,1})(d\boldsymbol{z})
\end{aligned}
$$

as $\boldsymbol{M}_{n,1}$ always realizes in $[-n, \boldsymbol{0}]$ and the probability in the integral is zero if $\boldsymbol{z} \not< \boldsymbol{x}$. The remark after lemma 3.1.13 in Falk (2019) states that we have for large enough n such that $\frac{\boldsymbol{z}+\boldsymbol{x}}{n} + \boldsymbol{1} \in [\boldsymbol{u}_0, \boldsymbol{1}] \iff \boldsymbol{z}+\boldsymbol{x} \in [n(\boldsymbol{u}_0-\boldsymbol{1}), \boldsymbol{0}] \iff \boldsymbol{z} \in [n(\boldsymbol{u}_0-\boldsymbol{1})-\boldsymbol{x}, -\boldsymbol{x}]$ the identity

$$P\left(U > \frac{z_1+x_1}{n}+1,\, V > \frac{z_2+x_2}{n}+1\right) = \wr\, \boldsymbol{1} - \frac{\boldsymbol{z}+\boldsymbol{x}}{n} - \boldsymbol{1} \,\wr_D = \frac{1}{n} \wr \boldsymbol{z}+\boldsymbol{x} \wr_D$$

and therefore

$$\begin{aligned} nA_n &= n\int_{[n(\boldsymbol{u}_0-\boldsymbol{1})-\boldsymbol{x},-\boldsymbol{x}]} \wr\, \frac{\boldsymbol{z}+\boldsymbol{x}}{n} \,\wr_D \left(P * \boldsymbol{M}_{n,1}\right)(d\boldsymbol{z}) + nR_n \\ &= \int_{[n(\boldsymbol{u}_0-\boldsymbol{1})-\boldsymbol{x},-\boldsymbol{x}]} \wr\, \boldsymbol{z}+\boldsymbol{x} \,\wr_D \left(P * \boldsymbol{M}_{n,1}\right)(d\boldsymbol{z}) + nR_n \\ &\to_{n\to\infty} \int_{(-\infty,-\boldsymbol{x}]} \wr\, \boldsymbol{z}+\boldsymbol{x} \,\wr_D \left(P * \boldsymbol{\eta}\right)(d\boldsymbol{z}) = E\left(\wr\, \boldsymbol{\eta}+\boldsymbol{x} \,\wr_D 1_{(\boldsymbol{\eta}+\boldsymbol{x}\le\boldsymbol{0})}\right) \end{aligned}$$

if we can show that $nR_n \to_{n\to\infty} 0$. This is the case because

$$\begin{aligned} &\left|\int_{[-n,-\boldsymbol{x}]\setminus[n(\boldsymbol{u}_0-\boldsymbol{1})-\boldsymbol{x},-\boldsymbol{x}]} P\left(U > \frac{z_1+x_1}{n}+1,\, V > \frac{z_2+x_2}{n}+1\right)\left(P * \boldsymbol{M}_{n,1}\right)(d\boldsymbol{z})\right| \\ &\le \int_{[-n,-\boldsymbol{x}]\setminus[n(\boldsymbol{u}_0-\boldsymbol{1})-\boldsymbol{x},-\boldsymbol{x}]} 1\left(P * \boldsymbol{M}_{n,1}\right)(d\boldsymbol{z}) \\ &= P\left(\boldsymbol{M}_{n,1} \in [-n,-\boldsymbol{x}] \setminus [n(\boldsymbol{u}_0-\boldsymbol{1})-\boldsymbol{x},-\boldsymbol{x}]\right) \\ &\le P\left(\boldsymbol{M}_{n,1} \in [-n,\boldsymbol{0}] \setminus [n(\boldsymbol{u}_0-\boldsymbol{1})-\boldsymbol{x},\boldsymbol{0}]\right) \\ &\le P\left(\frac{\boldsymbol{M}_{n,1}}{n} + \boldsymbol{1} \in [\boldsymbol{0},\boldsymbol{1}] \setminus [\boldsymbol{u}_0 - \boldsymbol{x}/n, \boldsymbol{1}]\right) \\ &= P\left(\max_{1\le j\le n-1} U_j < u_{0,1} - \frac{x_1}{n} \text{ or } \max_{1\le j\le n-1} V_j < u_{0,1} - \frac{x_2}{n}\right) \\ &\le P\left(\max_{1\le j\le n-1} U_j < u_{0,2}\right) + P\left(\max_{1\le j\le n-1} V_j < u_{0,1}\right) \\ &\le P\left(U < u_{0,1}\right)^{n-1} + P\left(V < u_{0,2}\right)^{n-1} \\ &\le \left[u_{0,1}\right]^{n-1} + \left[u_{0,2}\right]^{n-1} \end{aligned}$$

and $n \cdot \left[u_{0,i}\right]^{n-1} \to_{n\to\infty} 0$ for $i=1$ and $i=2$ as $\boldsymbol{u}_0 < \boldsymbol{1}$.

To find an expression for B_n, we will employ similar arguments, but have to establish a certain conditional probability first. We use the fact that (U_1, V_1) and (U_2, V_2) are independent to compute for $u_{0,1} < a, u < 1$ and $u_{0,2} < b, v < 1$

$$\begin{aligned}
&P\left(U_2 \le a,\, V_1 \le b \mid U_1 = u,\, V_2 = v\right)\\
&\quad = \lim_{\varepsilon\downarrow 0} \frac{P\left(U_2 \le a,\, V_1 \le b,\, U_1 \in [u, u+\varepsilon],\, V_2 \in [v, v+\varepsilon]\right)}{P\left(U_1 \in [u, u+\varepsilon],\, V_2 \in [v, v+\varepsilon]\right)}\\
&\quad = \lim_{\varepsilon\downarrow 0} \frac{P\left(V_1 \le b,\, U_1 \in [u, u+\varepsilon]\right)}{P\left(U_1 \in [u, u+\varepsilon]\right)} \times \frac{P\left(U_2 \le a,\ V_2 \in [v, v+\varepsilon]\right)}{P\left(V_2 \in [v, v+\varepsilon]\right)}\\
&\quad = \lim_{\varepsilon\downarrow 0} \frac{P\left(V_1 \le b,\, U_1 \le u+\varepsilon\right) - P\left(V_1 \le b,\, U_1 < u\right)}{\varepsilon}\\
&\qquad\qquad \times \frac{P\left(U_2 \le a,\ V_2 \le v+\varepsilon\right) - P\left(U_2 \le a,\ V_2 < v\right)}{\varepsilon}\\
&\quad = \lim_{\varepsilon\downarrow 0} \frac{\|(1-(u+\varepsilon), 1-b)\|_D - \|(1-b, 1-u)\|_D}{\varepsilon}\\
&\qquad\qquad \times \frac{\|(1-a, 1-(v+\varepsilon))\|_D - \|(1-a, 1-v)\|_D}{\varepsilon}\\
&\quad = \frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} 1-u \\ 1-b \end{pmatrix} \right\|_D \frac{\partial^+}{\partial_2} \left\| \begin{pmatrix} 1-a \\ 1-v \end{pmatrix} \right\|_D .
\end{aligned}$$

Similar to what we did with A_n, we first condition on $\boldsymbol{M}_{n,2} = \boldsymbol{z}$ to find that

$$\begin{aligned}
B_n = P\Bigg(&U_1 - \frac{x_1}{n} > U_2,\, U_1 - \frac{x_1}{n} - 1 > \max_{2<i\le n} U_i - 1,\\
&V_2 - \frac{x_2}{n} > V_1,\, V_2 - \frac{x_2}{n} - 1 > \max_{2<i\le n} V_i - 1\Bigg)\\
= \int_{[-n,-\boldsymbol{x}]} P\Bigg(&U_1 - \frac{x_1}{n} > U_2,\, U_1 > \frac{z_1}{n} + 1 + \frac{x_1}{n},\\
&V_2 - \frac{x_2}{n} > V_1,\, V_2 > \frac{z_2}{n} + 1 + \frac{x_2}{n}\Bigg) (P * \boldsymbol{M}_{n,2})(d\boldsymbol{z}).
\end{aligned}$$

To compute the integrand by conditioning on $U_1 = u$ and $V_2 = v$, we have to make sure that u and v as well as $u - x_1/n$ and $v - x_2/n$ stay in the correct boundaries for our formula above. This will be the case for $u > u_{0,1}$, $v > u_{0,2}$ and large enough n, therefore we can cover the part of the integration area where $\boldsymbol{z} \in (n(\boldsymbol{u}_0 - \boldsymbol{1}), -\boldsymbol{x})$. There, the probability is equal to

$$\int_{1+\frac{z_1+x_1}{n}}^{1} \int_{1+\frac{z_2+x_2}{n}}^{1} P\left(U_2 < u - \frac{x_1}{n},\, V_1 < v - \frac{x_2}{n} \ \middle|\ U_1 = u,\, V_2 = v\right) dv\, du$$

$$= \int_{1+\frac{z_1+x_1}{n}}^{1} \int_{1+\frac{z_2+x_2}{n}}^{1} \frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} 1-u \\ 1-\left(v-\frac{x_2}{n}\right) \end{pmatrix} \right\|_D \frac{\partial^+}{\partial_2} \left\| \begin{pmatrix} 1-\left(u-\frac{x_1}{n}\right) \\ 1-v \end{pmatrix} \right\|_D dv\, du$$
$$= \frac{1}{n^2} \int_0^{-(z_1+x_1)} \int_0^{-(z_2+x_2)} \frac{\partial^+}{\partial_1} \left\| \frac{1}{n} \begin{pmatrix} u \\ v+x_2 \end{pmatrix} \right\|_D \frac{\partial^+}{\partial_2} \left\| \frac{1}{n} \begin{pmatrix} u+x_1 \\ v \end{pmatrix} \right\|_D dv\, du$$

where we obtained the last equality by substituting both $u \to 1 - u/n$ and $v \to 1 - v/n$. As a final piece of notation, we define

$$D(u, v, x_1, x_2) := \frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} u \\ v+x_2 \end{pmatrix} \right\|_D \frac{\partial^+}{\partial_2} \left\| \begin{pmatrix} u+x_1 \\ v \end{pmatrix} \right\|_D .$$

and recall that

$$\frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} u/n \\ (v+x_2)/n \end{pmatrix} \right\|_D = \frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} u \\ v+x_2 \end{pmatrix} \right\|_D$$

by Remark 5.6. Hence, it is equal to our last integrand for all values of n.

In total, this leads to

$$\begin{aligned}
n^2 B_n &= \int_{(n(\boldsymbol{u}_0-\mathbf{1}),-\boldsymbol{x})} \int_0^{-(z_1+x_1)} \int_0^{-(z_2+x_2)} D(u, v, x_1, x_2)\, dv\, du\, \left(P * \boldsymbol{M}_{n,2}\right)(d\boldsymbol{z}) \\
&\quad + n^2 \tilde{R}_n \\
&\to_{n\to\infty} \int_{[-\infty,-\boldsymbol{x}]} \int_0^{-(z_1+x_1)} \int_0^{-(z_2+x_2)} D(u, v, x_1, x_2)\, dv\, du\, (P * \boldsymbol{\eta})(d\boldsymbol{z}) \\
&= E\left(\int_0^\infty \int_0^\infty D(u, v, x_1, x_2) 1_{(u \le -\eta_1 - x_1,\, v \le -\eta_2 - x_2)} 1_{(\boldsymbol{\eta} \le -\boldsymbol{x})}\, dv\, du \right) \\
&= E\left(\int_0^\infty \int_0^\infty D(u, v, x_1, x_2) 1_{(\eta_1 \le -u - x_1,\, \eta_2 \le -v - x_2)}\, dv\, du \right) \\
&= \int_0^\infty \int_0^\infty D(u, v, x_1, x_2) E\left(1_{(\eta_1 \le -u - x_1,\, \eta_2 \le -v - x_2)}\right)\, dv\, du \\
&= \int_0^\infty \int_0^\infty D(u, v, x_1, x_2) P\left(\eta_1 \le -u - x_1,\ \eta_2 \le -v - x_2\right)\, dv\, du \\
&= \int_0^\infty \int_0^\infty D(u, v, x_1, x_2) \exp\left(- \left\| \begin{pmatrix} u+x_1 \\ v+x_2 \end{pmatrix} \right\|_D \right)\, dv\, du.
\end{aligned}$$

$\tilde{R}_n$ covers the remaining integration area and $n^2 \tilde{R}_n \to 0$ for $n \to \infty$ by exactly the same arguments as we had $n\, R_n \to 0$ earlier for the remainder term of A_n.

Obviously, $n^2 B_n$ has the same limit as $n(n-1)B_n$ and, putting everything together, we can plug our findings into $\bar{H}_n(x,y) = nA_n + n(n-1)B_n$ and have proven the following bivariate result:

Theorem 5.11. *Consider the bivariate spacings $\boldsymbol{U}_{n:n} - \boldsymbol{U}_{n-1:n}$ for a sequence $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \dots$ of iid copies of $\boldsymbol{U}$ which follows a bivariate GPC, i.e., $\boldsymbol{u}$ has the distribution $C(u,v) = 1 - \|(1-u, 1-v)\|_D$ for $u_0 \le u \le 1$, $v_0 \le v \le 1$. Then the limiting distribution for $n \to \infty$ can be described by its survival probability for $x, y > 0$*

$$\begin{aligned}
&\bar{H}_n(x,y) = P(n(\boldsymbol{U}_{n:n} - \boldsymbol{U}_{n-1:n}) > (x,y)) \\
&\to \bar{H}(x,y) = E\left(\wr\wr\, (\eta_1 + x, \eta_2 + y) \,\wr\wr_D \, 1_{(\eta_1 + x \le 0,\, \eta_2 + y \le 0)}\right) \\
&\qquad + \int_0^\infty \int_0^\infty \frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} u \\ v+y \end{pmatrix} \right\|_D \frac{\partial^+}{\partial_2} \left\| \begin{pmatrix} u+x \\ v \end{pmatrix} \right\|_D \exp\left(- \left\| \begin{pmatrix} u+x \\ v+y \end{pmatrix} \right\|_D \right) dv\, du
\end{aligned}$$

where the bivariate rv $\boldsymbol{\eta}$ has df $G(\tilde{\boldsymbol{x}}) = \exp(-\|\tilde{\boldsymbol{x}}\|_D)$ for $\tilde{\boldsymbol{x}} \le \boldsymbol{0}$.

Example 5.12. We want to compute the limit above for some simple cases first. For $p = 1$ we have $\wr\wr \cdot \wr\wr_1 = 0$ and $\frac{\partial}{\partial_i} \|\cdot\|_D = 1$. Therefore, for all $x, y > 0$

$$\begin{aligned}
\bar{H}(x,y) &= 0 + \int_0^\infty \int_0^\infty 1 \times \exp\left(- \left\| \begin{pmatrix} u+x \\ v+y \end{pmatrix} \right\|_1 \right) dv\, du \\
&= \int_0^\infty \int_0^\infty \exp\left(-\left(u + x + v + y\right)\right) dv\, du \\
&= \exp\left(-\left(x+y\right)\right) \int_0^\infty \int_0^\infty \exp\left(-\left(u+v\right)\right) dv\, du \\
&= \exp\left(-(x+y)\right)
\end{aligned}$$

because $\int_0^\infty \exp(-u)\, du = 1$.

For $p = \infty$, we first observe that the double integral turns out to be zero: the partial derivatives are for general $\boldsymbol{z} \ge \boldsymbol{0} \in \mathbb{R}^d$ and $i = 1, \dots, d$

$$\frac{\partial^+}{\partial_i} \|\boldsymbol{z}\|_\infty = \begin{cases} 1 & \text{if } z_i = \max_{1 \le j \le d} z_j \\ 0 & \text{otherwise,} \end{cases}$$

therefore we have for $x, y > 0$

$$\frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} u \\ v+y \end{pmatrix} \right\|_\infty \frac{\partial^+}{\partial_2} \left\| \begin{pmatrix} u+x \\ v \end{pmatrix} \right\|_\infty = 1_{(u \ge v+y)} \cdot 1_{(v \ge u+x)} = 0$$

because the two conditions cannot be fulfilled simultaneously. Further, we have $\wr\wr (x,y) \wr\wr_\infty = \min(|x|,|y|)$. We can use the fact that $E(X) = \int_0^\infty P(X>t)\,dt$ for any non-negative rv X to compute the expectation for $x,y>0$ and get

$$\begin{aligned}
\bar{H}(x,y) &= E\left(\wr\wr (\eta_1+x,\eta_2+y) \wr\wr_\infty 1_{(\eta_1+x\le 0,\, \eta_2+y\le 0)}\right) + 0\\
&= E\left(\min(|\eta_1+x|,|\eta_2+y|)\, 1_{(\eta_1+x\le 0,\, \eta_2+y\le 0)}\right)\\
&= \int_0^\infty P\left(\min(-\eta_1-x,-\eta_2-y)\, 1_{(\eta_1+x\le 0,\, \eta_2+y\le 0)} > t\right)\, dt\\
&= \int_0^\infty P\left(-\max(\eta_1+x,\eta_2+y) > t, \eta_1+x\le 0,\, \eta_2+y\le 0\right)\, dt\\
&= \int_0^\infty P\left(\max(\eta_1+x,\eta_2+y) < -t\right)\, dt\\
&= \int_0^\infty P\left(\eta_1 < -t-x, \eta_2 < -t-y\right)\, dt\\
&= \int_0^\infty \exp\left(-\left\|(-t-x,-t-y)\right\|_\infty\right)\, dt\\
&= \left[-\exp\left(-\left\|(t+x,t+y)\right\|_\infty\right)\right]_0^\infty\\
&= \exp\left(-\left\|(x,y)\right\|_\infty\right).
\end{aligned}$$

Example 5.13. So we have just the expectation-part of the formula for $\|\cdot\|_\infty$ and just the integral-part for $\|\cdot\|_1$. For their combination in terms of the Marshall–Olkin D-norm $\|\cdot\|_{D_\lambda}$ with parameter $\lambda \in (0,1)$, we end up somewhere in between. It is a special case of a convex combination we mentioned in Example 5.10, so $\|\boldsymbol{z}\|_{D_\lambda} = \lambda \|\boldsymbol{z}\|_\infty + (1-\lambda)\|\boldsymbol{z}\|_1$ for general $\boldsymbol{z} \in \mathbb{R}^d$. This example also tells us that its partial derivatives are

$$\frac{\partial^+}{\partial_i}\|\boldsymbol{z}\|_{D_\lambda} = \lambda \frac{\partial^+}{\partial_i}\|\boldsymbol{z}\|_\infty + (1-\lambda)\frac{\partial^+}{\partial_i}\|\boldsymbol{z}\|_1 = \lambda 1_{\{z_i=\max_{1\le j\le d} z_j\}} + (1-\lambda)$$

and we know that $\wr\wr \boldsymbol{z} \wr\wr_{D_\lambda} = \lambda \min_{1\le i\le d}|z_i|$ from Falk (2019, Example 1.6.4). Similar to our computations above, the expectation is

$$\begin{aligned}
&E\left(\wr\wr (\eta_1+x,\eta_2+y) \wr\wr_{D_\lambda} 1_{(\eta_1+x\le 0,\, \eta_2+y\le 0)}\right)\\
&\quad= E\left(\lambda\min(|\eta_1+x|,|\eta_2+y|)\, 1_{(\eta_1+x\le 0,\, \eta_2+y\le 0)}\right)\\
&\quad= \int_0^\infty P\left(\max(\eta_1+x,\eta_2+y) < -\frac{t}{\lambda}\right)\, dt\\
&\quad= \lambda\int_0^\infty P\left(\eta_1 < -t-x, \eta_2 < -t-y\right)\, dt
\end{aligned}$$

$$= \lambda \left[-\frac{1}{2-\lambda} \exp\left(- \|(t+x, t+y)\|_{D_\lambda} \right) \right]_0^\infty$$
$$= \frac{\lambda}{2-\lambda} \exp\left(- \|(x,y)\|_{D_\lambda} \right)$$

where we used the substitution $t \mapsto \lambda t$ and the fact that $2-\lambda$ is the derivative of $\|(t+x,t+y)\|_{D_\lambda}$ with respect to t.

For the integral-part of our formula, we observe that the product of the partial derivatives is

$$\frac{\partial^+}{\partial_1} \left\| \begin{pmatrix} u \\ v+y \end{pmatrix} \right\|_{D_\lambda} \frac{\partial^+}{\partial_2} \left\| \begin{pmatrix} u+x \\ v \end{pmatrix} \right\|_{D_\lambda}$$
$$= \left[\lambda 1_{\{u \geq v+y\}} + (1-\lambda) \right] \left[\lambda 1_{\{v \geq u+x\}} + (1-\lambda) \right]$$
$$= \lambda^2 1_{\{u \geq v+y\}} 1_{\{v \geq u+x\}} + \lambda(1-\lambda)(1_{\{u \geq v+y\}} + 1_{\{v \geq u+x\}}) + (1-\lambda)^2$$
$$= \lambda(1-\lambda)(1_{\{u \geq v+y\}} + 1_{\{v \geq u+x\}}) + (1-\lambda)^2,$$

therefore we split up the integration accordingly. For the $(1-\lambda)^2$ part, we first look at

$$C := \int_0^\infty \int_0^\infty 1 \times \exp\left(- \left\| \begin{pmatrix} u+x \\ v+y \end{pmatrix} \right\|_{D_\lambda} \right) dv\, du$$
$$= e^{-(1-\lambda)(x+y)} \int_0^\infty \int_0^\infty \exp\left(-\lambda \max\left(u+x, v+y\right) - (1-\lambda)\left(u+v\right) \right) dv\, du.$$

We assume for now that $x \geq y$ and split up the inner integral at $v = u + x - y$ to obtain

$$\int_0^{u+x-y} \exp\left(-\lambda\left(u+x\right) - (1-\lambda)\left(u+v\right) \right) dv$$
$$= \exp\left(-\lambda x - u\right) \int_0^{u+x-y} \exp\left(-\left(1-\lambda\right) v \right) dv$$
$$= \frac{\exp\left(-\lambda x - u\right)}{1-\lambda} \left[1 - \exp\left(-(1-\lambda)\left(u+x-y\right) \right) \right]$$

and

$$\int_{u+x-y}^\infty \exp\left(-\lambda\left(v+y\right) - (1-\lambda)\left(u+v\right) \right) dv$$

$$\begin{aligned}
&= \exp\left(-\lambda y - (1-\lambda)u\right) \int_{u+x-y}^{\infty} \exp\left(-v\right)\, dv \\
&= \exp\left(-\lambda y - (1-\lambda)u - (u+x-y)\right),
\end{aligned}$$

leading to the three parts of the outer integral of

$$\int_0^\infty \frac{\exp\left(-\lambda x - u\right)}{1-\lambda}\, du = \frac{\exp\left(-\lambda x\right)}{1-\lambda} \int_0^\infty \exp\left(-u\right)\, du = \frac{\exp\left(-\lambda x\right)}{1-\lambda},$$

$$\begin{aligned}
-&\int_0^\infty \frac{\exp\left(-\lambda x - u\right)}{1-\lambda} \exp\left(-(1-\lambda)\left(u+x-y\right)\right)\, du \\
&= -\frac{\exp\left(-x + (1-\lambda)y\right)}{1-\lambda} \int_0^\infty \exp\left(-(2-\lambda)u\right)\, du \\
&= -\frac{\exp\left(-x + (1-\lambda)y\right)}{(1-\lambda)(2-\lambda)}
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^\infty \exp\left(-\lambda y - (1-\lambda)u - (u+x-y)\right)\, du \\
&= \exp\left(-x + (1-\lambda)y\right) \int_0^\infty \exp\left(-(2-\lambda)u\right)\, du \\
&= \frac{\exp\left(-x + (1-\lambda)y\right)}{2-\lambda}.
\end{aligned}$$

Combining those yields

$$\begin{aligned}
C &= e^{-(1-\lambda)(x+y)} \left[\frac{\exp\left(-\lambda x\right)}{1-\lambda} - \frac{\exp\left(-x+(1-\lambda)y\right)}{(1-\lambda)(2-\lambda)} + \frac{\exp\left(-x+(1-\lambda)y\right)}{2-\lambda}\right] \\
&= e^{-\lambda x-(1-\lambda)(x+y)} \left[\frac{1}{1-\lambda} - \frac{\lambda \exp\left(-x+\lambda x+(1-\lambda)y\right)}{(1-\lambda)(2-\lambda)}\right] \\
&= \exp\left(-\left\|(x,y)\right\|_{D_\lambda}\right) \left[\frac{1}{1-\lambda} - \frac{\lambda \exp\left(-(1-\lambda)(x-y)\right)}{(1-\lambda)(2-\lambda)}\right]
\end{aligned}$$

for the double integral without the indicator function and for $x \geq y$. Clearly, the result for $y \geq x$ is the almost the same, just with switched roles of x and y. We can unify the two cases by putting $|x-y|$ instead of $x-y$ in the exponential function.

Computing the double integrals with the indicator function is quite a bit easier as we do not have to split up the integration:

$$\begin{aligned}
B &:= \int_0^\infty \int_0^\infty 1_{\{v \geq u+x\}} \times \exp\left(-\left\|\begin{pmatrix} u+x \\ v+y \end{pmatrix}\right\|_{D_\lambda}\right) dv\, du \\
&= e^{-(1-\lambda)(x+y)} \int_0^\infty \int_{u+x}^\infty \exp\left(-\lambda \max\left(u+x, v+y\right) - (1-\lambda)\left(u+v\right)\right) dv\, du \\
&= e^{-(1-\lambda)(x+y)-\lambda y} \int_0^\infty \exp\left(-\left(1-\lambda\right)u\right) \int_{u+x}^\infty \exp\left(-v\right)\, dv\, du \\
&= e^{-(1-\lambda)x-y} \int_0^\infty \exp\left(-\left(1-\lambda\right)u\right) \exp\left(-u-x\right)\, du \\
&= e^{-(2-\lambda)x-y} \int_0^\infty \exp\left(-\left(2-\lambda\right)u\right)\, du \\
&= \frac{\exp\left(-\left(2-\lambda\right)x-y\right)}{2-\lambda} \\
&= \exp\left(-\left\|(x,y)\right\|_{D_\lambda}\right) \exp\left(-x-\lambda y+\lambda \max(x,y)\right)/(2-\lambda)
\end{aligned}$$

and in complete analogy, by switching the integration order,

$$\begin{aligned}
A &:= \int_0^\infty \int_0^\infty 1_{\{u \geq v+y\}} \times \exp\left(-\left\|\begin{pmatrix} u+x \\ v+y \end{pmatrix}\right\|_{D_\lambda}\right) dv\, du \\
&= e^{-(1-\lambda)(x+y)} \int_0^\infty \int_{v+y}^\infty \exp\left(-\lambda \max\left(u+x, v+y\right) - (1-\lambda)\left(u+v\right)\right) du\, dv \\
&= \exp\left(-\left\|(x,y)\right\|_{D_\lambda}\right) \exp\left(-y-\lambda x+\lambda \max(x,y)\right)/(2-\lambda).
\end{aligned}$$

Finally, we obtain our result with

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \frac{\partial^+}{\partial_1}\left\|\begin{pmatrix} u \\ v+y \end{pmatrix}\right\|_{D_\lambda} \frac{\partial^+}{\partial_2}\left\|\begin{pmatrix} u+x \\ v \end{pmatrix}\right\|_{D_\lambda} \exp\left(-\left\|\begin{pmatrix} u+x \\ v+y \end{pmatrix}\right\|_{D_\lambda}\right) dv\, du \\
&\quad = \lambda(1-\lambda)(A+B) + (1-\lambda)^2 C,
\end{aligned}$$

as

$$H(x,y) = \exp\left(-\left\|(x,y)\right\|_{D_\lambda}\right)\Bigg[\frac{\lambda}{2-\lambda}$$

$$+\frac{\lambda(1-\lambda)}{2-\lambda}\left(\exp\left(-x-\lambda y+\lambda\max(x,y)\right)+\exp\left(-y-\lambda x+\lambda\max(x,y)\right)\right)$$
$$\left.+(1-\lambda)-\frac{\lambda(1-\lambda)}{2-\lambda}\exp\left(-(1-\lambda)\left|x-y\right|\right)\right]. \tag{5.2}$$

Notice that the special cases $H(x,y) = \exp\left(-\left\|(x,y)\right\|_1\right)$ for $\lambda = 0$ and $H(x,y) = \exp\left(-\left\|(x,y)\right\|_\infty\right)$ for $\lambda = 1$ are included here. Some contour plots and further comments are part of the next and final example.

Example 5.14. For further p-norms, unfortunately, the formula seems very hard to evaluate analytically, even for specific values of p. However, numerical evaluation is possible. Computing precision is something to keep an eye out for if p is large or close to 1, but can be dealt with and is no issue for intermediate values of p. The expectation part can be evaluated by noting that the dual D-norm is always $\wr\wr\,(x,y)\,\wr\wr_D = x+y-\left\|(x,y)\right\|_D$ in the bivariate case and using the bivariate density of $\boldsymbol{\eta}$ which exists for $p\in(1,\infty)$.

Figure 5.1 shows the result for some selected values of p and also for several parameters λ of the Marshall–Olkin D-norm. We see that we have a smooth transition from the case of total independence to total dependence as p moves from 1 to ∞ and as λ takes its way from 0 to 1.

The values on the axes are not affected by the parameters. This is expected since they correspond to the marginal probabilities which are independent of the dependence modeled by λ or p. The symmetry is no surprise, either.

In the middle panel, we plotted the relative deviation between $H_n(x,y)$ from the Marshall–Olkin case (Equation 5.2) and $\exp\left(-\left\|(x,y)\right\|_{D_\lambda}\right)$ against the parameter λ. To be specific, the black lines are the maximum and minimum of $H_n(x,y)/\exp\left(-\left\|(x,y)\right\|_{D_\lambda}\right)$ over all $x,y\geq 0$. We see that the numerator is always smaller than or equal to the denominator and the maximum of 1 can obviously be achieved on the boundary whereas the minimum is attained away from the boundaries and, in particular, the origin. To get a better impression, the dashed red line is the maximum bounded away from the axes ($x,y>1$ to be exact) and the dotted red line is the minimum close to the origin ($x,y<1$). We know from Equation 5.1 that $\exp\left(-\left\|(x,y)\right\|_{D_\lambda}\right)$ is the asymptotic survival function of the last spacing. Therefore, what we observe is the close connection between the survival functions of the last and second-to-last spacing under the Marshall–Olkin model, especially close to the origin.

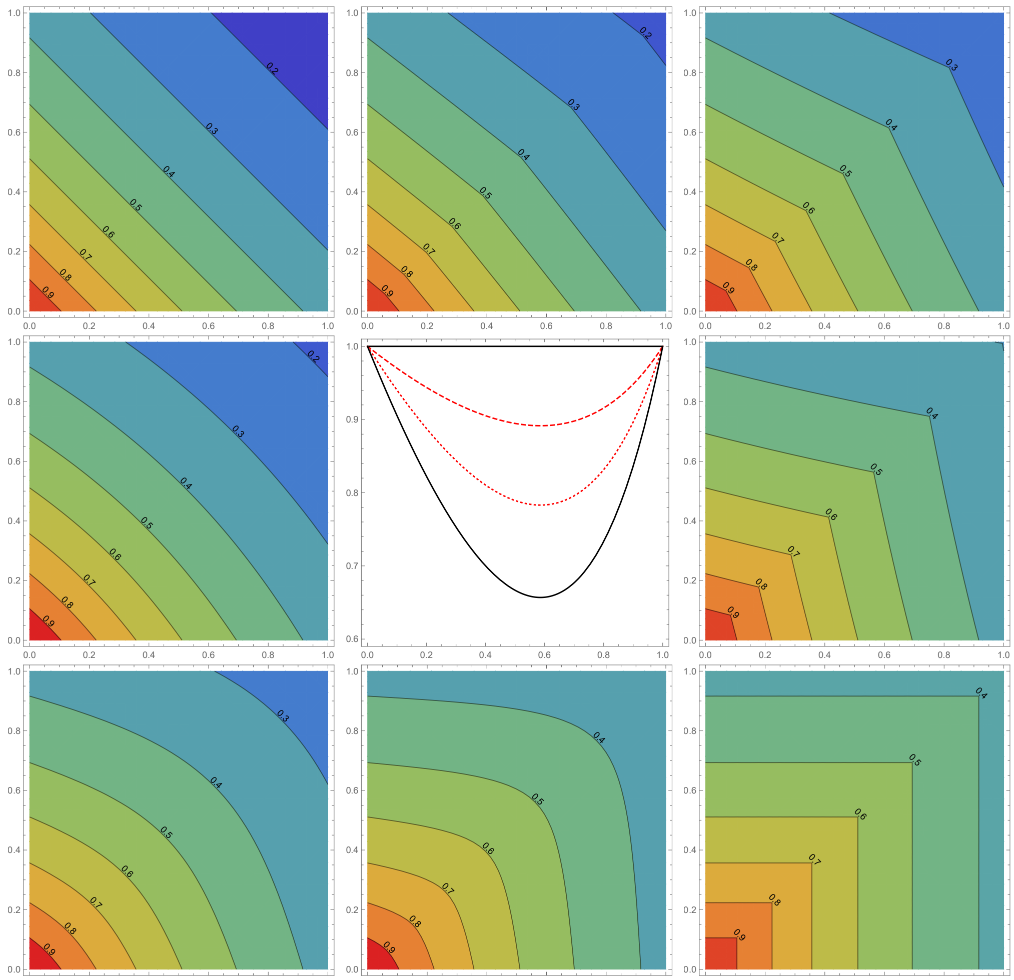

Figure 5.1: Top left panel is the bivariate survival function for the second to last spacing for $\|\cdot\|_1$ as the underlying D-norm. Bottom right panel is for $\|\cdot\|_\infty$, clockwise is the transition through the Marshall–Olkin norms with parameters $\lambda = 0.5, 0.75$ and 0.9 and counterclockwise the transition through the p-norms with parameters $p = 2, 4$ and 10. The middle panel shows the relative maximum and minimum deviation from the survival function to just $\exp\left(-\|\cdot\|_{D_\lambda}\right)$ for λ ranging from 0 to 1.

Chapter 6

Summary and Outlook

After introducing a bit of background, we looked into bivariate order statistics in Chapter 2. We found that the two components become asymptotically independent for certain combinations of indices and used an explicit formula for a conditional probability from Lemma 2.2 to deduce this result. It might be interesting to see if that formula can help in other situations as well.

In Chapter 3, we detected conditional independence in Archimedean copula models and we already started to extend this finding to the more general class of Archimax copulas. This was successful for some elements of the larger family, but more work is needed to check whether the result holds for arbitrary Archimax models as well. Furthermore, the idea of a simulation as sketched in Remark 3.8 deserves a closer look.

Generalized Pareto models were the subject of Chapter 4. We were able to derive an estimator for the exceedance probability over a given threshold. We could provide the corresponding confidence intervals and give a criterion to check the model assumption. The procedure was first backed up by means of a simulation study and then applied to real pollution data in Milan, Italy. Further work could go into the optimal choice of the parameter t_0 and extended comparison with existing algorithms.

We went on to compute derivatives of D-norms in the last chapter. We saw a way to compute the directional derivative as a suitable expectation in terms of its generator. Considering higher order derivatives would be a natural extension. A closer look at the condition that guaranteed the existence of the derivative could be interesting as well: our examples showed that it is not always a necessary condition. Finally, we concluded with some analysis of multivariate spacings. The last spacing turned out to be very simple and we were able to give an explicit formula for the survival function of the second-to-last spacing, but we had to limit ourselves to the bivariate case. It would be desirable to have the result for general dimension $d > 2$, though the formula might be quite complex there. The behavior of general multivariate spacings remains an open problem as well.

Bibliography

Agresti, A., and Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* **52**, 119–126. doi:10.2307/2685469.

Aulbach, S., Bayer, V., and Falk, M. (2012). A multivariate piecing-together approach with an application to operational loss data. *Bernoulli* **18**, 455–475. doi:10.3150/10-BEJ343.

Aulbach, S., Falk, M., and Fuller, T. (2018). Testing for a δ-neighborhood of a generalized Pareto copula. *Ann. Inst. Stat. Math.;* to appear. doi:10.1007/s10463-018-0657-x.

Aulbach, S., Falk, M., Hofmann, M., and Zott, M. (2015). Max-stable processes and the functional D-norm revisited. *Extremes* **18**, 191–212. doi:10.1007/s10687-014-0210-0.

Balkema, A. A., and Resnick, S. I. (1977). Max-infinite divisibility. *J. Appl. Probab.* **14**, 309–319. doi:10.2307/3213001.

Beirlant, J., Escobar-Bach, M., Goegebeur, Y., and Guillou, A. (2016). Bias-corrected estimation of stable tail dependence function. *Journal of Multivariate Analysis* **143**, 453–466.

Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics, Anniversary ed. Wiley, New York.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science* 101–117.

Brunekreef, B., and Holgate, S. T. (2002). Air pollution and health. *The lancet* **360**, 1233–1242.

Bücher, A. (2012). A note on nonparametric estimation of bivariate tail dependence. https://www.ruhr-uni-bochum.de/imperia/md/content/mathematik3/publications/taildep_buecher.pdf.

Buishand, T. A., de Haan, L., and Zhou, C. (2008). On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.* **2**, 624–642. doi:10.1214/08-AOAS159.

Charpentier, A., Fougères, A., Genest, C., and Nešlehová, J. (2014). Multivariate Archimax copulas. *J. Multivariate Anal.* **126**, 118–136. doi:10.1016/j.jmva.2013.12.013.

Charpentier, A., and Segers, J. (2009). Tails of multivariate Archimedean copulas. *J. Multivariate Anal.* **100**, 1521–1537. doi:10.1016/j.jmva.2008.12.015.

Cheng, S., de Haan, L., and Yang, J. (1997). Asymptotic distributions of multivariate intermediate order statistics. *Theory Probab. Appl.* **41**, 646–656. doi:10.1137/S0040585X97975733.

Clapp, L. J., and Jenkin, M. E. (2001). Analysis of the relationship between ambient levels of o3, no2 and no as a function of nox in the uk. *Atmospheric Environment* **35**, 6391–6405.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*, vol. 208. Springer.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157–172.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.), 117–130. D. Reidel, Dordrecht. doi:10.1007/978-94-017-3069-3_9.

Draisma, G., Drees, H., Ferreira, A., De Haan, L., et al. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10**, 251–280.

Einmahl, J. H., [de Haan], L., and Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Processes and their Applications* **70**, 143 – 171. doi:https://doi.org/10.1016/S0304-4149(97)00065-3.

Einmahl, J. H., Kiriliouk, A., and Segers, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes* 1–29.

Einmahl, J. H., Krajina, A., and Segers, J. (2012). An m-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics* **40**, 1764–1793.

Falk, M. (2019). *D-Norms. An Offspring of Multivariate Extreme Value Theory*. Springer International.

Falk, M., and Guillou, A. (2008). Peaks-over-threshold stability of multivariate generalized Pareto distributions. *J. Multivariate Anal.* **99**, 715–734. doi:10.1016/j.jmva.2007.03.009.

Falk, M., Hüsler, J., and Reiss, R.-D. (2004). *Laws of Small Numbers: Extremes and Rare Events*. 2nd ed. Birkhäuser, Basel. doi:10.1007/978-3-0348-7791-6.

Falk, M., Hüsler, J., and Reiss, R.-D. (2011). *Laws of Small Numbers: Extremes and Rare Events*. 3rd ed. Birkhäuser, Basel. doi:10.1007/978-3-0348-0009-9.

Falk, M., and Kohne, W. (1986). On the rate at which the sample extremes become independent. *Ann. Probab.* **14**, 1339–1346.

Falk, M., Padoan, S. A., and Wisheckel, F. (2019*a*). Conditional tail independence in archimedean copula models. *Journal of Applied Probability* **56**, 858869. doi:10.1017/jpr.2019.48.

Falk, M., Padoan, S. A., and Wisheckel, F. (2019*b*). Generalized pareto copulas: A key to multivariate extremes. *Journal of Multivariate Analysis* **174**, 104 538. doi:10.1016/j.jmva.2019.104538.

Falk, M., and Reiss, R.-D. (1988). Independence of order statistics. *Ann. Probab.* **16**, 854–862.

Falk, M., and Tichy, D. (2011). Asymptotic conditional distribution of exceedance counts: fragility index with different margins. *Ann. Inst. Stat. Math.* **64**, 1071–1085. doi:10.1007/s10463-011-0348-3.

FALK, M., and TICHY, D. (2012). Asymptotic conditional distribution of exceedance counts. *Adv. in Appl. Probab.* **44**, 270–291. doi:10.1239/aap/1331216653.

FALK, M., and WISHECKEL, F. (2017). Asymptotic independence of bivariate order statistics. *Statistics & Probability Letters* **125**, 91–98.

FALK, M., and WISHECKEL, F. (2018). Multivariate order statistics: the intermediate case. *Sankhya* **80**, 100–110. doi:10.1007/s13171-017-0099-1.

FISHER, R. A., and TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Cambridge Philos. Soc.* **24**, 180–190. doi:10.1017/S0305004100015681.

FRÉCHET, M. (1927). Sur la loi de probabilit de l'cart maximum. *Annales de la societe Polonaise de Mathematique* **6**, 93–116.

GALAMBOS, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association* **70**, 674–680.

GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. 2nd ed. Krieger, Malabar.

GELUK, J. L., DE HAAN, L., and DE VRIES, C. G. (2007). Weak & strong financial fragility. Tinbergen Institute Discussion Paper. TI 2007-023/2.

GENEST, C., and NEŠLEHOVÁ, J. (2012). Copula modeling for extremes. In *Encyclopedia of Environmetrics* (A. El-Shaarawi and W. Piegorsch, eds.), vol. 2, 530–541. Wiley, Chichester. doi:10.1002/9780470057339.vnn018.

GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)* **44**, 423–453. doi:10.2307/1968974.

GUERREIRO, C., ORTIZ, A. G., DE LEEUW, F., VIANA, M., and HORÁLEK, J. (2016). *Air Quality in Europe – 2016 Report*. Publications Office of the European Union.

GUILLOU, A., PADOAN, S. A., and RIZZELLI, S. (2018). Inference for asymptotically independent samples of extremes. *Journal of Multivariate Analysis* **167**, 114–135.

GUMBEL, E. J. (1946). On the independence of the extremes in a sample. *Ann. Math. Statist.* **17**, 78–81.

GUMBEL, E. J. (1958). *Statistics of Extremes.* Columbia University Press.

DE HAAN, L., and RESNICK, S. (1977). Limit theory for multivariate sample extremes. *Probab. Theory Related Fields* **40**, 317–337. doi:10.1007/BF00533086.

HEFFERNAN, J. E., and TAWN, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 497–546.

HUANG, X. (1992). *Statistics of Bivariate Extreme Values.* Ph.D. thesis, Tinbergen Institute Research Series.

HÜSLER, J., and LI, D. (2009). Testing asymptotic independence in bivariate extremes. *Journal of Statistical Planning and Inference* **139**, 990–998.

KLOSS, S. (2020). *Simulating Archimedean Copulas in Extreme Regions.* Master's thesis, University of Wuerzburg, Germany.

KLÜPPELBERG, C., KUHN, G., and PENG, L. (2008). Semi-parametric models for the multivariate tail dependence function–the asymptotically dependent case. *Scandinavian Journal of Statistics* **35**, 701–718.

KRUPSKII, P., and JOE, H. (2019). Nonparametric estimation of multivariate tail probabilities and tail dependence coefficients. *Journal of Multivariate Analysis* **172**, 147 – 161. doi:https://doi.org/10.1016/j.jmva.2019.02.013.

MCNEIL, A. J., and NEŠLEHOVÁ, J. (2009). Multivariate archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Ann. Statist.* **37**, 3059–3097. doi:10.1214/07-AOS556.

NELSEN, R. B. (2006). *An Introduction to Copulas.* Springer Series in Statistics, 2nd ed. Springer, New York. doi:10.1007/0-387-28678-0.

PICKANDS, J., III (1981). Multivariate extreme value distributions. *Proc. 43th Session ISI (Buenos Aires)* 859–878.

REISS, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer Series in Statistics. Springer, New York. doi:10.1007/978-1-4613-9620-8.

REISS, R.-D. (1993). *A Course on Point Processes*. Springer, New York. doi:10.1007/978-1-4613-9308-5.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*, *Applied Probability*, vol. 4. Springer, New York. doi:10.1007/978-0-387-75953-1. First Printing.

ROOTZÉN, H., and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12**, 917–930. doi:10.3150/bj/1161614952.

ROSSI, G., VIGOTTI, M. A., ZANOBETTI, A., REPETTO, F., GIANELLE, V., and SCHWARTZ, J. (1999). Air pollution and cause-specific mortality in milan, italy, 1980–1989. *Archives of Environmental Health: An International Journal* **54**, 158–164.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Pub. Inst. Stat. Univ. Paris* **8**, 229–231.

SKLAR, A. (1996). Random variables, distribution functions, and copulas – a personal look backward and forward. In *Distributions with fixed marginals and related topics* (L. Rüschendorf, B. Schweizer, and M. D. Taylor, eds.), *Lecture Notes – Monograph Series*, vol. 28, 1–14. Institute of Mathematical Statistics, Hayward, CA. doi:10.1214/lnms/1215452606.

SMITH, R. (1984). Thereshold methods for sample extremes in statistical extremes and applications, ed. *J. Tiago de Oliveira, Dordrecht: D. Reidel* 621–638.

VAN DER VAART, A. W. (2000). *Asymptotic statistics*, vol. 3. Cambridge university press.

VATAN, P. (1985). Max-infinite divisibility and max-stability in infinite dimensions. In *Probability in Banach Spaces V: Proceedings of the International Conference held in Medford, USA, July 16, 1984* (A. Beck, R. Dudley, M. Hahn, J. Kuelbs, and M. Marcus, eds.), *Lecture Notes in Mathematics*, vol. 1153, 400–425. Springer, Berlin. doi:10.1007/BFb0074963.

World Health Organization (2006). *Air quality guidelines: global update 2005*. 3rd ed. World Health Organization ISBN 9289021926.