Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Lehrstuhl für Mathematik VIII
Statistik

Contributions to Extreme Value Theory in Finite and Infinite Dimensions

With a Focus on Testing for Generalized Pareto Models

Dissertation zur Erlangung des Akademischen Grades eines
Doktors der Naturwissenschaften

vorgelegt von

Stefan Aulbach

aus

Würzburg

18. September 2015

Veröffentlichung im Februar 2016

There is no free lunch when it comes to high quantile estimation!

(Embrechts, Klüppelberg, and Mikosch, 1997, p. 349)

Preface and Acknowledgements

Some time ago I was asked to write an expert review about modeling certain univariate data beyond their maximum, including the estimation of tail probabilities. Of course, generalized Pareto distributions were the first tools that came to my mind, and I chose to base my review on the book Embrechts et al. (1997) since it provides an excellent mixture of theoretical background and practical advice. Although this review took (too) much time, delaying the handing in of the thesis you are about to read right now, the effort paid off in several ways: On the one hand, my senses for exploratory tools were sharpened, which improved the presentation of the simulation study at the end of this thesis. On the other hand, some quotes were burned into my memory that helped me recover from some setbacks, which are inevitable when someone writes a longer scientific work. Those quotes began with the impression that obtaining high quantile and tail estimates "[...] surely is a race we cannot win!", which was then reformulated in "this is a race which will be difficult to win!", and climaxed with the aforementioned statement "There is no free lunch when it comes to high quantile estimation!", cf. Embrechts et al. (1997, pp. 346, 349).

Obviously, stumbling blocks that are encountered in the univariate setting tend to grow to big rocks when the number of dimensions increases, and might even become huge mountains if the framework is extended to whole processes incorporating an infinite number of dimensions. Thankfully, I did not have to climb those mountains all alone and am very grateful to my advisor, Prof. Dr. Michael Falk, who helped me through several hard times during the past years and never stopped believing in me. I also thank him for quite a number of opportunities to present my results at conferences all over the world, which greatly helped getting input from other researchers around the globe and networking with them. Since parts of this thesis are in the process of publication in a scientific journal, I also owe valuable hints, from which my work has benefited a lot, to unknown referees. Moreover, I thank my former and present colleagues in Würzburg for their friendship and valuable discussions. In particular, I would like to mention Dr. Martin Hofmann, who laid major groundwork for my research, and Maximilian Zott, whose comments on selected parts of my thesis are highly appreciated. Furthermore, I thank my dear Canadian friend, Dr. Barbara Vona, for her kind support and her proof reading as a native speaker. I can only imagine how difficult it must have been to fight through a text written in a foreign scientific subject. Thank you a lot! All typos and false formulations that survived in the final text are due to my reluctance to follow her suggestions (including to dedicate this thesis to the Queen of England) or due to some just-in-time changes of the manuscript. Last but not least, I greatly thank my dear parents, Ulrike and Dieter, for their everlasting support and their sacrifices in order to give me the opportunity of a higher education. You are awesome and I am very proud of being your son!

Würzburg, September 2015 *Stefan Aulbach*

Contents

Some Notation

This is an overview of the most common notation used throughout this thesis. The following symbols have their typical meaning:

$\mathbb{N}, \mathbb{Q}, \mathbb{R}$	Set of all natural numbers excluding zero, of all rational numbers, and of all real numbers, respectively.
$\lvert x \rvert$	Absolute value of $x \in \mathbb{R}$.
$\lvert A \rvert$	Number of distinct elements of a set A.
$\mathbb{1}_A$	Indicator function of a set A, i. e. $\mathbb{1}_A(x) := 1$ if $x \in A$ and $\mathbb{1}_A(x) := 0$ else.
$\mathbb{0}_A$	Zero function of a set A, i. e. $\mathbb{0}_A(x) := 0$ for all $x \in A$.
$A \subset B$	A is a subset of or equal to the set B.
B^A	Set of all functions from A to B.
$f(A')$	$\{b \in B \mid \exists_{a \in A'}\ f(a) = b\}$ for $f \in B^A$ and $A' \subset A$.
$\mathcal{C}[0,1]$	Set of all continuous and real-valued functions on the interval $[0,1]$.

Since this thesis deals with multivariate as well as with functional random elements, it is quite convenient to define a common framework: Let I be an index set with $|I| \geq 1$ elements and $\mathcal{X}_t$ a non-empty set for each $t \in I$. In most cases we will choose $I = \{1, \ldots, d\}$ or $I = [0,1]$. Denote by

$$\mathop{\times}_{t \in I} \mathcal{X}_t := \begin{cases} \mathcal{X}_{t_0} & \text{if } I = \{t_0\} \\ \left\{ f \in \left(\bigcup_{t \in I} \mathcal{X}_t\right)^I \,\middle|\, \forall_{t \in I}\ f(t) \in \mathcal{X}_t \right\} & \text{if } |I| > 1 \end{cases}$$

the Cartesian product of $\mathcal{X}_t$, $t \in I$. Obviously $\mathcal{X}^I = \times_{t \in I} \mathcal{X}$ if $|I| > 1$ and $\mathcal{X}_t = \mathcal{X}$ for all $t \in I$. As usual, we also write $\times_{i=1}^d \mathcal{X}_i$ instead of $\times_{i \in \{1,\ldots,d\}} \mathcal{X}_i$ for $d \in \mathbb{N}$, and put $\mathcal{X}^d := \times_{i=1}^d \mathcal{X}$. If J is another non-empty index set satisfying $I \cap J = \emptyset$ and $\mathcal{X}_t$, $t \in J$, are further non-empty sets, then define

$$\mathop{\times}_{t \in I} \mathcal{X}_t \times \mathop{\times}_{t \in J} \mathcal{X}_t := \mathop{\times}_{t \in I \cup J} \mathcal{X}_t.$$

Occasionally, the index set of a factor in a Cartesian product is omitted. In these cases, the missing indices should be clear from the context, e. g.

$$A \times \bigtimes_{t \in I \setminus \{t_0\}} \mathcal{X}_t := \bigtimes_{t \in I} \mathcal{X}_t \quad \text{and} \quad A \times B := \Big\{ f \in (A \cup B)^{\{1,2\}} \;\Big|\; f(1) \in A, f(2) \in B \Big\}$$

where $t_0 \in I$, $\mathcal{X}_{t_0} = A$ and $A, B \neq \emptyset$.

An element $f \in \bigtimes_{t \in I} \mathcal{X}_t$ is also denoted by $(f(t))_{t \in I}$ or $(f(t))_{t \in T} \times (f(t))_{t \in T^c}$ for $T \subset I$ and $T^c = I \setminus T$. For this purpose we identify $(f(t))_{t \in I} \times (f(t))_{t \in \emptyset}$ with $(f(t))_{t \in I}$ and $\bigtimes_{t \in I} \mathcal{X}_t \times \bigtimes_{t \in \emptyset} \mathcal{X}_t$ with $\bigtimes_{t \in I} \mathcal{X}_t$. In particular, the column-vector $\boldsymbol{x} = (x_1, \ldots, x_d)^\mathsf{T} \in \mathcal{X}^d$ denotes the function $(x_i)_{i=1}^d := (x_i)_{i \in \{1,\ldots,d\}}$ and we define $h(x_1, \ldots, x_d) := h(\boldsymbol{x})$ for each function h on $\mathcal{X}^d$.

If the sets $\mathcal{X}_t$, $t \in I$, are in fact topological spaces, then $\mathbb{B}(\mathcal{X}_t)$ denotes the Borel-σ-algebra corresponding to $\mathcal{X}_t$, and $\mathbb{B}\big(\bigtimes_{t \in I} \mathcal{X}_t\big)$ is the Borel-σ-algebra of $\bigtimes_{t \in I} \mathcal{X}_t$ with respect to the product topology. We have in particular $\mathbb{B}(\mathcal{C}[0,1]) = \mathcal{C}[0,1] \cap \mathbb{B}(\mathbb{R}^{[0,1]}) = \big\{ B \cap \mathcal{C}[0,1] \mid B \in \mathbb{B}\big(\mathbb{R}^{[0,1]}\big) \big\}$. Unless stated otherwise, a space $\mathcal{X}$ shall be equipped with its Borel-σ-algebra.

Now consider $\mathcal{X} \subset \mathbb{R}$ and write $\mathbb{B}_d$ instead of $\mathbb{B}\big(\mathbb{R}^d\big)$, the corresponding Lebesgue-measure being λ_d. In order to keep notation short, define $\mathbf{0} := (0)_{i=1}^d$, $\mathbf{1} := (1)_{i=1}^d$, $\boldsymbol{\infty} := (\infty)_{i=1}^d$, and $\mathbf{e}_j := (1)_{i \in \{j\}} \times (0)_{i \in \{1,\ldots,d\} \setminus \{j\}}$, which is the j-th unit vector in $\mathbb{R}^d$. Moreover, all operations and relations such as $+$, $-$, $\cdot$, $/$, $<$, $\leq$ are meant pointwise, i. e. $f + g := \big(f(t) + g(t)\big)_{t \in I}$ for $f, g \in \mathcal{X}^I$ as well as

$$f_+ := \max\{\mathbb{0}_I, f\} := \big(\max\{0, f(t)\}\big)_{t \in I} \qquad \text{and} \qquad f_- := (-f)_+.$$

The same interpretation holds for the application of univariate functions to $f \in \mathcal{X}^I$, e. g. $\exp(f) := \big(\exp\big(f(t)\big)\big)_{t \in I}$. However, the pointwise reciprocal of f will be denoted by $\frac{\mathbb{1}_I}{f}$ and *not* by f^{-1}. The use of the symbol f^{-1} depends on the context and means, respectively, the preimage of a set under f, the inverse function of f, or the quantile function of f. There should be no risk of confusion.

For $f, g \in \mathcal{X}^I \subset \mathbb{R}^I$ the intervals $[f, g]$, $(f, g]$, $[f, g)$, and (f, g) are defined by

$$\begin{aligned} [f,g] &:= \{h \in \mathcal{X}^I \mid f \leq h \leq g\}, & [f,g) &:= \{h \in \mathcal{X}^I \mid f \leq h < g\}, \\ (f,g] &:= \{h \in \mathcal{X}^I \mid f < h \leq g\}, & (f,g) &:= \{h \in \mathcal{X}^I \mid f < h < g\}. \end{aligned}$$

Consider a function $\xi : \mathcal{X}^I \to \mathcal{Y}$ that maps $f \in \mathcal{X}^I$ to an element $\xi(f)$ of a metric space $\mathcal{Y}$. Then we write $\lim_{f \to f_0} \xi(f) = y$ for some $y \in \mathcal{Y}$ if $\lim_{n \to \infty} \xi(f_n) = y$ holds for any sequence $(f_n)_{n \in \mathbb{N}}$ in $\mathcal{X}^I$ with limit f. Analogously, we write $\lim_{f \to f_0 +} \xi(f) = y$ if

$\lim_{n\to\infty} \xi(f_n) = y$ for any sequence $(f_n)_{n\in\mathbb{N}}$ in $\{h \in \mathcal{X}^I \mid f \neq h \geq f\}$ with limit f, and $\lim_{f\to f_0-} \xi(f) = y$ if $\lim_{n\to\infty} \xi(f_n) = y$ for any sequence $(f_n)_{n\in\mathbb{N}}$ in $\{h \in \mathcal{X}^I \mid f \neq h \leq f\}$ with limit f.

Norms will play an important role throughout the present text. The usual p-norm is denoted by $\|\cdot\|_p$, i. e.

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{1/p} \quad \text{and} \quad \|f\|_p = \left(\int_0^1 |f(t)|^p \,\mathrm{d}t\right)^{1/p} \quad \text{for } 1 \leq p < \infty$$

as well as $\|\boldsymbol{x}\|_\infty = \max_{1\leq i\leq d} |x_i|$ and $\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|$ where $\boldsymbol{x} \in \mathbb{R}^d$, $f \in \mathbb{R}^{[0,1]}$. Note that $\|\cdot\|_p$ is for $1 \leq p < \infty$ actually not a norm on $\mathbb{R}^{[0,1]}$ but a pseudo-norm since $\|f\|_p = 0$ for all $f \in \mathbb{R}^{[0,1]}$ that attain the value zero λ_1-almost-everywhere.

When we consider the asymptotic behavior of two functions $f, g : \mathcal{X} \to \mathbb{R}$ defined on a normed vector space $(\mathcal{X}, \|\cdot\|)$, we write

$$\begin{aligned}
f(x) = \mathcal{O}(g(x)) \quad \text{as} \quad x \to x_0 \quad &:\Longleftrightarrow \quad \exists_{C>0,\delta>0}\, \forall_{x\in\mathcal{X},\|x-x_0\|\leq\delta}\ |f(x)| \leq C\,|g(x)| \\
f(x) = o(g(x)) \quad \text{as} \quad x \to x_0 \quad &:\Longleftrightarrow \quad \forall_{\varepsilon>0}\, \exists_{\delta>0}\, \forall_{x\in\mathcal{X},\|x-x_0\|\leq\delta}\ |f(x)| \leq \varepsilon\,|g(x)| \\
f(x) \sim g(x) \quad \text{as} \quad x \to x_0 \quad &:\Longleftrightarrow \quad \frac{f(x)}{g(x)} \to 1 \quad \text{as} \quad x \to x_0
\end{aligned}$$

where x_0 is a limit point of $\mathcal{X}$. If $\mathcal{X} \subset \mathbb{R}$ and $x_0 = \pm\infty$, then the condition $\|x - x_0\| \leq \delta$ is replaced with $x \geq \delta$ or $x \leq -\delta$, respectively.

If $(\mathcal{X}, \mathcal{A}, \mu)$ is a measure space and $T : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ a measurable mapping into a measurable space $(\mathcal{Y}, \mathcal{B})$, $(\mu * T)$ denotes the push forward measure of μ by T, i. e. $(\mu * T)(B) = \mu(\{x \in \mathcal{X} \mid T(x) \in B\})$ for $B \in \mathcal{B}$. Throughout this thesis, P is a probability measure on some suitable measurable space $(\Omega, \mathcal{A})$ and $\mathrm{E}(X)$ is, if existent, the expected value of a random element $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathbb{B}(\mathcal{X}))$ with respect to P. The distribution of X, i. e. the push forward measure $(\mathrm{P} * X)$, is also denoted by $\mathcal{L}_X$ whenever the underlying probability measure P is of minor interest. In particular, $\mathcal{L}_{X_1}$ and $\mathcal{L}_{X_2}$ may be based on two different probability spaces $(\Omega_1, \mathcal{A}_1, \mathrm{P}_1)$ and $(\Omega_2, \mathcal{A}_2, \mathrm{P}_2)$. We write $X_1 \overset{D}{=} X_2$ if $\mathcal{L}_{X_1}$ equals $\mathcal{L}_{X_2}$. By $X \sim \mathcal{L}$ we denote that X has distribution $\mathcal{L}$ and we also write $X \sim F$ if $\mathcal{L}$ has distribution function F. For a sequence $(X_n)_{n\in\mathbb{N}}$ of random elements, $X_n \overset{D}{\to} \mathcal{L}$ or $X_n \overset{D}{\to} F$ mean convergence in distribution towards $\mathcal{L}$ as $n \to \infty$; if $Y \sim \mathcal{L}$ then this is also denoted by $X_n \overset{D}{\to} Y$ as $n \to \infty$. As usual, we write $\mathcal{U}[0,1]$ for the uniform distribution on $[0,1]$, $\mathcal{B}(n,p)$ for the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1)$, $\mathcal{N}(\mu, \sigma^2)$ for the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and χ^2_n for the chi-square distribution with n degrees of freedom.

Introduction

Despite being a rather young mathematical discipline, extreme value theory has been subject to major change. However, it never lost its aim to describe rare but extreme events with prominent examples being flooding from the sea, heavy earthquakes, and financial collapses. Starting with the asymptotic distribution of the suitably standardized sample maximum, cf. the Fisher–Tippett–Gnedenko Theorem, the results have steadily been generalized from univariate observations to multivariate ones and even process data. More recently, the focus shifted more and more to all observations in the sample that exceed a certain high threshold, instead of considering the maximum only. Applied to the aforementioned examples, one would consider only waves exceeding the height of a certain dike, earthquakes having at least a certain intensity, and, after applying a simple transformation, share prices falling below some low threshold. It turned out that the previous results on maxima could be carried over to this new framework, leading to so-called generalized Pareto distributions, which are the only reasonable probability distributions suited for modeling observations above a high threshold.

Probably due to this "exclusiveness", it seems to be widely accepted to just apply generalized Pareto models to observed data, at least for multivariate and process data, without an a priori check whether these kind of models are actually suitable for the data under consideration. Since there indeed are probability distributions for which a generalized Pareto model might fail, there is still a certain gap in the process of statistical inference. This thesis therefore aims at providing a statistical test for the hypothesis that the data are in a certain neighborhood of a generalized Pareto distribution. In this context, also some punctual contributions to extreme value theory in general will be considered, focusing on finite dimensional and on functional observations. By using a notation based on certain norms, called D-norms, the inherent similarities of finite dimensional extreme value theory and extreme value theory for continuous processes will be particularly stressed. Moreover, D-norms provide an elegant way to express the most central terms of extreme value theory in general, such as "max-stable distribution", "domain of attraction", and "generalized Pareto distribution".

In Chapter 1 we will briefly review — both in the finite dimensional context and in the one of continuous functions on $[0,1]$ — how generalized Pareto distributions are based on classical extreme value theory, i. e. asymptotic results for maxima. This gives us the opportunity to introduce the notation of D-norms and to link it with other representations found in the standard literature. Moreover, the term "copula" will be carried over to stochastic processes with continuous sample paths.

Based on the characteristic excursion stability of a generalized Pareto distribution, Chapter 2 defines certain neighborhoods, called δ-neighborhoods, of a generalized Pareto distribution. Due to a decomposition of a distribution into its univariate margins and a copula, we will then derive a test for the hypothesis that the copula underlying the observed data is in such a δ-neighborhood. This will be done by considering finite dimensional observations, and then generalizing the results to process data. Finally, it will be shown that both frameworks can be linked consistently if a continuous process can only be observed at a finite grid of observation points, and if the fineness of this grid increases. Each of these steps — finite dimensional test, functional test, and the linkage of both — will be done for copula data first, before more general data are considered.

Since the derivation of the asymptotic distribution of the test statistic will require certain technical restrictions, Chapter 3 analyzes these assumptions in more detail. It provides in particular some examples of copulas that are in a δ-neighborhood, i. e. the null hypothesis is true, and of copulas that do not satisfy the null hypothesis. We will consider moreover a simple approach how a finite dimensional copula can be extended to a functional one, and we will give some practical advice how to choose the free parameters incorporated in the test statistics.

Finally, Chapter 4 compares the in total three different test statistics with another test found in the literature that has a similar null hypothesis, which will be done by means of a simulation study. This thesis ends with a short summary of the results and an outlook to further open questions.

1 Basics

This chapter is dedicated to provide elementary terms and concepts needed for the main part of this thesis. We will shortly recall some main results concerning max-stable distributions and generalized Pareto distributions; we refer to Beirlant et al. (2004), de Haan and Ferreira (2006), and Falk et al. (2011) for more details and further reading. Subsequent chapters will mainly rely on the fact that observations above a high threshold can be reasonably modeled only by means of generalized Pareto distributions, where Section 1.1 considers the finite dimensional framework and Section 1.2 deals with the space of continuous functions on $[0, 1]$.

The similarities of both settings, the finite dimensional one and the functional one, are particularly stressed by using a kind of non-standard notation, which nevertheless is thoroughly founded on standard literature and eases the insight into the theoretical results of later chapters. In particular, all of the terms "max-stable distribution", "generalized Pareto distribution", and "domain of attraction" can be broken down to conditions on a certain class of norms. This fact will be exploited in Chapter 2, where these conditions will be sharpened in order to derive statistical tests for certain neighborhoods of a generalized Pareto distribution.

1.1 Finite Dimensional Extreme Value Theory

We start with the uni- and multivariate case. Let $X, X_1, X_2, \dots$ be i. i. d. random variables with distribution function F, i. e. $F(x) = \mathrm{P}(X \le x)$, $x \in \mathbb{R}$. The maximum in the sample $X_1, \dots, X_n$ obviously has the distribution function

$$\mathrm{P}\left(\max_{1\le i\le n} X_i \le x\right) = \left(\mathrm{P}(X \le x)\right)^n = F^n(x), \qquad x \in \mathbb{R}. \tag{1.1.1}$$

It is well-known that those distributions that are, in a certain sense, stable with respect to exponentiation are crucial for the definition of generalized Pareto distributions. Note that F^n converges pointwise to $\mathbb{1}_{[\omega(F),\infty)}$ as $n \to \infty$, where $\omega(F) := \sup\{x \in \mathbb{R} \mid F(x) < 1\}$ is the *upper endpoint* of F. This fact, however, does not provide sufficient information about the shape of F in a neighborhood of $\omega(F)$. An approach resulting in a *non-degenerate* limit, i. e. the limit is not of the form $\mathbb{1}_{[y,\infty)}$ for some $y \in \mathbb{R} \cup \{\infty\}$, is more desirable.

Definition 1.1.2 Let F and G be (univariate) distribution functions where G is non-degenerate. Then F is in the *domain of attraction* of G if for each $n \in \mathbb{N}$ there are *norming constants* $a_n > 0$ and $b_n \in \mathbb{R}$ satisfying

$$F^n(a_n x + b_n) \to G(x) \qquad \text{as} \qquad n \to \infty \tag{1.1.3}$$

for all continuity points x of G. In this case G is referred to as a *(univariate) extreme value distribution* or a *(univariate) max-stable distribution* (MSD) and we write $F \in \mathcal{D}(G)$.

Equation (1.1.1) shows that (1.1.3) is a condition on the asymptotic distribution of the suitably standardized maximum $(\max_{1 \le i \le n} X_i - b_n)/\, a_n$ among $X_1, \ldots, X_n$. The distribution functions G that may appear as a limit in (1.1.3) are well-known; they were identified by Fisher and Tippett (1928) and Gnedenko (1943). Furthermore Khintchine's convergence theorem, cf. Leadbetter et al. (1983, Theorem 1.2.3), shows that the limit is, in a certain sense, uniquely determined, which is part (ii) of the following result:

Theorem 1.1.4 *The class of all univariate MSDs is given by* $\{G_{\gamma;\mu,\sigma} \mid \gamma, \mu \in \mathbb{R},\, \sigma > 0\}$ *where* $G_{\gamma;\mu,\sigma}$ *is a distribution function defined by*

$$G_{\gamma;\mu,\sigma}(x) := \exp\left(-\left(1 + \gamma\, \frac{x - \mu}{\sigma}\right)^{-1/\gamma}\right), \qquad 1 + \gamma\, \frac{x - \mu}{\sigma} > 0, \tag{1.1.5}$$

and $G_{0;\mu,\sigma}(x)$ *is interpreted as* $\exp\Bigl(-\exp\Bigl(-\frac{x-\mu}{\sigma}\Bigr)\Bigr)$, $x \in \mathbb{R}$. *Moreover we have for any distribution function* F*:*

(i) $F \in \mathcal{D}(G_{\gamma;\mu,\sigma})$ *if and only if* $F \in \mathcal{D}(G_\gamma)$ *where* $G_\gamma := G_{\gamma;0,1}$.

(ii) $F \in \mathcal{D}(G_{\gamma_1})$ *and* $F \in \mathcal{D}(G_{\gamma_2})$ *imply* $\gamma_1 = \gamma_2$.

The representation (1.1.5) is commonly known as the *von Mises parametrization*, due to von Mises (1936); cf. Jenkinson (1955). Note that all univariate MSDs are continuous and, if $F \in \mathcal{D}(G)$, (1.1.3) holds for all $x \in \mathbb{R}$.

According to Theorem 1.1.4 (i) we may, without loss of generality, restrict ourselves to MSDs of the form G_γ, $\gamma \in \mathbb{R}$. The remaining parameter γ holds all the essential information about the upper tail of $F \in \mathcal{D}(G_\gamma)$:

Definition 1.1.6 The parameter γ in Theorem 1.1.4 is referred to as the *extreme value index* of the MSD G_γ.

Note that all MSDs share the property

$$G_\gamma(x) = G_\gamma^n(a_n(\gamma)\, x + b_n(\gamma)), \qquad x \in \mathbb{R},$$

where $a_n(0) = 1$, $b_n(0) = \log(n)$ and $a_n(\gamma) = n^\gamma$, $b_n(\gamma) = \gamma^{-1}(n^\gamma - 1)$ for $\gamma \neq 0$. This means that, if $Y_1, \dots, Y_n$ are i. i. d. with distribution function G_γ, the suitably standardized maximum among $Y_1, \dots, Y_n$ has the same distribution as the original data, cf. (1.1.1). Thus G_γ is, roughly speaking, stable with respect to taking the maximum. On the other hand, if F is a non-degenerate distribution function such that for all $n \in \mathbb{N}$ there are $a_n > 0$ and $b_n \in \mathbb{R}$ satisfying $F(x) = F^n(a_n\, x + b_n)$ for all $x \in \mathbb{R}$, then F is an MSD in the sense of Definition 1.1.2; cf. Leadbetter et al. (1983, Theorem 1.3.1). This reasoning justifies the term "max-stable" distributions in the mentioned definition.

Now consider $d \in \mathbb{N}$ and a d-variate distribution function $F : \mathbb{R}^d \to [0, \infty)$, i. e. there is a random vector $\boldsymbol{X} = (X_1, \dots, X_d)^\intercal$ in $\mathbb{R}^d$ satisfying

$$F(\boldsymbol{x}) = \mathrm{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$$

for all $\boldsymbol{x} = (x_1, \dots, x_d)^\intercal \in \mathbb{R}^d$. Again we aim at deriving a characterization of the shape of F close to its *upper endpoint* $\boldsymbol{\omega}(F) := (\omega(F_1), \dots, \omega(F_d))^\intercal$, where F_i is the *i-th margin* of F, i. e. $F_i(x) := \mathrm{P}(X_i \leq x)$, $x \in \mathbb{R}$. If $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots$ are independent copies of $\boldsymbol{X}$ then the distribution function of the standardized maximum is

$$\mathrm{P}\left(\frac{\max_{1\leq j\leq d}\{\boldsymbol{X}^{(j)}\} - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{x} \right) = F^n(\boldsymbol{a}_n\, \boldsymbol{x} + \boldsymbol{b}_n), \qquad x \in \mathbb{R}^d, \tag{1.1.7}$$

where $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$ are sequences in $(0, \infty)^d$ and $\mathbb{R}^d$, respectively, and the maximum is taken componentwise:

$$\max_{1\leq j\leq d}\{\boldsymbol{X}^{(j)}\} = \left(\max_{1\leq j\leq d}\{X_i^{(j)}\} \right)_{i=1}^d .$$

As before we focus on non-degenerate limits of (1.1.7) as $n \to \infty$:

Definition 1.1.8 Let F and G be d-variate distribution functions where G is *non-degenerate*, i. e. all margins of G are non-degenerate. F is in the *domain of attraction* of G if for each $n \in \mathbb{N}$ there are *norming vectors* $\boldsymbol{a}_n > \boldsymbol{0}$, $\boldsymbol{b}_n \in \mathbb{R}^d$ such that

$$F^n(\boldsymbol{a}_n\, \boldsymbol{x} + \boldsymbol{b}_n) \to G(\boldsymbol{x}) \qquad \text{as} \qquad n \to \infty \tag{1.1.9}$$

for all continuity points of G. In this case G is referred to as a *(d-variate) max-stable distribution* (MSD) and we write $F \in \mathcal{D}(G)$.

It is easy to verify that any MSD G is continuous: If (1.1.9) holds with $\boldsymbol{a}_n = (a_n^{(1)}, \dots, a_n^{(d)})^\intercal$ and $\boldsymbol{b}_n = (b_n^{(1)}, \dots, b_n^{(d)})^\intercal$ then the multivariate mapping theorem, see e. g. Billingsley (2012, Theorem 29.2), implies

$$F_i^n(a_n^{(i)}\, x + b_n^{(i)}) \to G_i(x) \qquad \text{as} \qquad n \to \infty \tag{1.1.10}$$

for each $i \in \{1, \dots, d\}$ and each continuity point of G_i, i. e. each margin of F is in the domain of attraction of the corresponding margin of G. Now Theorem 1.1.4 shows that each margin of G is continuous and, thus, G is continuous as well; cf. Galambos (1978, Theorem 5.2.1).

The term "max-stable distribution" has exactly the same interpretation as in the univariate case:

Theorem 1.1.11 *A non-degenerate distribution function G is an MSD if and only if it is max-stable, i. e. for each $n \in \mathbb{N}$ there exist norming vectors $\boldsymbol{a}_n > \boldsymbol{0}$, $\boldsymbol{b}_n \in \mathbb{R}^d$ such that*

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

This result can, e. g., be found in Resnick (1987, Proposition 5.9). Note that the cited result shows in particular that a non-degenerate distribution function G is max-stable if and only if there are functions $\boldsymbol{a}, \boldsymbol{b} : (0, \infty) \to \mathbb{R}^d$ satisfying $\boldsymbol{a} > \boldsymbol{0}$ and

$$G^s(\boldsymbol{a}(s)\, \boldsymbol{x} + \boldsymbol{b}(s)) = G(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d,\ s > 0.$$

Sklar's Theorem and Max-Stable Distributions

Next, we characterize (1.1.9) by terms of a well-known decomposition theorem of multivariate distribution functions, which are split into their margins and a copula, see Theorem 1.1.13 below. This decomposition will turn out to be a crucial tool for later chapters as it allows the assumption without loss of generality that each margin of an MSD corresponds to the standard negative exponential distribution.

Definition 1.1.12 A *copula* C is a d-variate distribution function where each margin of C is the uniform distribution on $(0, 1)$, i. e. $C_i(x) = x$, $x \in (0, 1)$, for each $i \in \{1, \dots, d\}$.

Now we state the aforementioned decomposition theorem, which is taken from Schweizer and Sklar (2005, Theorem 6.2.4 and Theorem 6.2.5) and Nelsen (2006, Theorem 2.10.9). It was introduced by Sklar (1959) but a crucial tool for its proof, Theorem 6.2.6 in Schweizer and Sklar (2005), was established in Moore and Spruill (1975) and Deheuvels (1978); cf. Sklar (1996). Another notable supplement is Rüschendorf (2009).

Theorem 1.1.13 (Sklar's Theorem) *Let F be a d-variate distribution function with margins $F_1, \dots, F_d$. Then there exists a copula C satisfying*

$$F(\boldsymbol{x}) = C(F_1(x_1), \dots, F_d(x_d)) \quad \textit{for all} \quad \boldsymbol{x} = (x_1, \dots, x_d)^\intercal \in \mathbb{R}^d. \tag{1.1.14}$$

Furthermore, the restriction of C to the domain $\bigtimes_{i=1}^d F_i(\mathbb{R})$ is uniquely determined and has the representation

$$C(\boldsymbol{u}) = F\big(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\big) \quad \textit{for all} \quad \boldsymbol{u} = (u_1, \dots, u_d)^\intercal \in \bigtimes_{i=1}^d F_i(\mathbb{R}) \tag{1.1.15}$$

where $F_i^{-1}(u) = \inf\{x \in \mathbb{R} \mid F_i(x) \geq u\}$, $0 < u < 1$, is the quantile function of F_i.

If conversely $F_1, \dots, F_d$ are univariate distribution functions and C is a copula then F defined by (1.1.14) *is a d-variate distribution function with margins $F_1, \dots, F_d$.*

Although a copula of a distribution function F is not uniquely determined in general, (1.1.15) implies that it *is* unique on the relevant domain. Therefore, we will call any copula satisfying (1.1.15) *the* copula of F, denoted by C_F, where the subscript may be omitted if there is no risk of confusion. Note that the copula C_G of an MSD G is always uniquely determined but it is *not* max-stable in the sense of Theorem 1.1.11. However, Theorem 1.1.11 implies the property

$$C_G^n(\boldsymbol{u}^{1/n}) = C_G(\boldsymbol{u}) \quad \text{for all } \boldsymbol{u} \in [\boldsymbol{0}, \boldsymbol{1}] \text{ and } n \in \mathbb{N}, \tag{1.1.16}$$

which is dual to the max-stability of G, see e. g. Galambos (1978, Theorems 5.2.1 and 5.2.4).

Definition 1.1.17 The copula of an MSD is called an *extreme value copula* (EVC).

For a further discussion as well as examples of EVCs we refer to Gudendorf and Segers (2010) and Ribatet and Sedki (2013). As we have already seen in (1.1.10),

$$F_i^n\big(a_n^{(i)}\, x + b_n^{(i)}\big) \to_{n\to\infty} G_i(x), \qquad x \in \mathbb{R},\ 1 \leq i \leq d,$$

is a necessary condition for $F \in \mathcal{D}(G)$. We obtain furthermore

$$C_F^n\Big(F_1\big(a_n^{(1)}\, x_1 + b_n^{(1)}\big), \dots, F_d\big(a_n^{(d)}\, x_d + b_n^{(d)}\big)\Big) \to_{n\to\infty} C_G(G_1(x_1), \dots, G_d(x_d))$$

by applying Theorem 1.1.13 to (1.1.9), which suggests in conjunction with (1.1.16) to analyze the copula and the margins of F separately. Indeed this procedure is well-established:

Theorem 1.1.18 (Galambos, 1978, Theorem 5.2.3; Deheuvels, 1978, 1984) *A d-variate distribution function F is in the domain of attraction of an MSD G if and only if the i-th margin of F is in the domain of attraction of the i-th margin of G for each $i \in \{1, \dots, d\}$ together with*

$$C_F^n(\boldsymbol{u}^{1/n}) \to C_G(\boldsymbol{u}) \quad \text{as} \quad n \to \infty, \qquad \boldsymbol{u} \in (0,1)^d, \tag{1.1.19}$$

where C_F and C_G denote the copulas of F and G, respectively.

Based on this crucial observation, we are able to justify restricting ourselves to a certain kind of standard MSDs. Although the next result is not new, its proof is stated since — using only well-known arguments from the theory of copulas — it might be of interest of its own.

Lemma 1.1.20 (Aulbach et al., 2012*a*; cf. de Haan and Ferreira, 2006, Theorem 6.1.1) *With the notations of Theorem 1.1.18,* (1.1.19) *and*

$$C_F^n\left(\mathbf{1} + \frac{1}{n}\boldsymbol{x}\right) \to_{n\to\infty} C_G(\exp(\boldsymbol{x})), \qquad \boldsymbol{x} \le \mathbf{0}, \tag{1.1.21}$$

are equivalent. Furthermore, (1.1.21) *holds if and only if*

$$\frac{1 - C_F(\mathbf{1} + t\boldsymbol{x})}{t} \to_{t\to 0+} -\log(C_G(\exp(\boldsymbol{x}))), \qquad \boldsymbol{x} \le \mathbf{0}, \tag{1.1.22}$$

is true.

Proof. Taylor's formula and Nelsen (2006, Theorem 2.10.7) imply

$$\begin{aligned}
&\left|C_F(\exp(n^{-1}\boldsymbol{x})) - C_F(\mathbf{1} + n^{-1}\boldsymbol{x})\right| \\
&= \left|C_F\left(1 + \frac{x_1}{n} + o\left(\frac{x_1}{n}\right), \dots, 1 + \frac{x_d}{n} + o\left(\frac{x_d}{n}\right)\right) - C_F\left(1 + \frac{x_1}{n}, \dots, 1 + \frac{x_d}{n}\right)\right| \\
&\le \sum_{i=1}^d o\left(\frac{x_i}{n}\right) = o\left(\frac{\|\boldsymbol{x}\|}{n}\right) \quad \text{as} \quad n \to \infty
\end{aligned}$$

pointwise for each $\boldsymbol{x} = (x_1, \dots, x_d)^\intercal \le \mathbf{0}$ and for any norm $\|\cdot\|$ on $\mathbb{R}^d$. Thus

$$\begin{aligned}
C_F^n\left(\mathbf{1} + \frac{1}{n}\boldsymbol{x}\right) &= \left(C_F(\exp(n^{-1}\boldsymbol{x})) + o\left(\frac{\|\boldsymbol{x}\|}{n}\right)\right)^n \\
&= C_F^n(\exp(n^{-1}\boldsymbol{x}))\left(1 + \frac{1}{n}\frac{n\, o(n^{-1}\|\boldsymbol{x}\|)}{C_F(\exp(n^{-1}\boldsymbol{x}))}\right)^n \quad \text{as } n \to \infty
\end{aligned}$$

which implies the first assertion. We proceed as in the proof of de Haan and Ferreira (2006, Theorem 1.1.2): A Taylor expansion of the function $(0,1] \ni y \mapsto -\log(y)$ at $y = 1$ yields that (1.1.21) is equivalent to

$$n\left(1 - C_F\left(\mathbf{1} + \frac{1}{n}\boldsymbol{x}\right)\right) \to_{n\to\infty} -\log\big(C_G(\exp(\boldsymbol{x}))\big), \qquad \boldsymbol{x} \le \mathbf{0}.$$

Since $\lfloor\frac{1}{t}\rfloor \le \frac{1}{t} \le (1 + \lfloor\frac{1}{t}\rfloor^{-1})\lfloor\frac{1}{t}\rfloor$ for $t > 0$, where $\lfloor\cdot\rfloor$ denotes the integer part, we obtain

$$\frac{1 - C_F\left(\mathbf{1} + \lfloor\frac{1}{t}\rfloor^{-1}\boldsymbol{x}\right)}{t} \to_{t\to 0+} -\log\big(C_G(\exp(\boldsymbol{x}))\big), \qquad \boldsymbol{x} \le \mathbf{0},$$

and (1.1.22) follows from

$$0 \le \frac{1}{\lfloor\frac{1}{t}\rfloor} - t = \frac{\frac{1}{t} - \lfloor\frac{1}{t}\rfloor}{\lfloor\frac{1}{t}\rfloor\frac{1}{t}} \le \frac{1}{\lfloor\frac{1}{t}\rfloor\frac{1}{t}} = o(t) \quad \text{as} \quad t \to 0+$$

and Nelsen (2006, Theorem 2.10.7). □

The previous considerations suggest to focus on the following standard case, which will be crucial throughout the rest of this thesis:

Definition 1.1.23 An MSD G is called a *standard* MSD if all margins of G coincide with $\exp(x)$ for $x \le 0$, the standard negative exponential distribution.

It is obvious that G is a standard MSD if and only if it can be written as $G(\boldsymbol{x}) = C_G(\exp(\boldsymbol{x}))$, $\boldsymbol{x} \in \mathbb{R}^d$, where C_G is a copula satisfying (1.1.16).

Remark 1.1.24 A common approach in the literature is to consider *simple* MSDs instead of standard ones, i. e. the margins are assumed to be standard Fréchet instead of standard negative exponential, $G_i(x) = \exp(-x^{-1})$ for $x > 0$, cf. de Haan and Ferreira (2006, Theorem 6.1.1). This is due to the fact that the Fréchet distribution is, according to Theorem 1.1.4, the prototype of a probability distribution with a heavy upper tail. However, Theorem 1.1.13 and Theorem 1.1.18 show that both cases can easily be transformed into one another: Let G be an MSD with copula C. If G is standard max-stable then

$$G\left(-\frac{\mathbf{1}}{\boldsymbol{x}}\right) = C\left(\exp\left(-\frac{\mathbf{1}}{\boldsymbol{x}}\right)\right) = C^n\left(\exp\left(-\frac{1}{n}\frac{\mathbf{1}}{\boldsymbol{x}}\right)\right) = G^n\left(-\frac{1}{n}\frac{\mathbf{1}}{\boldsymbol{x}}\right), \qquad \boldsymbol{x} > \mathbf{0},$$

is simple max-stable. Conversely, if G is simple max-stable, then

$$G\left(-\frac{\mathbf{1}}{\boldsymbol{x}}\right) = C(\exp(\boldsymbol{x})) = C^n\left(\exp\left(\frac{1}{n}\boldsymbol{x}\right)\right) = G^n\left(-n\frac{\mathbf{1}}{\boldsymbol{x}}\right), \qquad \boldsymbol{x} < \mathbf{0},$$

is standard max-stable. In both cases $G(-\mathbf{1}/\cdot)$ has the same copula as G.

D-Norms and Generators

From the previous reasoning it is clear that any MSD G may be written as

$$G(\boldsymbol{x}) = C(G_1(x_1), \dots, G_d(x_d)), \qquad \boldsymbol{x} = (x_1, \dots, x_d)^\intercal \in \mathbb{R}^d,$$

with univariate MSDs $G_1, \dots, G_d$ and an EVC C, which is involved in the limit in (1.1.22). Theorem 1.1.26 below will show that this limit actually defines a norm $\|\cdot\| = -\log C(\exp(-|\cdot|))$ on $\mathbb{R}^d$. This kind of norms, which will be crucial for the definition of generalized Pareto distributions, is generated by a certain class of random vectors:

Definition 1.1.25 Let $\boldsymbol{Z}$ be a random vector in $[\boldsymbol{0}, \boldsymbol{\infty})$ that satisfies $\mathrm{E}(\boldsymbol{Z}) = \boldsymbol{1}$. Then $\|\cdot\|_D$ defined by

$$\|\boldsymbol{x}\|_D := \mathrm{E}(\|\boldsymbol{x}\boldsymbol{Z}\|_\infty) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^d$$

is called a *D-norm* with *generator* $\boldsymbol{Z}$. Furthermore, two generators are *equivalent* if they give rise to the same D-norm. The value $\|\boldsymbol{1}\|_D = \mathrm{E}(\|\boldsymbol{Z}\|_\infty)$ is also referred to as the *generator constant* of $\boldsymbol{Z}$.

It is quite easy to verify that any D-norm $\|\cdot\|_D$ is actually a norm having the property $\|\cdot\|_\infty \le \|\cdot\|_D \le \|\cdot\|_1$, cf. Hofmann (2009, Lemma 5.1.3). The bounds are D-norms themselves with generators $\boldsymbol{Z}^{(\infty)} = \boldsymbol{1}$ and $\boldsymbol{Z}^{(1)}$ satisfying $\mathrm{P}(\boldsymbol{Z}^{(1)} = d\,\mathbf{e}_i) = \frac{1}{d}$, $i = 1, \dots, d$. Moreover, for any $p \in (1, \infty)$ the p-norm $\|\cdot\|_p$ is generated by $\boldsymbol{Z}^{(p)} = \frac{1}{\Gamma(1-p^{-1})}(X_1, \dots, X_d)^\intercal$ where Γ denotes the gamma function and $X_1, \dots, X_d$ are independent and Fréchet-distributed with parameter p, i.e. $\mathrm{P}(X_1 \le x) = \exp(-x^{-p})$ for $x > 0$. Note that there is a one-to-one relation between D-norms and standard MSDs:

Theorem 1.1.26 (Balkema–de Haan–Resnick–Vatan)

(i) For any standard MSD G there is a D-norm $\|\cdot\|_D$ such that

$$G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_D) \quad \textit{for all} \quad \boldsymbol{x} \in (-\boldsymbol{\infty}, \boldsymbol{0}]. \tag{1.1.27}$$

Conversely, each D-norm $\|\cdot\|_D$ defines a standard MSD G by (1.1.27).

(ii) Let $\|\cdot\|$ be any norm on $\mathbb{R}^d$. For each D-norm $\|\cdot\|_D$ on $\mathbb{R}^d$ there exist $r > 0$ and a generator $\boldsymbol{Z}$ of $\|\cdot\|_D$ such that $\|\boldsymbol{Z}\| = r$ with probability one. Moreover, r and the distribution of $\boldsymbol{Z}$ are uniquely determined.

The previous result was derived from Vatan (1985, Theorem 3.9), which itself is stated for simple MSDs with additional scaling parameters. Note that a generator $\boldsymbol{Z}$ of a

D-norm $\|\cdot\|_D$ is not uniquely determined in general since $X\boldsymbol{Z}$ is a generator of $\|\cdot\|_D$ as well whenever $X \geq 0$ is a univariate random variable with $\mathrm{E}(X) = 1$ such that X and $\boldsymbol{Z}$ are independent. Moreover, two equivalent generators do not necessarily have the same copula:

Example 1.1.28 Consider $d \geq 2$ independent and identically gamma distributed random variables $V_1, \dots, V_d$, i. e. there is some $\alpha > 0$ such that each V_i has the Lebesgue-density $\gamma_\alpha(v) = \frac{v^{\alpha-1}}{\Gamma(\alpha)} \exp(-v)\, \mathbb{1}_{[0,\infty)}(v)$. Then both,

$$\boldsymbol{Z}^{(1)} = \left(\frac{d\, V_i}{\sum_{j=1}^d V_j} \right)_{i=1}^d \quad \text{and} \quad \boldsymbol{Z}^{(2)} = \left(\frac{V_i}{\alpha} \right)_{i=1}^d,$$

are generators, and the independence of $\frac{1}{d}\boldsymbol{Z}^{(1)}$ and $\sum_{j=1}^d V_j$ shows

$$\mathrm{E}\Big(\|\boldsymbol{x}\boldsymbol{Z}^{(1)}\|_\infty\Big) = \frac{\mathrm{E}(\sum_{j=1}^d V_j)}{d\,\alpha}\, \mathrm{E}\Big(\|\boldsymbol{x}\boldsymbol{Z}^{(1)}\|_\infty\Big) = \mathrm{E}\Big(\|\boldsymbol{x}\boldsymbol{Z}^{(2)}\|_\infty\Big) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^d.$$

We refer to Aulbach et al. (2015*b*, Section 4) for details.

Although Theorem 1.1.26 (ii) yields that the distribution of $\boldsymbol{Z}$ is unique if $\|\boldsymbol{Z}\|$ is almost surely constant, it is in general a non-trivial task to compute a generator with this property. But once r and the distribution $(\mathrm{P} * \boldsymbol{Z})$ in Theorem 1.1.26 (ii) have been identified, a simple integral transformation yields a $(\mathrm{P} * \boldsymbol{Z})$-density of an equivalent generator $\boldsymbol{Z}^*$ which is almost surely constant with respect to another norm $\|\cdot\|^*$, cf. Beirlant et al. (2004, Section 8.2.3).

Example 1.1.29

(i) Every D-norm on $\mathbb{R}^d$ has a generator $\boldsymbol{Z}$ such that $\|\boldsymbol{Z}\|_1 = d$ with probability one.

(ii) For each finite dimensional D-norm $\|\cdot\|_D$ there is a generator $\boldsymbol{Z}$ and some $r > 0$ such that $\|\boldsymbol{Z}\|_\infty = r$ with probability one. We have in particular $\|\mathbf{1}\|_D = \mathrm{E}(\|\boldsymbol{Z}\|_\infty) = r$.

(iii) Apart from the cases $p = 1$ and $p = \infty$ there is — to the best of the author's knowledge — no generator $\boldsymbol{Z}$ of $\|\cdot\|_p$ known such that $\|\boldsymbol{Z}\|$ is almost surely constant, no matter how $\|\cdot\|$ is chosen.

Remark 1.1.30 Consider a standard MSD G, a norm $\|\cdot\|$ on $\mathbb{R}^d$, and the positive part of the corresponding unit sphere $\mathcal{S}_+ = \{\boldsymbol{x} \in [\mathbf{0}, \infty) \mid \|\boldsymbol{x}\| = 1\}$. If r and $\boldsymbol{Z}$ are given as in Theorem 1.1.26 (ii), then the finite measure $\sigma = r\,(\mathrm{P} * \boldsymbol{Z})(r\,\cdot)$ on $(\mathcal{S}_+, \mathbb{B}(\mathcal{S}_+))$ is known as the *spectral measure* of G, where $rB := \{\boldsymbol{x} \in \mathbb{R}^d \mid \frac{1}{r}\boldsymbol{x} \in B\}$ for $B \in \mathbb{B}(\mathcal{S}_+)$. The characterization of an MSD in terms of its spectral measure goes back to Balkema and Resnick (1977), de Haan and Resnick (1977), and Vatan (1985). We refer to Vatan (1985) for an extensive historical overview.

The "D" in "D-norm" is an abbreviation for "dependence". Note that a transformation of the margins of a standard MSD as in Remark 1.1.24 does not alter the corresponding EVC, cf. Theorem 1.1.13. Each EVC C has therefore the representation

$$C(\boldsymbol{u}) = \exp(-\|\log(\boldsymbol{u})\|_D) \quad \text{for all} \quad \boldsymbol{u} \in (\boldsymbol{0}, \boldsymbol{1}] \tag{1.1.31}$$

where $\|\cdot\|_D$ denotes a suitable D-norm. As motivated by the following result, the generator constant measures the degree of dependence of the margins of an MSD.

Lemma 1.1.32 (Takahashi, 1988; cf. Falk et al., 2011, Theorem 4.4.1) *Let C be a d-variate EVC with corresponding D-norm $\|\cdot\|_D$. Then:*

(i) $C(\boldsymbol{u}) = \prod_{i=1}^d u_i$ for all $\boldsymbol{u} = (u_1, \dots, u_d)^\intercal \in [\boldsymbol{0}, \boldsymbol{1}]$ if and only if $\|\boldsymbol{1}\|_D = d$.

(ii) $C(\boldsymbol{u}) = \min\{u_1, \dots, u_d\}$ for all $\boldsymbol{u} = (u_1, \dots, u_d)^\intercal \in [\boldsymbol{0}, \boldsymbol{1}]$ if and only if $\|\boldsymbol{1}\|_D = 1$.

Recall that $\|\cdot\|_\infty \leq \|\cdot\|_D \leq \|\cdot\|_1$, i. e. $1 \leq \|\boldsymbol{1}\|_D \leq d$ where the both extreme cases are characterized by Lemma 1.1.32.

Remark 1.1.33 The D-norm of a standard MSD is also known as the *stable tail dependence function* — introduced by Huang (1992) as the limit in (1.1.22) — whereas the generator constant is also called *extremal coefficient*, cf. Smith (1990).

Generalized Pareto Distributions

While MSDs are the natural choice to model the suitably standardized maximum in an i. i. d. sample, breaking all the data down to just one observation, another approach focuses on all "large" data in the sample. Therefore, we say that a vector $\boldsymbol{x} \in \mathbb{R}^d$ *exceeds* a threshold $\boldsymbol{x}_0 \in \mathbb{R}^d$ if $\boldsymbol{x} \not\leq \boldsymbol{x}_0$, i. e. at least one component of $\boldsymbol{x}$ is larger than the corresponding component of $\boldsymbol{x}_0$. The following distributions will be crucial.

Definition 1.1.34 A d-variate distribution function W is referred to as a *generalized Pareto distribution* (GPD) if there exist $\boldsymbol{x}_0 < \boldsymbol{\omega}(W)$ and an MSD G such that $W(\boldsymbol{x}) = 1 + \log(G(\boldsymbol{x}))$ for all $\boldsymbol{x} \geq \boldsymbol{x}_0$. If G is a standard MSD then W is called a *standard* GPD.

This definition is according to Falk et al. (2011, Section 5.1) and extends the one by Kaufmann and Reiss (1995) to arbitrary dimensions. Other definitions are given in Tajvidi (1996, Paper B), Beirlant et al. (2004, Section 8.3.1) and Rootzén and Tajvidi (2006). Note that, after a transformation of the margins, all of the in total three definitions coincide close to the upper endpoint of the distribution; we refer to Michel (2006, Remark 2.2.3) and Beirlant et al. (2004, Section 8.3.1) for details.

In the uni- and bivariate cases $1 + \log(G(\boldsymbol{x}))$, $\log(G(\boldsymbol{x})) \geq -1$, already defines a distribution function, cf. Kaufmann and Reiss (1995) and Falk et al. (2011, Lemma 5.1.1). Although this is not true for $d \geq 3$ — cf. Michel (2008, Theorem 6) and Hofmann (2009, Theorem 2.2.2) — Hofmann (2009, Theorem 6.2.1) and Falk et al. (2011, Lemma 5.1.5) show that for any MSD, there exists a corresponding GPD, and in particular:

Theorem 1.1.35 (Hofmann, 2009; Falk et al., 2011) *Let G be a standard MSD with D-norm $\|\cdot\|_D$. Then there is a corresponding standard GPD W satisfying*

$$W(\boldsymbol{x}) = 1 - \|\boldsymbol{x}\|_D \quad \text{for} \quad \boldsymbol{x} \in \left[-\frac{1}{d}, 0\right]^d. \tag{1.1.36}$$

Furthermore, if $\boldsymbol{X}$ has distribution function G, we obtain

$$\lim_{r \to 0+} \mathrm{P}(\boldsymbol{X} \leq r\boldsymbol{x} \mid \boldsymbol{X} \not\leq r\boldsymbol{t}) = 1 - \frac{\|\boldsymbol{x}\|_D}{\|\boldsymbol{t}\|_D} \quad \text{for} \quad \boldsymbol{x} \in [\boldsymbol{t}, \boldsymbol{0}], \tag{1.1.37}$$

where $\boldsymbol{t} < \boldsymbol{0}$ is chosen arbitrarily.

Equation (1.1.37) suggests to model exceedances over a high threshold by means of a GPD, which is in complete accordance with the univariate results of Balkema and de Haan (1974) and Pickands (1975) — cf. Reiss and Thomas (2007, Section 1.4) — and the multivariate ones by Tajvidi (1996, Paper B, Section 4), Beirlant et al. (2004, Section 8.3.1) and Rootzén and Tajvidi (2006). For an approach that takes this into account we refer to Aulbach et al. (2012*a*) and Aulbach et al. (2012*b*).

In particular, if $\boldsymbol{X}$ has the distribution function W in (1.1.36), then

$$\mathrm{P}(\boldsymbol{X} \leq r\boldsymbol{x} \mid \boldsymbol{X} \not\leq r\boldsymbol{t}) = 1 - \frac{\|\boldsymbol{x}\|_D}{\|\boldsymbol{t}\|_D} \quad \text{for all} \quad \boldsymbol{x} \in [\boldsymbol{t}, \boldsymbol{0}] \tag{1.1.38}$$

whenever $r > 0$ and $\boldsymbol{t} < \boldsymbol{0}$ satisfy $r\boldsymbol{t} \in \left[-\frac{1}{d}, 0\right]^d$. In this case, we end up with the *excursion stability* of a standard GPD

$$\mathrm{P}\left(\frac{\boldsymbol{X}}{\|\boldsymbol{t}\|_D} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X} \not\leq \boldsymbol{t}\right) = 1 - \|\boldsymbol{x}\|_D \quad \text{for} \quad \boldsymbol{t} \in \left[-\frac{1}{d}, 0\right]^d \text{ and } \boldsymbol{x} \in [\boldsymbol{t}, \boldsymbol{0}],$$

cf. de Haan and Ferreira (2006), i. e. the conditional distribution of $\frac{\boldsymbol{X}}{\|\boldsymbol{t}\|_D}$ given $\boldsymbol{X} \not\leq \boldsymbol{t}$ coincides with the distribution of $\boldsymbol{X}$ in the upper tail. For further concepts of excursion stability, we refer to Falk et al. (2011, Sections 5.3 and 5.4) and Falk and Guillou (2008).

Taking Lemma 1.1.20 into account, we obtain a characterization of the domain of attraction condition of Theorem 1.1.18 in terms of D-norms. All relevant components — MSD, GPD, and the corresponding domain of attraction — can thus be broken down to certain D-norm conditions, making these norms quite an appealing tool.

Theorem 1.1.39 (Aulbach et al., 2012*a*, Corollary 2.2) *Let F be a d-variate distribution function with copula C_F and margins $F_1, \dots, F_d$. If G is an MSD with copula $C_G = \exp(-\|\log(\cdot)\|_D)$ and margins $G_1, \dots, G_d$ such that $F_i \in \mathcal{D}(G_i)$ for each $i \in \{1, \dots, d\}$, then the following assertions hold:*

(i) $F \in \mathcal{D}(G)$ implies $C_F(\boldsymbol{u}) = 1 - \|\boldsymbol{u} - \boldsymbol{1}\|_D + r(\boldsymbol{u})$, $\boldsymbol{u} \in [\boldsymbol{0}, \boldsymbol{1}]$, where the remainder satisfies

$$r(\boldsymbol{1}) = 0 \qquad \text{and} \qquad \lim_{t \to 0+} \sup_{\substack{\boldsymbol{u} \in [0,1]^d \setminus \{\boldsymbol{1}\} \\ \|\boldsymbol{u} - \boldsymbol{1}\| < t}} \frac{|r(\boldsymbol{u})|}{\|\boldsymbol{u} - \boldsymbol{1}\|} = 0 \tag{1.1.40}$$

for an arbitrary norm $\|\cdot\|$.

(ii) If there is some norm $\|\cdot\|^*$ such that $C_F(\boldsymbol{u}) = 1 - \|\boldsymbol{u} - \boldsymbol{1}\|^* + r(\boldsymbol{u})$, $\boldsymbol{u} \in [\boldsymbol{0}, \boldsymbol{1}]$, where the remainder satisfies (1.1.40) *for some norm $\|\cdot\|$, then $F \in \mathcal{D}(G)$ and $\|\cdot\|_D = \|\cdot\|^*$.*

We close this section stating a simple method of computing a random vector that follows a standard GPD. It was proved in Buishand et al. (2008, Section 2.2) for the bivariate case and extended to an arbitrary dimension by Aulbach et al. (2012*a*, Proposition 2.4).

Theorem 1.1.41 (Buishand et al., 2008; Aulbach et al., 2012*a*)

(i) Let W be a d-variate standard GPD with corresponding D-norm $\|\cdot\|_D$. Then there are a generator $\boldsymbol{Z} = (Z_1, \dots, Z_d)^\intercal$ of $\|\cdot\|_D$ satisfying $\mathrm{P}(\boldsymbol{Z} \leq d\boldsymbol{1}) = 1$ and a vector $\boldsymbol{x}_0 \in [-\frac{1}{d}, 0)^d$ such that

$$W(\boldsymbol{x}) = \mathrm{P}\left(-U \frac{\boldsymbol{1}}{\boldsymbol{Z}} \leq \boldsymbol{x}\right) \quad \text{for all} \quad \boldsymbol{x} \in [\boldsymbol{x}_0, \boldsymbol{0}]$$

where the random variable $U \sim \boldsymbol{\mathcal{U}}[0,1]$ is independent of $\boldsymbol{Z}$.

(ii) Let $U \sim \boldsymbol{\mathcal{U}}[0,1]$ be independent of a generator $\boldsymbol{Z}$. If there is a vector $\boldsymbol{c} \geq \boldsymbol{1}$ such that $\mathrm{P}(\boldsymbol{Z} \leq \boldsymbol{c}) = 1$, then the random random vector $-U \frac{\boldsymbol{1}}{\boldsymbol{Z}}$ follows a standard GPD and the corresponding D-norm is given by $\boldsymbol{Z}$.

If one drops the condition that the generator is bounded, then the distribution function of $-U\frac{\mathbf{1}}{\mathbf{Z}}$ is *not* a GPD itself but somewhat close to a standard GPD, cf. Theorem 1.2.26 below. Note that, in order to avoid dividing by zero, $-U\frac{\mathbf{1}}{\mathbf{Z}}$ may be substituted by $\max\{m\mathbf{1}, -U\frac{\mathbf{1}}{\mathbf{Z}}\}$ for an arbitrary negative constant $m < 0$.

Although Theorem 1.1.41 can be used to carry over the term "GPD" into the space of continuous functions, cf. Buishand et al. (2008, Section 2.3), the following section deals with a slightly different approach.

1.2 Extreme Value Theory in $C[0,1]$

Now we extend the results of the previous section to an uncountably infinite number of dimensions, namely the space $C[0,1]$ of continuous real-valued functions defined on the unit interval $[0,1] \subset \mathbb{R}$. Note that the distribution of a stochastic process $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ with continuous sample paths is determined by its finite dimensional projections. The identity

$$P(X_{t_1} \le x_1, \dots, X_{t_d} \le x_d) = \lim_{n\to\infty} P\left(\boldsymbol{X} \le \sum_{i=1}^{d} x_i \mathbb{1}_{\{t_i\}} + n\mathbb{1}_{[0,1]\setminus\{t_1,\dots,t_d\}}\right),$$

which holds for all $d \in \mathbb{N}$ and $(t_1, x_1), \dots, (t_d, x_d) \in [0,1] \times \mathbb{R}$, suggests therefore to define the distribution function of $\boldsymbol{X}$ as follows.

Definition 1.2.1 Let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ and put

$$\mathcal{E}[0,1] := \{f \in \mathbb{R}^{[0,1]} \mid f \text{ is bounded and has a finite set of discontinuities}\}.$$

Then we call the function $F : \mathcal{E}[0,1] \to [0,1]$ defined by $F(f) := P(\boldsymbol{X} \le f)$ the *distribution function* of $\boldsymbol{X}$. Furthermore the distribution function F_t of X_t is referred to as the t-th *margin* of F for $t \in [0,1]$. We say that F is *non-degenerate* if all of its margins are non-degenerate.

The definition of max-stability carries over, cf. Giné et al. (1990):

Definition 1.2.2 Let $\boldsymbol{\eta} = (\eta_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ such that its distribution function G is non-degenerate. $\boldsymbol{\eta}$ is called a *max-stable process* (MSP) and G a *max-stable distribution* (MSD) if for each $n \in \mathbb{N}$ there exist functions $a_n \in C[0,1] \cap (0,\infty)^{[0,1]}$ and $b_n \in C[0,1]$ such that

$$\frac{\max_{1\le i\le n}\{\boldsymbol{\eta}^{(i)}\} - b_n}{a_n} \overset{D}{=} \boldsymbol{\eta}$$

where $\boldsymbol{\eta}^{(1)}, \boldsymbol{\eta}^{(2)}, \dots$ are independent copies of $\boldsymbol{\eta}$. If in particular $P(\eta_t \le x) = \exp(x)$ for all $x \leq 0$ and $t \in [0,1]$, then $\boldsymbol{\eta}$ and G are referred to as a *standard* MSP and a *standard* MSD, respectively.

Remark 1.2.3 From Giné et al. (1990, Corollary 3.4) — cf. Hofmann (2012, Lemma 2.2) and de Haan and Ferreira (2006, Theorem 9.4.1) — we know

$$\mathrm{P}\left(\inf_{t\in[0,1]} \xi_t > 0\right) = 1 = \mathrm{P}\left(\sup_{t\in[0,1]} \eta_t < 0\right)$$

whenever $\boldsymbol{\xi}$ is a *simple* MSP, i. e. $\mathrm{P}(\xi_t \leq x) = \exp(-\frac{1}{x})$ for all $x > 0$ and $t \in [0,1]$, and $\boldsymbol{\eta}$ is a standard MSP. In this case $-\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{\xi}}$ and $-\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{\eta}}$ are standard max-stable and simple max-stable, respectively, cf. Remark 1.1.24.

As before, standard MSDs will be characterized by means of D-norms, which are defined analogously to Definition 1.1.25.

Definition 1.2.4 Let $\boldsymbol{Z}$ be a stochastic process in $C[0,1]\cap[0,\infty)^{[0,1]}$ that satisfies $\mathrm{E}(\boldsymbol{Z}) = \mathbb{1}_{[0,1]}$ and $\mathrm{E}(\|\boldsymbol{Z}\|_\infty) < \infty$. Then $\|\cdot\|_D$ defined by

$$\|f\|_D := \mathrm{E}(\|f\,\boldsymbol{Z}\|_\infty) \quad \text{for all} \quad f \in \mathcal{E}[0,1]$$

is called a *D-norm* with *generator* $\boldsymbol{Z}$. Furthermore, two generators are *equivalent* if they give rise to the same D-norm. The value $\|\mathbb{1}_{[0,1]}\|_D = \mathrm{E}(\|\boldsymbol{Z}\|_\infty)$ is also referred to as the *generator constant* of $\boldsymbol{Z}$.

Note the additional requirement $\mathrm{E}(\|\boldsymbol{Z}\|_\infty) < \infty$ of a generator $\boldsymbol{Z}$, which is trivial if $\boldsymbol{Z}$ is finite dimensional; cf. de Haan and Ferreira (2006, Corollary 9.4.5). As in the finite dimensional setting, it is easily verified that any D-norm $\|\cdot\|_D$ is a norm satisfying

$$\|\cdot\|_\infty \leq \|\cdot\|_D \leq \|\mathbb{1}_{[0,1]}\|_D \,\|\cdot\|_\infty\,, \tag{1.2.5}$$

cf. Hofmann (2012, Lemma 2.6). Giné et al. (1990, Proposition 3.2) implies furthermore a functional version of Theorem 1.1.26:

Theorem 1.2.6 (Giné et al., 1990)

(i) For any standard MSD G there is a D-norm $\|\cdot\|_D$ such that

$$G(f) = \exp(-\|f\|_D) \quad \textit{for all} \quad f \in \mathcal{E}[0,1] \cap (-\infty,0]^{[0,1]}. \tag{1.2.7}$$

Conversely, each D-norm $\|\cdot\|_D$ defines a standard MSD G by (1.2.7).

(ii) *For each D-norm $\|\cdot\|_D$ there exist $r > 0$ and a generator $\boldsymbol{Z}$ of $\|\cdot\|_D$ such that $\|\boldsymbol{Z}\|_\infty = r$ with probability one.*

Contrary to Theorem 1.1.26, this result makes no assertion whether r and the distribution of $\boldsymbol{Z}$ in (ii) are uniquely determined. Furthermore, (ii) just considers the sup-norm $\|\cdot\|_\infty$ instead of an arbitrary norm $\|\cdot\|$, which seems to be less general. It is, however, natural to restrict oneself to the sup-norm since the open balls with respect to $\|\cdot\|_\infty$ generate the underlying Borel-σ-algebra $\mathbb{B}(C[0,1])$; while all norms are equivalent in the finite dimensional framework, this is not true in $C[0,1]$. Equation (1.2.5) shows moreover that all D-norms are equivalent to the sup-norm.

Copula Processes and the Domain of Attraction

In Section 1.1 copulas have been quite a useful tool to characterize the domain of attraction of an MSD, and thus motivated to consider standard MSDs only. Now we deal with a functional extension.

Definition 1.2.8 A stochastic process $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ in $\mathbb{R}^{[0,1]}$ is called a *copula process* if $U_t \sim \mathcal{U}[0,1]$ holds for all $t \in [0,1]$. We say that $\boldsymbol{U}$ is a copula process of a stochastic process $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ if $\left(F_t^{-1}(U_t)\right)_{t\in[0,1]}$ has the same distribution as $\boldsymbol{X}$, where F_t^{-1} denotes the quantile function of X_t.

Although it may appear natural to require that a copula process of a sample continuous process is in $C[0,1]$ as well, we will see in Section 3.3 that there are rather simple processes that do not have a continuous copula process. However, any process in $C[0,1]$ with continuous marginal distributions does have a copula process in $C[0,1]$, which is easy to prove.

Lemma 1.2.9 *Let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ where F_t is the distribution function of X_t, $t \in [0,1]$. If all F_t, $t \in [0,1]$, are continuous, then $\boldsymbol{U} = (F_t(X_t))_{t\in[0,1]}$ is a copula process of $\boldsymbol{X}$, and $\boldsymbol{U}$ is in $C[0,1]$.*

In particular, if $\boldsymbol{X}$ and $\boldsymbol{U}$ are given as in Lemma 1.2.9, then $\boldsymbol{X}$ can be reconstructed from $\boldsymbol{U}$ and F_t, $t \in [0,1]$: Since all margins F_t, $t \in [0,1]$, are continuous, we have $\left(F_q^{-1}(U_q)\right)_{q\in\mathbb{Q}\cap[0,1]} = (X_q)_{q\in\mathbb{Q}\cap[0,1]}$ with probability one. But then $\boldsymbol{X}$ is already completely determined because all of its sample paths are continuous.

Definition 1.2.10 Denote by $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ a stochastic process in $C[0,1]$ and by F the distribution function of $\boldsymbol{X}$. If all margins F_t, $t \in [0,1]$, of F are continuous, then we call $(F_t(X_t))_{t\in[0,1]}$ *the* copula process of $\boldsymbol{X}$, and its distribution function the *copula* of F.

As shown by Giné et al. (1990, Corollary 3.6) and Hofmann (2012, Proposition 2.10), any MSD can be transformed into a standard MSD and vice versa. This is done by transforming the margins, whereas the dependence structure between the margins remains the same; cf. the discussion following Remark 1.1.30. The copula of an MSD depends therefore on the corresponding D-norm but *not* on the margins of that MSD. According to (1.1.31), we define a functional extreme value copula as follows, cf. Ribatet and Sedki (2013, Section 3).

Definition 1.2.11 A copula C is called an *extreme value copula* (EVC) if it has the representation

$$C(f) = \exp(-\|\log(f)\|_D) \quad \text{for all} \quad f \in \mathcal{E}[0,1] \cap (0,1)^{[0,1]}$$

with respect to some D-norm $\|\cdot\|_D$.

While the case of complete dependence is characterized by the condition $\|\mathbb{1}_{[0,1]}\|_D = 1$, which is analogous to Section 1.1, there is no standard MSP in $C[0,1]$ which corresponds to the case of independence. In fact it is checked easily that $\{\|\mathbb{1}_{[0,1]}\|_D \mid \|\cdot\|_D \text{ is a } D\text{-norm}\} = [1,\infty)$.

Lemma 1.2.12 (Hofmann, 2012, Lemma 2.12; cf. Lemma 1.1.32) *Any D-norm $\|\cdot\|_D$ satisfies $\|\cdot\|_D = \|\cdot\|_\infty$ if and only if $\|\mathbb{1}_{[0,1]}\|_D = 1$. Moreover, for any generator $\boldsymbol{Z}$ of $\|\cdot\|_\infty$ there is some univariate random variable $Z \geq 0$ satisfying $\mathrm{E}(Z) = 1$ such that $\boldsymbol{Z} = Z\,\mathbb{1}_{[0,1]}$ with probability one. Similarly, if $\boldsymbol{\eta}$ is a standard MSP with D-norm $\|\cdot\|_\infty$, then $\boldsymbol{\eta} = \eta\,\mathbb{1}_{[0,1]}$ with probability one where η is a standard negative exponential random variable.*

As in the multivariate context, copulas can be used to characterize the domain of attraction of a functional MSD. While convergence of distribution functions, for all continuity points of the limit distribution, is equivalent to weak convergence in the finite dimensional case, the extension to function space is twofold: The domain of attraction of an MSD may be defined by weak convergence of suitably standardized maxima, or by convergence of their corresponding distribution functions.

Definition 1.2.13 (Aulbach et al., 2013) Let F be the distribution function of some stochastic process $\boldsymbol{X}$ in $C[0,1]$. If G is an MSD with corresponding MSP $\boldsymbol{\eta}$ and if for each $n \in \mathbb{N}$ there are $a_n \in C[0,1] \cap (0,\infty)^{[0,1]}$ and $b_n \in C[0,1]$ such that

$$F^n(a_n f + b_n) \to_{n\to\infty} G(f) \quad \text{for all} \quad f \in \mathcal{E}[0,1] \tag{1.2.14}$$

then F is in the *domain of attraction* of G and we write $F \in \mathcal{D}(G)$ or $\boldsymbol{X} \in \mathcal{D}(\boldsymbol{\eta})$. If $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots$ are independent copies of $\boldsymbol{X}$ and

$$\frac{\max_{1\le i\le n}\{\boldsymbol{X}^{(i)}\} - b_n}{a_n} \xrightarrow{D} \boldsymbol{\eta} \quad \text{as} \quad n\to\infty \tag{1.2.15}$$

then we write $F \in \mathcal{D}_w(G)$ or $\boldsymbol{X} \in \mathcal{D}_w(\boldsymbol{\eta})$, according to de Haan and Lin (2001).

Note that $F \in \mathcal{D}_w(G)$ — i. e. the standardized maximum of the $\boldsymbol{X}^{(i)}$ converges weakly to $\boldsymbol{\eta}$ — is a sufficient condition for (1.2.14). We refer to Hofmann (2012) for a comparison of $F \in \mathcal{D}(G)$, $F \in \mathcal{D}_w(G)$, and other types of convergence.

The domain of attraction condition (1.2.14) can be decomposed into a condition on the margins and a copula condition, yielding a functional analogue of Theorem 1.1.18 and Lemma 1.1.20; see Theorem 1.2.18 below. A similar decomposition of (1.2.15) can be found in de Haan and Ferreira (2006, Theorem 9.2.1), cf. de Haan and Lin (2001, Theorem 2.8), and is stated here for easier reference. In the following, however, we will focus on the more general type of convergence, i. e. (1.2.14).

Theorem 1.2.16 (de Haan and Ferreira, 2006; cf. de Haan and Lin, 2001) *Let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$, and denote its distribution function by F. Suppose that all margins F_t, $t\in[0,1]$, of F are continuous. Moreover, let $\boldsymbol{U}$ be the copula process of $\boldsymbol{X}$. If $\boldsymbol{\eta}$ is an MSP with corresponding MSD G and margins G_t, $t\in[0,1]$, then the following assertions hold:*

(i) If $\boldsymbol{X} \in \mathcal{D}_w(\boldsymbol{\eta})$ then $\boldsymbol{U} \in \mathcal{D}_w(\bar{\boldsymbol{\eta}})$, where $\bar{\boldsymbol{\eta}} = \big(\log(G_t(\eta_t))\big)_{t\in[0,1]}$ is a standard MSP, and there are functions $a_n \in C[0,1]\cap(0,\infty)^{[0,1]}$ and $b_n \in C[0,1]$ such that

$$\sup_{t\in[0,1]} \left|F_t^n(a_n(t)\,x + b_n(t)) - G_t(x)\right| \to_{n\to\infty} 0 \quad \textit{for all} \quad x\in\mathbb{R}. \tag{1.2.17}$$

(ii) If $\boldsymbol{U} \in \mathcal{D}_w(\bar{\boldsymbol{\eta}})$ for some standard MSP $\bar{\boldsymbol{\eta}}$ and if (1.2.17) holds for some $a_n \in C[0,1]\cap(0,\infty)^{[0,1]}$ and $b_n\in C[0,1]$, then we have $\boldsymbol{X}\in\mathcal{D}_w(\boldsymbol{\eta})$ and $\boldsymbol{\eta} \overset{D}{=} \big(G_t^{-1}(\exp(\bar{\eta}_t))\big)_{t\in[0,1]}$.

For the more general case, the prerequisites are somewhat stricter:

Theorem 1.2.18 (Aulbach et al., 2015*a*, Theorem 2.1) *Let $\boldsymbol{X}$ be a stochastic process in $C[0,1]$ with distribution function F, continuous margins F_t, $t\in[0,1]$, and copula C. Moreover, let $\boldsymbol{\eta}$ be an MSP with corresponding MSD G and margins G_t, $t\in[0,1]$. Suppose that there are functions $a_n\in C[0,1]\cap(0,\infty)^{[0,1]}$ and $b_n\in C[0,1]$ satisfying*

$$\sup_{t\in[0,1]} \left|F_t^n(a_n(t)\,f(t) + b_n(t)) - G_t(f(t))\right| \to_{n\to\infty} 0 \quad \textit{for all} \quad f\in\mathcal{E}[0,1]\,. \tag{1.2.19}$$

Then we have:

(i) *If* $F \in \mathcal{D}(G)$, *then*

$$C^n\left(\mathbb{1}_{[0,1]} + \frac{1}{n} f\right) \to \exp(-\|f\|_D) \quad as \quad n \to \infty \tag{1.2.20}$$

holds for all $f \in \mathcal{E}[0,1] \cap (-\infty, 0]^{[0,1]}$ *where* $\|\cdot\|_D$ *is the D-norm of the standard MSP* $(\log(G_t(\eta_t)))_{t\in[0,1]}$.

(ii) *If* (1.2.20) *holds for some D-norm* $\|\cdot\|_D$ *and all* $f \in \mathcal{E}[0,1] \cap (-\infty, 0]^{[0,1]}$, *then* $F \in \mathcal{D}(G)$. *In this case we have* $\boldsymbol{\eta} \stackrel{D}{=} (G_t^{-1}(\exp(\bar{\eta}_t)))_{t\in[0,1]}$ *for any standard MSP* $(\bar{\eta}_t)_{t\in[0,1]}$ *with D-norm* $\|\cdot\|_D$.

Similar to Section 1.1, both preceding results decompose the corresponding domain of attraction condition into a condition on the margins and a condition on the copula. In the following we will focus on the copula process condition (1.2.20), which means to consider standard MSDs and their D-norms.

Generalized Pareto Distributions

As in the finite dimensional framework one obtains a suitable model for large observations, i. e. data exceeding a high threshold, by considering generalized Pareto processes, which are defined analogously to Definition 1.1.34:

Definition 1.2.21 A distribution function W on $\mathcal{E}[0,1]$ is called a *standard generalized Pareto distribution* (GPD) if there are a D-norm $\|\cdot\|_D$ and a constant $x_0 < 0$ such that $W(f) = 1 - \|f\|_D$ holds for all $f \in \mathcal{E}[0,1] \cap [x_0, 0]^{[0,1]}$. In this case, a stochastic process in $C[0,1]$ with distribution function W is referred to as a *standard generalized Pareto process* (GPP).

The both results from Theorem 1.1.35 and Theorem 1.1.41 carry over to stochastic processes. The proof of the functional version of Theorem 1.1.35 is easy, and therefore omitted.

Theorem 1.2.22 *Let* $\boldsymbol{\eta}$ *be a standard MSP with D-norm* $\|\cdot\|_D$. *The following assertions hold:*

(i) *If* $\boldsymbol{Z}$ *is a generator of* $\|\cdot\|_D$ *that satisfies* $\mathrm{P}(\boldsymbol{Z} \le c\mathbb{1}_{[0,1]}) = 1$ *for some* $c \ge 1$, *there is a standard GPD W such that*

$$W(f) = 1 - \|f\|_D \quad for \quad f \in \mathcal{E}[0,1] \cap \left[-\frac{1}{c}, 0\right]^{[0,1]}.$$

(ii) *For each* $g \in \mathcal{E}[0,1] \cap (-\infty,0]^{[0,1]} \setminus \{0\}^{[0,1]}$ *we have*

$$\lim_{r\to 0+} \mathrm{P}(\boldsymbol{\eta} \le r\, f \mid \boldsymbol{\eta} \not\le r\, g) = 1 - \frac{\|f\|_D}{\|g\|_D} \quad \text{for} \quad f \in \mathcal{E}[0,1] \cap [g, \mathbb{0}_{[0,1]}]$$

(iii) *If* $\boldsymbol{V}$ *is a standard GPP with D-norm* $\|\cdot\|_D$*, there is for each* $g \in \mathcal{E}[0,1] \cap (-\infty,0]^{[0,1]} \setminus \{0\}^{[0,1]}$ *some* $r_0 > 0$ *such that*

$$\mathrm{P}(\boldsymbol{V} \le r\, f \mid \boldsymbol{V} \not\le r\, g) = 1 - \frac{\|f\|_D}{\|g\|_D} \quad \text{for} \quad f \in \mathcal{E}[0,1] \cap [g, \mathbb{0}_{[0,1]}] \text{ and } r \in (0, r_0].$$

Note that part (iii) is the excursion stability of a functional GPD, cf. (1.1.38). The generation of a stochastic process that follows a given GPD, cf. Theorem 1.1.41, has already been considered in Aulbach et al. (2012*b*, Section 4):

Theorem 1.2.23 (Aulbach et al., 2012*b*)

(i) *For any D-norm* $\|\cdot\|_D$ *there exist* $x_0 \in [-1,0)$ *and a generator* $\boldsymbol{Z}$ *of* $\|\cdot\|_D$ *satisfying* $\mathrm{P}\big(\boldsymbol{Z} \le \frac{1}{|x_0|} \mathbb{1}_{[0,1]}\big) = 1$ *such that*

$$\mathrm{P}\left(-U \frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}} \le f\right) = 1 - \|f\|_D \quad \text{for all} \quad f \in \mathcal{E}[0,1] \cap [x_0, 0]^{[0,1]} \tag{1.2.24}$$

where the random variable $U \sim \mathcal{U}[0,1]$ *is independent of* $\boldsymbol{Z}$*.*

(ii) *Let* $U \sim \mathcal{U}[0,1]$ *be independent of a generator* $\boldsymbol{Z}$ *with corresponding D-norm* $\|\cdot\|_D$*. If there is a function* $g \in \mathcal{E}[0,1] \cap [1,\infty)^{[0,1]}$ *such that* $\mathrm{P}(\boldsymbol{Z} \le g) = 1$*, then* (1.2.24) *holds for some* $x_0 < 0$*.*

Note that $-U\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}}$ takes values in $[-\infty,0]^{[0,1]} \setminus \{-\infty\}^{[0,1]}$, which is not a subset of $C[0,1]$. In order to obtain a standard GPP, one may cut off the lower part of this process: If $h \in C[0,1] \cap (-\infty,0)^{[0,1]}$ then

$$\boldsymbol{V} := \max\left\{h, -U\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}}\right\} \tag{1.2.25}$$

is a standard GPP which corresponds to the D-norm generated by $\boldsymbol{Z}$.

Since Theorem 1.2.23 focuses on almost surely bounded generators, Example 1.1.28 raises the question whether there is a similar result for unbounded generators. The following simple extension of Theorem 1.2.23 shows that the above boundary condition is crucial, and slightly sharpens the area in which the representation $\mathrm{P}\big(-U\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}} \le \cdot\big) = 1 - \|\cdot\|_D$ holds; cf. the proof of Hofmann (2009, Theorem 6.2.1).

Theorem 1.2.26 *Consider an arbitrary generator $\boldsymbol{Z} = (Z_t)_{t\in[0,1]}$ together with its D-norm $\|\cdot\|_D$. If $U \sim \boldsymbol{\mathcal{U}}[0,1]$ is independent of $\boldsymbol{Z}$, then we have*

$$\mathrm{P}\left(-U\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}} \le f\right) = 1 - \|f\|_D + r(f) \quad \text{for} \quad f \in \mathcal{E}[0,1] \cap (-\infty,0]^{[0,1]}$$

where the remainder $r(f)$ is non-negative and satisfies

$$\sup_{g\in\mathcal{E}[0,1]\cap[f,\mathbb{0}_{[0,1]}]} r(g) \to 0 \quad \text{as} \quad \|f\|_\infty \to 0. \tag{1.2.27}$$

Furthermore $\mathrm{P}(\boldsymbol{Z} \le h) = 1$ holds for some $h \in \mathcal{E}[0,1] \cap [1,\infty)^{[0,1]}$ if and only if $r(f) = 0$ for all $f \in \mathcal{E}[0,1] \cap \left[-\frac{\mathbb{1}_{[0,1]}}{h}, \mathbb{0}_{[0,1]}\right]$.

Proof. Since U and $\boldsymbol{Z}$ are independent, we get

$$\begin{aligned}
r(f) &:= \mathrm{P}\left(-U\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}} \le f\right) - 1 + \|f\|_D \\
&= \mathrm{P}(U \ge \|f\,\boldsymbol{Z}\|_\infty) - \mathrm{E}(1 - \|f\,\boldsymbol{Z}\|_\infty) \\
&= \int_{C[0,1]} 1 - \mathrm{P}(U \le \|f\,z\|_\infty)\,(\mathrm{P} * \boldsymbol{Z})(\mathrm{d}z) - \mathrm{E}(1 - \|f\,\boldsymbol{Z}\|_\infty) \\
&= \mathrm{E}\Big((1 - \|f\,\boldsymbol{Z}\|_\infty)\,\mathbb{1}_{[0,1]}(\|f\,\boldsymbol{Z}\|_\infty)\Big) - \mathrm{E}(1 - \|f\,\boldsymbol{Z}\|_\infty) \\
&= \mathrm{E}\big[(\|f\,\boldsymbol{Z}\|_\infty - 1)\,\mathbb{1}_{(1,\infty)}(\|f\,\boldsymbol{Z}\|_\infty)\big] \ge 0 \quad \text{for} \quad f \in \mathcal{E}[0,1] \cap (-\infty,0]^{[0,1]}.
\end{aligned}$$

Because of $\left|(\|f\,\boldsymbol{Z}\|_\infty - 1)\,\mathbb{1}_{(1,\infty)}(\|f\,\boldsymbol{Z}\|_\infty)\right| \le 1 + \|f\,\boldsymbol{Z}\|_\infty \le 1 + \|\boldsymbol{Z}\|_\infty$ for all $f \in \mathcal{E}[0,1] \cap \left[-\mathbb{1}_{[0,1]}, \mathbb{0}_{[0,1]}\right]$, the dominated convergence theorem implies (1.2.27). Furthermore we have

$$r\left(-\frac{\mathbb{1}_{[0,1]}}{h}\right) = 0 \quad \text{for some} \quad h \in \mathcal{E}[0,1] \cap [1,\infty)^{[0,1]}$$

if and only if

$$1 = \mathrm{P}\left(\left\|-\frac{\mathbb{1}_{[0,1]}}{h}\boldsymbol{Z}\right\|_\infty \le 1\right) = \mathrm{P}(\boldsymbol{Z} \le h),$$

which completes the proof. □

Although our definition of a (standard) GPP differs from the one introduced by Buishand et al. (2008, Section 2.3), cf. Aulbach et al. (2013, Examples 1 and 5), Theorem 1.2.23 shows that one might switch between both definitions without loss of generality. The difference is that in the sense of Definition 1.2.21, a standard GPP is any stochastic process in $C[0,1]$ whose distribution function has the representation $1 - \|\cdot\|_D$ in its upper tail, whereas the definition in Buishand et al. (2008, Section 2.3) considers processes of the form (1.2.25) only, cf. Aulbach et al. (2012*b*). We also refer to Ferreira and de Haan (2014), Aulbach et al. (2015*a*), and Dombry and Ribatet (2015) for further recent results on GPPs.

In the finite dimensional case the domain of attraction has been characterized by univariate domain of attraction conditions together with a GPD-approximation of a copula, cf. Theorem 1.1.39. While Theorem 1.2.18 addresses an analogous decomposition into a copula condition and some kind of uniform domain of attraction of the margins, a corresponding GPD-approximation is easily obtained from (1.2.20), cf. Aulbach et al. (2013, Proposition 8).

Theorem 1.2.28 (Aulbach et al., 2013) *For a functional copula C and a standard MSD G with D-norm $\|\cdot\|_D$, the assertions $C \in \mathcal{D}(G)$ and*

$$C\Big(\mathbb{1}_{[0,1]} + t\,f\Big) = 1 - t\,\|f\|_D + o(t) \quad \text{for } f \in \mathcal{E}[0,1] \cap (-\infty, 0]^{[0,1]} \text{ as } t \to 0+$$

are equivalent.

The preceding result is the functional version of Lemma 1.1.20. Another necessary yet not sufficient condition for $\boldsymbol{U} \in \mathcal{D}(\boldsymbol{\eta})$, which is of the form of Theorem 1.1.39 (i), will be considered in Lemma 2.1.1. This condition will then be sharpened such that $\boldsymbol{U} \in \mathcal{D}(\boldsymbol{\eta})$ follows and will serve as the basis for the statistical tests in the following chapter.

2 Testing for Generalized Pareto Models

So far we have considered two different frameworks that share very similar properties: the finite dimensional and the functional one. As we have seen in Chapter 1, GPDs are in both setups natural tools for modeling exceedances above a high threshold. In particular, the finite dimensional marginal distribution functions of an MSD and a GPD are, respectively, MSDs and GPDs themselves.

Now consider we aim at modeling those exceedances adequately for a given data set. Chapter 1 shows that if the threshold is sufficiently large, the distribution of the exceedances should be close to a GPD. As the number of dimensions of our observed data increases, or if the data are functional, adapting a model based on a GPD gets more and more complicated since the dependencies among the single components need to be modeled, too; see e. g. Aulbach et al. (2012*a*) and Aulbach et al. (2012*b*). A first step in the model selection procedure could be to examine how close the data are to a GPD. To this end, Section 2.1 derives certain neighborhoods of a GPD, which have a polynomial rate of convergence towards a GPD. The subsequent sections will present a goodness-of-fit test with the null hypothesis that the observed data originate from a distribution which belongs to such a neighborhood.

The test itself will be developed in several steps: Section 2.2 first introduces the testing procedure in finite dimensions when copula data are observed, and then extends the results to general data. Similarly, the approach is carried over to stochastic processes in $C[0,1]$ by Section 2.3. Section 2.4 shows that the test consistently links both frameworks, the finite dimensional and the functional one: We will assume that the data are actually generated by some continuous process which cannot be observed as a whole but at a finite set of observation points only. If the number of observation points tends to infinity in a certain manner, we end up with the test statistic of Section 2.3.

The highlight of this chapter will be that the asymptotic distribution of our test statistic is the same for *all* frameworks under consideration — no matter whether we observe finite or infinite dimensional data and whether these emerge from a copula or not. In order to increase its readability, this chapter focuses on the derivation of the test itself; a discussion of the strength of the technical prerequisites is deferred to Chapter 3, apart from minor exceptions.

2.1 Null Hypothesis

Contrary to the multivariate framework in Section 1.1, where each random vector can be decomposed into its univariate marginal distributions and a copula random vector, there are stochastic processes in $C[0,1]$ which do *not* have a copula process in $C[0,1]$; cf. Section 3.3. However, recall that a stochastic process in $C[0,1]$ with continuous marginal distributions does have a corresponding copula process, cf. Lemma 1.2.9, and that the class of all those processes is sufficient for many applications. For example, think of a dike that prevents flooding from the sea. Assume that the sea level is observed at each point of the length of this dike along the coast. As the waves approach the coast, the sea level increases and decreases continuously. Attempting to model the distribution of the sea level at a single observation point discontinuously would mean that certain levels would appear with a strictly positive probability, whereas slightly lower and slightly larger wave heights would have probability zero, which seems unnatural.

In what follows we will focus on stochastic processes that do have a continuous copula process. Under this assumption, which will be reviewed in Section 3.3 below, we have by Theorem 1.2.18 that we can examine the marginal distributions of a stochastic process and the corresponding continuous copula process separately, cf. Theorem 1.1.18 and Theorem 1.2.16. Precisely, if we look for a suited probabilistic model for the upper tail of the distribution of a stochastic process $\boldsymbol{X} = (X_t)_{t\in[0,1]}$, the corresponding copula process $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ in $C[0,1]$ should, in presence of (1.2.19), be modeled such that $\boldsymbol{U}$ is in the domain of attraction of a standard MSP. In particular, Theorem 1.2.28 yields the following necessary condition for $\boldsymbol{U} \in \mathcal{D}(\boldsymbol{\eta})$, which is quite similar to Theorem 1.1.39 (i).

Lemma 2.1.1 *Consider a functional copula C and a standard MSD G with corresponding D-norm $\|\cdot\|_D$. Then $C \in \mathcal{D}(G)$ implies $C(f) = 1 - \|f - \mathbb{1}_{[0,1]}\|_D + r(f)$ for each $f \in \mathcal{E}[0,1] \cap [0,1]^{[0,1]}$ where the remainder satisfies*

$$r(\mathbb{1}_{[0,1]}) = 0 \quad \text{and} \quad \lim_{t\to 0+} \sup_{\substack{f\in\mathcal{E}[0,1]\cap[0,1]^{[0,1]}\\ \|f-\mathbb{1}_{[0,1]}\|_\infty < t}} |r(f)| = 0. \tag{2.1.2}$$

Proof. Consider a copula process $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ in $C[0,1]$ with distribution function C. If $g \in \mathcal{E}[0,1] \cap (-\infty,0]^{[0,1]}$ satisfies $\|g\|_\infty = 1$, then there is for each $n \in \mathbb{N}$ some $t_n \in [0,1]$ such that $g(t_n) \in \left[-1, -1 + \frac{1}{n}\right]$. This gives $g \le -\left(1 - \frac{1}{n}\right)\mathbb{1}_{\{t_n\}}$, and thus

$$\mathrm{P}\Big(\boldsymbol{U} \not\le \mathbb{1}_{[0,1]} + t\,g\Big) \ge \mathrm{P}\left(\boldsymbol{U} \not\le \mathbb{1}_{[0,1]} - t\left(1 - \frac{1}{n}\right)\mathbb{1}_{\{t_n\}}\right) = \mathrm{P}\left(U_{t_n} > 1 - t\left(1 - \frac{1}{n}\right)\right)$$

for arbitrary $t \in (0,1)$ and $n \in \mathbb{N}$. We obtain on the one hand $\mathrm{P}(\boldsymbol{U} \not\leq \mathbb{1}_{[0,1]} + t\, g) \geq t$ and

$$\frac{\mathrm{P}\Big(\boldsymbol{U} \not\leq \mathbb{1}_{[0,1]} + t\, g\Big) - t\, \|g\|_D}{t} \geq 1 - \|g\|_D \geq 1 - \big\|\mathbb{1}_{[0,1]}\big\|_D$$

for all $t \in (0,1)$, cf. (1.2.5). On the other hand, Theorem 1.2.28 gives

$$\begin{aligned} \frac{\mathrm{P}\Big(\boldsymbol{U} \not\leq \mathbb{1}_{[0,1]} + t\, g\Big) - t\, \|g\|_D}{t} &\leq \frac{\mathrm{P}\Big(\boldsymbol{U} \not\leq (1-t)\, \mathbb{1}_{[0,1]}\Big) - t\, \big\|\mathbb{1}_{[0,1]}\big\|_D}{t} + \big\|\mathbb{1}_{[0,1]}\big\|_D - \|g\|_D \\ &\leq o(1) + \big\|\mathbb{1}_{[0,1]}\big\|_D - 1 \quad \text{as} \quad t \to 0+ \end{aligned}$$

since $\|g\|_D \geq \|g\|_\infty = 1$. We conclude for $f \in \mathcal{E}[0,1] \cap ([0,1]^{[0,1]} \setminus \{1\}^{[0,1]})$

$$\begin{aligned} |r(f)| &= \Big|\mathrm{P}\Big(\boldsymbol{U} \not\leq \mathbb{1}_{[0,1]} + (f - \mathbb{1}_{[0,1]})\Big) - \big\|f - \mathbb{1}_{[0,1]}\big\|_D\Big| \\ &\leq \sup_{\substack{g \in \mathcal{E}[0,1] \cap (-\infty,0]^{[0,1]} \\ \|g\|_\infty = 1}} \Big|\mathrm{P}\Big(\boldsymbol{U} \not\leq \mathbb{1}_{[0,1]} + \big\|f - \mathbb{1}_{[0,1]}\big\|_\infty g\Big) - \big\|f - \mathbb{1}_{[0,1]}\big\|_\infty \|g\|_D\Big| \\ &< \big\|f - \mathbb{1}_{[0,1]}\big\|_\infty \big\|\mathbb{1}_{[0,1]}\big\|_D \end{aligned}$$

whenever $\big\|f - \mathbb{1}_{[0,1]}\big\|_\infty$ is sufficiently small, which implies (2.1.2). $\square$

So far, we have seen that $C \in \mathcal{D}(G)$ implies a certain approximation of the upper tail of C in terms of a GPD. Recall that by Theorem 1.1.39 a sharper version of (2.1.2) is, in the multivariate framework, necessary *and* sufficient for a copula to be in the domain of attraction of a standard MSD. In order to emphasize the similarities of both frameworks under consideration, the finite dimensional one and the functional one, it is convenient to introduce some further notation.

Definition 2.1.3 For $\mathcal{I} = [0,1]^d$ or $\mathcal{I} = \mathcal{E}[0,1] \cap [0,1]^{[0,1]}$ we define

$$\mathcal{B}_+(x,r) := \{y \in \mathcal{I} \mid 0 < \|y - x\|_\infty < r\} \quad \text{for} \quad x \in \mathcal{I} \text{ and } r > 0,$$

i. e. we take, with respect to the sup-norm, the open ball in $\mathcal{I}$ with center x and radius r excluding x.

Motivated by Theorem 1.1.39 and Lemma 2.1.1, we now assume that the copula C underlying the observed data has the expansion

$$C(f) = 1 - \big\|f - \mathbb{1}_{[0,1]}\big\| + r(f) \quad \text{for all} \quad f \in \mathcal{E}[0,1] \cap [0,1]^{[0,1]}$$

with some norm $\|\cdot\|$, and that the remainder satisfies

$$r(\mathbb{1}_{[0,1]}) = 0 \quad \text{and} \quad \lim_{t\to 0+} \sup_{f\in\mathcal{B}_+(\mathbb{1}_{[0,1]},t)} \frac{|r(f)|}{\|f-\mathbb{1}_{[0,1]}\|_\infty} = 0, \tag{2.1.4}$$

which corresponds to the condition in Theorem 1.1.39. If $\|\cdot\|$ and $\|\cdot\|_\infty$ are equivalent, we obtain

$$\begin{aligned} C^n\left(\mathbb{1}_{[0,1]} + \frac{1}{n}f\right) &= \left[1 - \frac{1}{n}\|f\| + r\left(\mathbb{1}_{[0,1]} + \frac{1}{n}f\right)\right]^n \\ &= \left[1 - \frac{\|f\|}{n}\left(1 - \frac{r\left(\mathbb{1}_{[0,1]} + \frac{1}{n}f\right)}{\frac{1}{n}\|f\|_\infty}\frac{\|f\|_\infty}{\|f\|}\right)\right]^n \\ &\to \exp(-\|f\|) \quad \text{as} \quad n\to\infty \end{aligned}$$

for $f \in \mathcal{E}[0,1] \cap ((-\infty,0]^{[0,1]} \setminus \{0\}^{[0,1]})$ since

$$\left|\frac{r\left(\mathbb{1}_{[0,1]} + \frac{1}{n}f\right)}{\frac{1}{n}\|f\|_\infty}\right| \le \sup_{g\in\mathcal{B}_+\left(\mathbb{1}_{[0,1]},\frac{1}{n-1}\|f\|_\infty\right)} \frac{|r(g)|}{\|g-\mathbb{1}_{[0,1]}\|_\infty} \to 0 \quad \text{as} \quad n\to\infty.$$

This proves that $\|\cdot\|$ is a D-norm and that $\boldsymbol{U}$ is in the domain of attraction of the standard MSD with D-norm $\|\cdot\|$, cf. Aulbach et al. (2013, Remark 2) and Theorem 1.1.39. Due to (1.2.5), the assumption of $\|\cdot\|$ and $\|\cdot\|_\infty$ being equivalent cannot be dropped.

In what follows we will focus on the rate of convergence of the remainder specified in (2.1.4). Definition 2.1.5 therefore distinguishes the both cases where the remainder vanishes and where the rate of convergence is of polynomial order.

Definition 2.1.5 A (finite dimensional) copula C is called a *(finite dimensional) generalized Pareto copula* (GPC) if there is a (finite dimensional) D-norm $\|\cdot\|_D$ and $\boldsymbol{u}_0 \in [\boldsymbol{0},\boldsymbol{1})$ satisfying

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

Then for $\delta > 0$ we call the set $\mathcal{D}_\delta(C)$ of all distribution functions F satisfying

$$\exists_{c,\varepsilon>0}\,\forall_{\boldsymbol{u}\in\mathcal{B}_+(\boldsymbol{1},\varepsilon)} \left|\frac{1-F(\boldsymbol{u}) - \|\boldsymbol{u}-\boldsymbol{1}\|_D}{\|\boldsymbol{u}-\boldsymbol{1}\|_\infty}\right| \le c\,\|\boldsymbol{u}-\boldsymbol{1}\|_\infty^\delta$$

the *(finite dimensional) δ-neighborhood* of the GPC C.

Similarly, a (functional) copula C is called a *(functional) GPC* if there exist a D-norm $\|\cdot\|_D$ and $f_0 \in \mathcal{E}[0,1] \cap [0,1)^{[0,1]}$ such that

$$C(f) = 1 - \|f-\mathbb{1}_{[0,1]}\|_D \quad \text{for all} \quad f\in\mathcal{E}[0,1]\cap\left[f_0,\mathbb{1}_{[0,1]}\right].$$

Then for $\delta > 0$ the set $\mathcal{D}_\delta(C)$ of all distribution functions F having the property

$$\exists_{c,\varepsilon>0}\,\forall_{f\in\mathcal{B}_+(\mathbb{1}_{[0,1]},\varepsilon)}\ \left|\frac{1-F(f)-\left\|f-\mathbb{1}_{[0,1]}\right\|_D}{\left\|f-\mathbb{1}_{[0,1]}\right\|_\infty}\right| \le c\left\|f-\mathbb{1}_{[0,1]}\right\|_\infty^\delta$$

is called the *(functional) δ-neighborhood* of C.

Obviously, a distribution function C is a GPC if and only if $C(\cdot-\mathbf{1})$ or $C(\cdot-\mathbb{1}_{[0,1]})$, respectively, is a standard GPD *and* the univariate margins of C correspond to the uniform distribution on $[0,1]$. It is easy to verify that any finite dimensional projection of a GPC is a GPC as well. Moreover, if there exists $\delta>0$ such that $F\in\mathcal{D}_\delta(C)$ for some GPC C, then any finite dimensional projection of F is also in the δ-neighborhood of the corresponding projection of C. Any stochastic process $\boldsymbol{X}$ in $C[0,1]$ with distribution function $F\in\mathcal{D}_\delta(C)$ satisfies in particular

$$\mathrm{P}\Big(\boldsymbol{X}<\mathbb{1}_{[0,1]}\Big)=1+\lim_{m\to\infty}\left[F\left(\left(1-\frac{1}{m}\right)\mathbb{1}_{[0,1]}\right)-C\left(\left(1-\frac{1}{m}\right)\mathbb{1}_{[0,1]}\right)\right]=1.$$

We shortly summarize some further properties of δ-neighborhoods:

Lemma 2.1.6 *Let C, C_* be GPCs with corresponding D-norms $\|\cdot\|_D, \|\cdot\|_{D,*}$. Further let F be a distribution function.*

(i) We have $\mathcal{D}_{\delta_1}(C)\subset\mathcal{D}_{\delta_2}(C)$ for $0<\delta_2<\delta_1$.

(ii) Any standard GPD W with D-norm $\|\cdot\|_D$ satisfies $W(\cdot+\mathbf{1})\in\mathcal{D}_\delta(C)$ for all $\delta>0$ in the finite dimensional case and $W(\cdot+\mathbb{1}_{[0,1]})\in\mathcal{D}_\delta(C)$ for all $\delta>0$ in the functional case.

(iii) If there is some $\delta>0$ with $F\in\mathcal{D}_\delta(C)$, then $F\in\mathcal{D}(G)$ where G is the standard MSD with D-norm $\|\cdot\|_D$.

(iv) If $F\in\mathcal{D}_\delta(C)$ and $F\in\mathcal{D}_{\delta_*}(C_*)$ for some $0<\delta\le\delta_*$, then $F\in\mathcal{D}_{\delta_*}(C)$ and $\|\cdot\|_D=\|\cdot\|_{D,*}$.

(v) If $C\in\mathcal{D}_{\delta_0}(C_*)$ for some $\delta_0>0$, then $C_*\in\mathcal{D}_\delta(C)$ for all $\delta>0$ and $\|\cdot\|_D=\|\cdot\|_{D,*}$.

Proof. Parts (i) and (ii) are obvious. We focus on the functional cases of the remaining assertions; the finite dimensional ones are proven similarly.

The assumption $F\in\mathcal{D}_\delta(C)$ implies that there exist $K,\varepsilon>0$ such that the function $r:\mathcal{E}[0,1]\cap(-\infty,0]^{[0,1]}\to\mathbb{R}$ given by $r(f)=F(\mathbb{1}_{[0,1]}+\frac{1}{n}f)-1+\frac{1}{n}\|f\|_D$ satisfies

$$\frac{\|f\|_\infty}{\|f\|_D}\frac{|r(f)|}{\frac{1}{n}\|f\|_\infty}\le K\left(\frac{1}{n}\|f\|_D\right)^\delta\quad\text{whenever}\quad 0<\frac{1}{n}\|f\|_\infty<\varepsilon.$$

Now the arguments prior to Definition 2.1.5 yield

$$F^n\left(\mathbb{1}_{[0,1]} + \frac{1}{n} f\right) = \left[1 - \frac{1}{n} \|f\|_D \left[1 + O\left(\left(\frac{1}{n} \|f\|_D\right)^{\delta}\right)\right]\right]^n \to \exp(-\|f\|_D)$$

as $n \to \infty$ pointwise for all $f \in \mathcal{E}[0,1] \cap (-\infty, 0]^{[0,1]}$, as asserted in (iii).

For part (iv) it suffices to show $\|\cdot\|_D = \|\cdot\|_{D,*}$ since then $\mathcal{D}_{\delta_*}(C) = \mathcal{D}_{\delta_*}(C_*)$ follows by Definition 2.1.5. Note that $F \in \mathcal{D}_\delta(C)$ and $F \in \mathcal{D}_{\delta_*}(C_*)$ imply the existence of $\varepsilon, c, c_* > 0$ such that

$$\begin{aligned}\left|\frac{\|f\|_{D,*} - \|f\|_D}{\|f\|_\infty}\right| &\leq \left|\frac{1 - F(\mathbb{1}_{[0,1]} + f) - \|f\|_D}{\|f\|_\infty}\right| + \left|\frac{1 - F(\mathbb{1}_{[0,1]} + f) - \|f\|_{D,*}}{\|f\|_\infty}\right| \\ &\leq (c + c_*) \|f\|_\infty^\delta \quad \text{for all} \quad f \in \mathcal{B}_+(\mathbb{0}_{[0,1]}, \varepsilon).\end{aligned}$$

For $f \in \mathcal{E}[0,1] \setminus \{\mathbb{0}_{[0,1]}\}$ we obtain

$$\left|\frac{\|f\|_{D,*} - \|f\|_D}{\|f\|_\infty}\right| = \left|\frac{\left\|\frac{\eta}{\|f\|_\infty} f\right\|_{D,*} - \left\|\frac{\eta}{\|f\|_\infty} f\right\|_D}{\left\|\frac{\eta}{\|f\|_\infty} f\right\|_\infty}\right| \leq (c + c_*)\eta^\delta \quad \text{for all} \quad \eta \in (0, \varepsilon),$$

which proves $\|\cdot\|_D = \|\cdot\|_{D,*}$. Part (v) is, due to (ii), a special case of (iv). □

The preceding result shows that δ-neighborhoods are consistently supplemented by shifted standard GPDs. Therefore the convention $u^\infty = 0$ for $u \in [0,1)$ leads to the following identification:

Definition 2.1.7 Let C be a GPC with (finite dimensional or functional) D-norm $\|\cdot\|_D$ and denote by $\mathcal{D}_\infty(C)$ the set of all distribution functions F such that $F(\cdot - \mathbf{1})$ or $F(\cdot - \mathbb{1}_{[0,1]})$, respectively, is a standard GPD with D-norm $\|\cdot\|_D$. Then we call $\mathcal{D}_\infty(C)$ the ∞-*neighborhood* of C.

As outlined above, δ-neighborhoods provide a consistent approach of modeling data by specifying how close the underlying distribution function is to a GPC. These kind of models will serve as the null hypothesis for the tests in the subsequent sections.

Hypothesis 2.1.8 There exist $\delta \in (0, \infty]$ and a GPC such that the copula underlying the observed data is in the δ-neighborhood of this GPC.

Remark 2.1.9 A similar condition has also been considered in Einmahl et al. (2006), where a test for the bivariate extreme value condition is performed. Precisely, Hypothesis 2.1.8 corresponds to Einmahl et al. (2006, Equation 2.5) and assures that the underlying bivariate D-norm $l(x_1, x_2) := \|(x_1, x_2)^\intercal\|_D$ is estimated consistently. Since the cited authors use probabilities of the form $\mathrm{P}(1 - F_1(X_1) \le tx_1, 1 - F_2(X_2) \le tx_2)$ instead of copula probabilities, their function R has, in our notation, the representation $R(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_D$ for $\boldsymbol{x} \in [0, \infty)^2$. However, Einmahl et al. (2006) rely on a certain representation of the bivariate spectral measure, cf. Remark 1.1.30, which does not seem to extend to higher dimensions in an obvious manner. We also refer to Aulbach and Falk (2012), who tested for standard GPPs instead of δ-neighborhoods.

While (2.1.4) assumes a locally uniform approximation of a distribution function by means of a GPC, δ-neighborhoods require a certain quality of that approximation. As seen before, any distribution function which is in a δ-neighborhood of a GPC C with D-norm $\|\cdot\|_D$ is also in the domain of attraction of the standard MSD G with D-norm $\|\cdot\|_D$, i. e. $\bigcup_{\delta\in(0,\infty]} \mathcal{D}_\delta(C) \subset \mathcal{D}(G)$. Moreover, Falk and Reiss (2002, Theorem 1.1), Falk et al. (2011, Theorem 5.5.5), and Aulbach et al. (2015*a*, Proposition 3.7) have shown that a δ-neighborhood, roughly, collects all those distributions with a certain polynomial rate of convergence towards that MSD. For convenience we state the functional version of this result; its proof is analogous to Falk and Reiss (2002, Theorem 1.1).

Lemma 2.1.10 (Aulbach et al., 2015*a*) *Let F be the distribution function of a stochastic process in $C[0,1] \cap (-\infty, 1]^{[0,1]}$, and let C be a GPC with corresponding D-norm $\|\cdot\|_D$.*

(i) If $\delta \in (0, 1]$ and F is in the δ-neighborhood of C, then we have

$$\sup_{f\in\bar{E}[0,1]\cap(-\infty,0]^{[0,1]}} \left| F^n\left(\mathbb{1}_{[0,1]} + \frac{1}{n} f\right) - \exp(-\|f\|_D)\right| = O\left(\frac{1}{n^\delta}\right) \tag{2.1.11}$$

as $n \to \infty$.

(ii) Suppose that there is $\varepsilon > 0$ such that the derivative $h_f(c) := \frac{\partial}{\partial c} F(\mathbb{1}_{[0,1]} + c\,|f|)$ exists for all $c \in (-\varepsilon, 0)$ and $f \in \mathcal{B}_- := \{g \in \bar{E}[0,1] \cap (-\infty, 0]^{[0,1]} \mid \|g\|_\infty = 1\}$. Moreover let $H_f(c) := F(\mathbb{1}_{[0,1]} + c\,|f|)$, $c < 0$, satisfy the von Mises condition

$$r_f(c) := \frac{-c\, h_f(c)}{1 - H_f(c)} - 1 \to_{c\to 0-} 0 \quad \text{for all} \quad f \in \mathcal{B}_-$$

and additionally

$$\sup_{f\in\mathcal{B}_-} \left|\int_c^0 \frac{r_f(t)}{t}\, \mathrm{d}t\right| \to_{c\to 0-} 0.$$

If (2.1.11) holds for some $\delta \in (0, 1]$, then F is in the δ-neighborhood of C.

Note that the family $\{H_f \mid f \in \mathcal{B}_-\}$ of univariate distribution functions on $(-\infty, 0]$ completely determines the distribution function $H(f) := F\big(\mathbb{1}_{[0,1]} + f\big)$, $f \in \mathcal{E}[0,1] \cap (-\infty, 0]^{[0,1]}$. It is called the *spectral decomposition* of H, cf. Falk et al. (2011, Section 5.4).

2.2 Testing for Finite Dimensional δ-Neighborhoods

Now consider that we observe multivariate data and our aim is to check whether the observed dependencies can be modeled by a copula that satisfies the null hypothesis derived in Section 2.1, which is restated here for convenience. For examples of copulas that satisfy this hypothesis and for examples of those that do not, we refer to Chapter 3.

Hypothesis 2.1.8 There exist $\delta \in (0, \infty]$ and a GPC such that the copula underlying the observed data is in the δ-neighborhood of this GPC.

Assume that our data consist of independent realizations of a random vector $\boldsymbol{X} = (X_1, \dots, X_d)^\intercal$ with arbitrary distribution function F. Due to Theorem 1.1.13, there is a copula C such that $F(\boldsymbol{x}) = C\big(F_1(x_1), \dots, F_d(x_d)\big)$ for $\boldsymbol{x} = (x_1, \dots, x_d)^\intercal \in \mathbb{R}^d$, where F_i denotes the i-th margin of F. If $\boldsymbol{U} = (U_1, \dots, U_d)^\intercal$ has distribution function C, then we obtain obviously

$$\mathrm{P}\left(\left(F_i^{-1}(U_i)\right)_{i=1}^{d} \le \boldsymbol{x}\right) = \mathrm{P}\left(\bigcap_{i=1}^{d}\{U_i \le F_i(x_i)\}\right) = F(\boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^d,$$

i. e. we may assume $\boldsymbol{X} = \big(F_i^{-1}(U_i)\big)_{i=1}^{d}$ without loss of generality.

At first we will consider the case that F is a copula itself, i. e. our data actually consist of independent realizations of the random vector $\boldsymbol{U}$. After having derived a test statistic for Hypothesis 2.1.8 in this framework, we will generalize our results to distribution functions with continuous margins, utilizing the empirical counterpart of a copula.

Copula Data

Assume we observe independent copies $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$ of a random vector $\boldsymbol{U} = (U_1, \dots, U_d)^\intercal$ which is distributed according to a copula C. If Hypothesis 2.1.8 is true, there exist a D-norm $\|\cdot\|_D$ and $\delta \in (0, \infty]$, $K > 0$, $\varepsilon \in (0, 1]$ such that

$$\big|\mathrm{P}(\boldsymbol{U} \not\le \boldsymbol{u}) - \|\boldsymbol{u} - \boldsymbol{1}\|_D\big| \le K\,\|\boldsymbol{u} - \boldsymbol{1}\|_\infty^{1+\delta} \quad \text{for all} \quad \boldsymbol{u} \in \mathcal{B}_+(\boldsymbol{1}, \varepsilon);$$

recall the convention $u^\infty = 0$ for $u \in [0, 1)$ and Definition 2.1.7. We obtain thus

$$\sup_{\boldsymbol{u}\in[(1-c)\mathbf{1},\mathbf{1}]} \left|\mathrm{P}(\boldsymbol{U} \not\leq \boldsymbol{u}) - \|\boldsymbol{u}-\mathbf{1}\|_D\right| \leq Kc^{1+\delta} \quad \text{for all} \quad c \in (0,\varepsilon), \tag{2.2.1}$$

i. e. the probability that $\boldsymbol{U}$ exceeds a threshold $\boldsymbol{u}$ can be approximated by $\|\boldsymbol{u}-\mathbf{1}\|_D$, uniformly for all $\boldsymbol{u} \geq (1-c)\mathbf{1}$.

A natural estimator of $\mathrm{P}(\boldsymbol{U} \not\leq (1-c)\mathbf{1})$ is the relative frequency of those random vectors among $\boldsymbol{U}^{(1)},\dots,\boldsymbol{U}^{(n)}$ which are not less than or equal to the threshold vector $(1-c)\mathbf{1}$. Therefore consider

$$S_{\boldsymbol{U}}(c) := \frac{1}{d}\sum_{i=1}^{d} \mathbb{1}_{(1-c,1]}(U_i), \quad c \in (0,1), \tag{2.2.2}$$

which is the mean number of those components of $\boldsymbol{U}$ that exceed the value $1-c$. Actually, $S_{\boldsymbol{U}}(c)$ is a discrete version of the sojourn time that the random vector $\boldsymbol{U}$ spends above the threshold $1-c$; see Falk and Hofmann (2011) for details as well as Section 2.3.

Definition 2.2.3 Let C be a d-variate copula and $\boldsymbol{U} = (U_1,\dots,U_d)^\intercal$ a random vector in $\mathbb{R}^d$ with distribution function C. For $c \in (0,1)$ we call $S_{\boldsymbol{U}}(c)$ in (2.2.2) the *sojourn time* of $\boldsymbol{U}$ above $(1-c)\mathbf{1}$, and

$$N_{\boldsymbol{U}}^{(n)}(c) := \sum_{i=1}^{n} \mathbb{1}_{(0,1]}(S_{\boldsymbol{U}^{(i)}}(c)), \quad c \in (0,1),$$

the *number of exceedances* above $(1-c)\mathbf{1}$ among independent copies $\boldsymbol{U}^{(1)},\dots,\boldsymbol{U}^{(n)}$ of $\boldsymbol{U}$; i. e. we count how many of the duplicates of $\boldsymbol{U}$ have at least one component that is greater than $1-c$.

Since $N_{\boldsymbol{U}}^{(n)}(c)$ is $\mathcal{B}(n,p(c))$-distributed with $p(c) = \mathrm{P}(\boldsymbol{U} \not\leq (1-c)\mathbf{1})$, an obvious estimator of $\mathrm{P}(\boldsymbol{U} \not\leq (1-c)\mathbf{1})$ is

$$\frac{1}{n} N_{\boldsymbol{U}}^{(n)}(c) \to \mathrm{P}(\boldsymbol{U} \not\leq (1-c)\mathbf{1}) \quad \text{with probability one as } n \to \infty.$$

In presence of Hypothesis 2.1.8, (2.2.1) shows that $\frac{1}{n} N_{\boldsymbol{U}}^{(n)}(c)$ actually estimates $c\,\|\mathbf{1}\|_D$ whenever $c < \varepsilon$. In order to test Hypothesis 2.1.8, we require $c = c_n \in (0,1)$ to depend on the sample size n and to satisfy $c_n \to 0$ as $n \to \infty$.

Lemma 2.2.4 *For $\delta \in (0,\infty]$ let C be a (finite dimensional) copula that is in the δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$. If $\boldsymbol{U}$ has the distribution function C, we obtain*

$$(nc_n)^{\frac{1}{2}}\left(\frac{N_{\boldsymbol{U}}^{(n)}(c_n)}{nc_n} - \|\mathbf{1}\|_D\right) \xrightarrow{D} \boldsymbol{N}(0,\|\mathbf{1}\|_D) \quad \textit{as} \quad n \to \infty \tag{2.2.5}$$

for any sequence $(c_n)_{n\in\mathbb{N}}$ in $(0,1)$ with $c_n \to 0$, $nc_n \to \infty$, and $nc_n^{1+2\delta} \to 0$ as $n \to \infty$.

Proof. Lemma 1.1.20 and $nc_n \to \infty$ give

$$n\big[1 - C((1-c_n)\mathbf{1})\big] = nc_n \frac{1 - C((1-c_n)\mathbf{1})}{c_n} \to \infty \quad \text{as} \quad n \to \infty.$$

Thus Lindeberg's central limit theorem — see e. g. Billingsley (2012, Theorem 27.2) — implies

$$\sum_{i=1}^{n} \frac{\mathbb{1}_{(0,1]}\big(S_{\boldsymbol{U}^{(i)}}(c_n)\big) - 1 + C((1-c_n)\mathbf{1})}{\sqrt{n\, C((1-c_n)\mathbf{1})\big[1 - C((1-c_n)\mathbf{1})\big]}} \xrightarrow{D} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty$$

where $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$ are independent copies of $\boldsymbol{U}$. Moreover there is some $K > 0$ satisfying

$$(nc_n)^{\frac{1}{2}} \left| \frac{1 - C((1-c_n)\mathbf{1})}{c_n} - \|\mathbf{1}\|_D \right| \le K \left(nc_n^{1+2\delta} \right)^{\frac{1}{2}}$$

whenever n is sufficiently large, cf. (2.2.1). Since $nc_n^{1+2\delta} \to 0$ as $n \to \infty$, the assertion follows from $C((1-c_n)\mathbf{1}) \to 1$, $\frac{1}{c_n}\big[1 - C((1-c_n)\mathbf{1})\big] \to \|\mathbf{1}\|_D$, and Slutsky's theorem; see e. g. Gut (2013, Theorem 11.4). □

Remark 2.2.6 If $\delta = \infty$, i. e. C is a GPC itself, the condition $0 = nc_n^{\infty} \to 0$ is trivial and we have in particular $\frac{1}{c_n}\big[1 - C((1-c_n)\mathbf{1})\big] = \|\mathbf{1}\|_D$ whenever n is sufficiently large.

Note that the test statistic in (2.2.5) still depends on the usually unknown generator constant $\|\mathbf{1}\|_D$. That is we need to estimate $\|\mathbf{1}\|_D$ from the data in such a way that we can still exploit the asymptotic normal distribution in Lemma 2.2.4. Recall the uniform approximation (2.2.1), which holds if Hypothesis 2.1.8 is true. Thus we can consider the $k \in \mathbb{N}$ different thresholds $(1 - \frac{c}{j})\mathbf{1}$, $j = 1, \dots, k$, simultaneously. This leads to a generalization of Lemma 2.2.4, namely Corollary 2.2.9, which will be the crucial tool for the proofs of the main results of this section, Theorem 2.2.10 and Theorem 2.2.12.

In order to reuse Corollary 2.2.9 in subsequent sections, we introduce an additional sequence $(m_n)_{n\in\mathbb{N}}$ with $m_n \le n$ and $m_n \to \infty$ as $n \to \infty$. The case $m_n < n$ will be of particular interest when we consider more general data. The proof of Corollary 2.2.9 relies on the following rather general tool, which will turn out to be useful in the functional framework as well.

Lemma 2.2.7 *Consider $k \in \mathbb{N}$ as well as two sequences $(m_n)_{n\in\mathbb{N}}$ and $(p_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ and $(0,1)$, respectively, which satisfy $p_n \to 0$ and $m_n p_n \to \infty$ as $n \to \infty$. For each $n \in \mathbb{N}$ let $X_1^{(n)}, \dots, X_{m_n}^{(n)}$ be independent random elements with events $A_1^{(n)} \supset \dots \supset A_k^{(n)}$ such that for all $i \in \{1, \dots, m_n\}$ and $j \in \{1, \dots, k\}$*

$$\mathrm{P}\Big(X_i^{(n)} \in A_j^{(n)}\Big) = \frac{p_n}{j}(1 + r_j(n))$$

with remainders that satisfy $r_j(n) = o\Big((m_n p_n)^{-\frac{1}{2}}\Big)$ *as* $n \to \infty$*. Then we have*

$$\left(\sum_{i=1}^{m_n} \frac{j\, \mathbb{1}_{A_j^{(n)}}(X_i^{(n)}) - p_n}{(m_n p_n)^{\frac{1}{2}}} \right)_{j=1}^{k} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{M}\boldsymbol{M}^\intercal) \quad \textit{as} \quad n \to \infty$$

where

$$\boldsymbol{M} = \Big(\mathbb{1}_{[\ell,\infty)}(j)\Big)_{1 \le j,\ell \le k} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}. \tag{2.2.8}$$

Proof. Let $\boldsymbol{t} = (t_1, \dots, t_k)^\intercal \in \mathbb{R}^k \setminus \{\mathbf{0}\}$ and define the symmetric and positive semidefinite $k \times k$ matrix $\boldsymbol{\Sigma}_n = (\sigma_{j\ell}^{(n)})_{1 \le j,\ell \le k}$ by

$$\begin{aligned} \sigma_{j\ell}^{(n)} &:= \mathrm{Cov}\Big(j\, \mathbb{1}_{A_j^{(n)}}(X_i^{(n)}), \ell\, \mathbb{1}_{A_\ell^{(n)}}(X_i^{(n)})\Big) \\ &= j\ell \left(\frac{p_n(1 + r_{\max\{j,\ell\}}(n))}{\max\{j,\ell\}} - \frac{p_n^2(1 + r_j(n))(1 + r_\ell(n))}{j\ell} \right) \\ &= \min\{j,\ell\}\, p_n \left(1 + r_{\max\{j,\ell\}}(n) - \frac{p_n(1 + r_j(n))(1 + r_\ell(n))}{\min\{j,\ell\}} \right). \end{aligned}$$

Since $\boldsymbol{M}\boldsymbol{M}^\intercal = (\min\{j,\ell\})_{1 \le j,\ell \le k}$ is positive definite and $\frac{1}{p_n}\sigma_{j\ell}^{(n)} \to \min\{j,\ell\}$ as $n \to \infty$, $\boldsymbol{t}^\intercal \boldsymbol{\Sigma}_n \boldsymbol{t}$ is strictly positive for large n. We obtain furthermore

$$\boldsymbol{t}^\intercal \boldsymbol{\Sigma}_n \boldsymbol{t} = \mathrm{Var}\left(\sum_{j=1}^{k} t_j j\, \mathbb{1}_{A_j^{(n)}}(X_i^{(n)}) \right) \quad \text{for} \quad i = 1, \dots, m_n$$

and

$$\sup_{1 \le i \le m_n} \left| \sum_{j=1}^{k} \frac{t_j\Big(j\, \mathbb{1}_{A_j^{(n)}}(X_i^{(n)}) - p_n(1 + r_j(n))\Big)}{(m_n \boldsymbol{t}^\intercal \boldsymbol{\Sigma}_n \boldsymbol{t})^{\frac{1}{2}}} \right| \le \sum_{j=1}^{k} \frac{|t_j|}{j\,(m_n p_n)^{\frac{1}{2}}} \left(\frac{p_n}{\boldsymbol{t}^\intercal \boldsymbol{\Sigma}_n \boldsymbol{t}} \right)^{\frac{1}{2}} \to 0$$

with probability one as $n \to \infty$. Lindeberg's central limit theorem and Slutsky's theorem show therefore

$$\begin{aligned} \boldsymbol{t}^\intercal \left(\sum_{i=1}^{m_n} \frac{j\, \mathbb{1}_{A_j^{(n)}}(X_i^{(n)}) - p_n}{(m_n p_n)^{\frac{1}{2}}} \right)_{j=1}^{k} &= \left(\frac{\boldsymbol{t}^\intercal \boldsymbol{\Sigma}_n \boldsymbol{t}}{p_n} \right)^{\frac{1}{2}} \sum_{i=1}^{m_n} \sum_{j=1}^{k} \frac{t_j\Big(j\, \mathbb{1}_{A_j^{(n)}}(X_i^{(n)}) - p_n(1 + r_j(n))\Big)}{(m_n \boldsymbol{t}^\intercal \boldsymbol{\Sigma}_n \boldsymbol{t})^{\frac{1}{2}}} \\ &\quad + \sum_{j=1}^{k} t_j (m_n p_n)^{\frac{1}{2}}\, r_j(n) \end{aligned}$$

$$\xrightarrow{D} \mathcal{N}(0, \boldsymbol{t}^\intercal \boldsymbol{M}\boldsymbol{M}^\intercal \boldsymbol{t}) \quad \text{as} \quad n \to \infty.$$

Now the assertion follows from the Cramér-Wold Theorem, see e. g. Billingsley (2012, Theorem 29.4). □

The desired generalization of Lemma 2.2.4 is now a simple application of Lemma 2.2.7:

Corollary 2.2.9 *Let $\delta \in (0, \infty]$ and $k \in \mathbb{N}$. Furthermore let C be a (finite dimensional) copula which is in the δ-neighborhood of some GPC with D-norm $\|\cdot\|_D$. Consider a random vector $\boldsymbol{U}$ with distribution function C. If $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{N}$ and $(0,1)$, respectively, such that $c_n \to 0$, $m_n c_n \to \infty$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$, then we have*

$$\frac{1}{k}\sum_{j=1}^{k} \frac{j}{m_n c_n} N_{\boldsymbol{U}}^{(m_n)}\Big(\frac{c_n}{j}\Big) \to \|\boldsymbol{1}\|_D \quad \text{in probability as } n \to \infty$$

and

$$\left((m_n c_n)^{\frac{1}{2}}\left(\frac{j}{m_n c_n} N_{\boldsymbol{U}}^{(m_n)}\Big(\frac{c_n}{j}\Big) - \|\boldsymbol{1}\|_D\right)\right)_{j=1}^{k} \xrightarrow{D} \mathcal{N}(\boldsymbol{0}, \|\boldsymbol{1}\|_D \, \boldsymbol{M}\boldsymbol{M}^\intercal) \quad \text{as} \quad n \to \infty$$

where $\boldsymbol{M}$ is given in (2.2.8).

Proof. We have on the one hand

$$(m_n c_n)^{\frac{1}{2}}\left(\frac{j}{m_n c_n} N_{\boldsymbol{U}}^{(m_n)}\Big(\frac{c_n}{j}\Big) - \|\boldsymbol{1}\|_D\right) = \sum_{i=1}^{m_n} \frac{j\, \mathbb{1}_{(0,1]}\Big(S_{\boldsymbol{U}^{(i)}}\big(\frac{c_n}{j}\big)\Big) - c_n \|\boldsymbol{1}\|_D}{(m_n c_n)^{\frac{1}{2}}}$$

for all $j \in \{1, \dots, k\}$, where $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(m_n)}$ denote independent copies of $\boldsymbol{U}$. Since on the other hand $\mathbb{1}_{(0,1]}(S_{\boldsymbol{U}}(c)) = \mathbb{1}_{[\boldsymbol{0},\boldsymbol{1}]\setminus[\boldsymbol{0},(1-c)\boldsymbol{1}]}(\boldsymbol{U})$, $c \in (0,1)$, and

$$\mathrm{P}\Big(\boldsymbol{U} \nleq \Big(1 - \frac{c_n}{j}\Big)\boldsymbol{1}\Big) = \frac{c_n}{j}\, \|\boldsymbol{1}\|_D \left[1 + O\bigg(\Big(\frac{c_n}{j}\Big)^{\delta}\bigg)\right] \quad \text{as} \quad n \to \infty$$

for all $j \in \{1, \dots, k\}$, Lemma 2.2.7 implies the second assertion. This yields in particular

$$(m_n c_n)^{\frac{1}{2}}\left(\frac{j}{m_n c_n} N_{\boldsymbol{U}}^{(m_n)}\Big(\frac{c_n}{j}\Big) - \|\boldsymbol{1}\|_D\right) \xrightarrow{D} \mathcal{N}(0, j\, \|\boldsymbol{1}\|_D) \quad \text{as} \quad n \to \infty$$

for each $j \in \{1, \dots, k\}$, and thus $\frac{j}{m_n c_n} N_{\boldsymbol{U}}^{(m_n)}(\frac{c_n}{j}) \to \|\boldsymbol{1}\|_D$ in probability, which completes the proof. □

The matrix $\boldsymbol{M}$ in (2.2.8) refers to a well-known stochastic process: Let $\boldsymbol{B} = (B_t)_{t\in[0,\infty)}$ be a *standard Brownian motion*, i. e. $\mathrm{P}(B_0 = 0) = 1$, all sample paths of $\boldsymbol{B}$ are continuous, and the increments $B_{t_i} - B_{t_{i-1}}$, $i = 1, \dots, n$, are independent and $\mathcal{N}(0, t_i - t_{i-1})$-distributed whenever $n \in \mathbb{N}$ and $0 \le t_0 \le t_1 \le \dots \le t_n$; see e. g. Schilling and Partzsch (2014). Then the random vector $(B_j - B_{j-1})_{j=1}^k$ is k-dimensional standard normally distributed and we have $(B_j)_{j=1}^k = \boldsymbol{M}(B_j - B_{j-1})_{j=1}^k$. Corollary 2.2.9 therefore has the following interpretation: Assume that we observe independent data from a copula satisfying Hypothesis 2.1.8, and consider the vector of exceedance frequencies above certain high threshold vectors. If this random vector is normalized properly, the result has asymptotically the same distribution as $(B_j)_{j=1}^k$, a standard Brownian motion which is evaluated for integer arguments. Motivated by the usual chi-square goodness-of-fit test, we obtain a first test for Hypothesis 2.1.8 by diagonalizing the covariance matrix of $(B_j)_{j=1}^k$.

Theorem 2.2.10 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \ge 2$. Furthermore let C be a (finite dimensional) copula which is in the δ-neighborhood of some GPC with D-norm $\|\cdot\|_D$. If the random vector $\boldsymbol{U}$ has distribution function C and $(c_n)_{n\in\mathbb{N}}$ is a sequence in $(0,1)$ satisfying $c_n \to 0$, $nc_n \to \infty$, and $nc_n^{1+2\delta} \to 0$ as $n \to \infty$, we obtain*

$$T_n := \frac{\sum_{j=1}^k \Big(j\, N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{j}\big) - \frac{1}{k}\sum_{\ell=1}^k \ell\, N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{\ell}\big)\Big)^2}{\frac{1}{k}\sum_{\ell=1}^k \ell\, N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{\ell}\big)} \xrightarrow{D} \sum_{j=1}^{k-1} \lambda_j \xi_j^2 \quad as \quad n \to \infty$$

where

$$\lambda_j = \frac{1}{4\sin^2\Big(\frac{j}{k}\frac{\pi}{2}\Big)}, \qquad j = 1, \dots, k-1,$$

and $\xi_1, \dots, \xi_{k-1}$ are independent and standard normally distributed random variables.

Proof. Let $\boldsymbol{I}_k$ be the $k \times k$ unit matrix and let $\boldsymbol{E}_k = (1)_{1\le i,j\le k}$ be the $k \times k$-matrix where all entries equal 1. Then $\boldsymbol{P} := \boldsymbol{I}_k - \frac{1}{k}\boldsymbol{E}_k$ is a projection matrix, i. e. $\boldsymbol{P} = \boldsymbol{P}^\intercal = \boldsymbol{P}^2$, satisfying $\boldsymbol{P}(x)_{i=1}^k = \boldsymbol{0}$ for all $x \in \mathbb{R}$. Thus Corollary 2.2.9 and Slutsky's theorem show that T_n is asymptotically equivalent to

$$\begin{aligned}
&\frac{1}{nc_n\,\|\boldsymbol{1}\|_D} \sum_{j=1}^k \left(j\, N_{\boldsymbol{U}}^{(n)}\Big(\frac{c_n}{j}\Big) - \frac{1}{k}\sum_{\ell=1}^k \ell\, N_{\boldsymbol{U}}^{(n)}\Big(\frac{c_n}{\ell}\Big)\right)^2 \\
&= \frac{1}{nc_n\,\|\boldsymbol{1}\|_D} \begin{pmatrix} 1\cdot N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{1}\big) - nc_n\,\|\boldsymbol{1}\|_D \\ \vdots \\ k\cdot N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{k}\big) - nc_n\,\|\boldsymbol{1}\|_D \end{pmatrix}^\intercal \left(\boldsymbol{I}_k - \frac{1}{k}\boldsymbol{E}_k\right) \begin{pmatrix} 1\cdot N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{1}\big) - nc_n\,\|\boldsymbol{1}\|_D \\ \vdots \\ k\cdot N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{k}\big) - nc_n\,\|\boldsymbol{1}\|_D \end{pmatrix}
\end{aligned}$$

$$= \boldsymbol{Y}_n^\mathsf{T} \boldsymbol{P} \boldsymbol{Y}_n$$

where $\boldsymbol{Y}_n = (Y_{n,1}, \dots, Y_{n,k})^\mathsf{T}$ with

$$Y_{n,j} = \frac{1}{(nc_n \, \|\mathbf{1}\|_D)^{\frac{1}{2}}} \Big(j \, N_{\boldsymbol{U}}^{(n)}\Big(\frac{c_n}{j}\Big) - nc_n \, \|\mathbf{1}\|_D \Big), \qquad j = 1, \dots, k.$$

Corollary 2.2.9 and the multivariate mapping theorem show

$$T_n \xrightarrow{D} \boldsymbol{\xi}^\mathsf{T} \boldsymbol{M}^\mathsf{T} \boldsymbol{P} \boldsymbol{M} \boldsymbol{\xi} \quad \text{as} \quad n \to \infty$$

with a k-dimensional standard normal random vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)^\mathsf{T}$. It is well-known, see e. g. Anderson and Stephens (1997, Section 4) or Fortiana and Cuadras (1997), that the eigenvalues of

$$\boldsymbol{M}^\mathsf{T} \boldsymbol{P} \boldsymbol{M} = \Big(\min\{i-1, j-1\} - \frac{(i-1)(j-1)}{k} \Big)_{1 \le i,j \le k}$$

are

$$\lambda_j = \frac{1}{4 \sin^2\big(\frac{j}{k} \frac{\pi}{2}\big)} \quad \text{for} \quad j = 1, \dots, k-1, \quad \text{and} \quad \lambda_k = 0$$

with corresponding orthonormal eigenvectors

$$\boldsymbol{r}_j = \sqrt{\frac{2}{k}} \Big(\sin\Big(\frac{(i-1)j\pi}{k}\Big)\Big)_{i=1}^k \quad \text{for} \quad j = 1, \dots, k-1, \quad \text{and} \quad \boldsymbol{r}_k = \mathbf{e}_1.$$

This implies $T_n \xrightarrow{D} \boldsymbol{\xi}^\mathsf{T} \operatorname{diag}(\lambda_1, \dots, \lambda_{k-1}, 0) \, \boldsymbol{\xi}$ as $n \to \infty$, which completes the proof. □

In the simple cases $k = 2$ and $k = 3$ we have, respectively, $\lambda_1 = \frac{1}{2}$ and $\lambda_1 = 1$, $\lambda_2 = \frac{1}{3}$. For higher values of k, the distribution function of $\sum_{j=1}^{k-1} \lambda_j \xi_j^2 = \lambda_{k-1} \sum_{j=1}^{k-1} \frac{\lambda_j}{\lambda_{k-1}} \xi_j^2$ may be computed from Robbins and Pitman (1949, Theorem 1). For simulation techniques we refer to Duchesne and Lafaye de Micheaux (2010). As discussed in Remark 2.2.6, the condition $nc_n^{1+\delta} \to 0$ is obsolete for $\delta = \infty$.

Remark 2.2.11 Additionally to the discussion following Corollary 2.2.9, the proof of Theorem 2.2.10 shows

$$\sum_{j=1}^{k-1} \lambda_j \xi_j^2 \overset{D}{=} \sum_{j=1}^{k} \Big(B_j - \frac{1}{k} \sum_{\ell=1}^{k} B_\ell \Big)^2 \quad \text{for} \quad k \ge 2$$

where $(B_t)_{t \in [0,\infty)}$ is a standard Brownian motion. Computing expected values, we obtain the nice by-product

$$\sum_{j=1}^{k-1} \frac{1}{4\sin^2\left(\frac{j}{k}\frac{\pi}{2}\right)} = \frac{(k-1)(k+1)}{6} \quad \text{for} \quad k \geq 2$$

and, using characteristic functions, it is straightforward to prove that

$$\frac{1}{(k-1)(k+1)} \sum_{j=1}^{k-1} \lambda_j \xi_j^2 \xrightarrow{D} \frac{1}{\pi^2} \sum_{j=1}^{\infty} \frac{1}{j^2} \xi_j^2 \quad \text{as} \quad k \to \infty.$$

Taking expectations on both sides motivates the well-known equality $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$.

The previous arguments suggest to replace k in Theorem 2.2.10 with a sequence $(k_n)_{n\in\mathbb{N}}$. If $k_n \to \infty$ as $n \to \infty$ at a proper rate of convergence, it should be possible to reproduce the limit distribution in Remark 2.2.11. Although this might be of theoretical interest, we avoid doing so for several reasons: On the one hand, a data set will typically contain not too many exceedances above a high threshold. If c_n is sufficiently small to detect a δ-neighborhood, i. e. the threshold is sufficiently large, there will be even less data that exceed $\left(1 - \frac{c_n}{k}\right)\mathbf{1}$. This means that we would need a very large sample size in order to increase k *and* to assure that there are still sufficiently many exceedances in the outer most extremal region. On the other hand, it will be necessary to introduce another parameter as soon as we consider more general data, cf. Lemma 2.2.7. While Theorem 2.2.10 allows to choose, e. g., $k = 2$ or $k = 3$ independently of the sample size n, obtaining reasonable values for the parameters would probably become even harder if k depended on n as well, cf. Section 3.4.

However, considering a modification of the inverse matrix of $\boldsymbol{M}$ in (2.2.8) leads to alternative test statistics for Hypothesis 2.1.8. Their asymptotic distributions are compared to the one in Theorem 2.2.10 easier to handle, where one of them will not even depend on k. We refer to Chapter 4 for a comparison of the in total three tests.

Theorem 2.2.12 *Let $\delta \in (0, \infty]$ and $k \in \mathbb{N}$, $k \geq 2$. Furthermore let the d-dimensional random vector $\boldsymbol{U}$ have the distribution function C such that C is a copula which is in the δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$. We obtain*

$$\mathcal{T}_n := \frac{\sum_{j=1}^{k-1}\left((j+1)\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j+1}\right) - j\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right)\right)^2}{\frac{1}{k}\sum_{j=1}^{k} j\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right)} \xrightarrow{D} \chi_{k-1}^2 \quad \textit{as} \quad n \to \infty$$

and

$$\tau_n := \frac{k\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{k}\right) - N_{\boldsymbol{U}}^{(n)}(c_n)}{\left(\frac{k-1}{k}\sum_{j=1}^{k} j\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right)\right)^{\frac{1}{2}}} \xrightarrow{D} \mathcal{N}(0,1) \quad as \quad n\to\infty$$

for any sequence $(c_n)_{n\in\mathbb{N}}$ *in* $(0,1)$ *with* $c_n \to 0$, $nc_n\to\infty$, *and* $nc_n^{1+2\delta}\to 0$ *as* $n\to\infty$.

Proof. Corollary 2.2.9 shows that $\boldsymbol{Y}_n = (Y_{n,1},\dots,Y_{n,k})^\intercal \xrightarrow{D} \mathcal{N}(\boldsymbol{0}, \boldsymbol{M}\boldsymbol{M}^\intercal)$ as $n\to\infty$ where

$$Y_{n,j} = \frac{1}{(nc_n\,\|\boldsymbol{1}\|_D)^{\frac{1}{2}}}\left(j\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right) - nc_n\,\|\boldsymbol{1}\|_D\right) \quad \text{for} \quad j=1,\dots,k$$

and $\boldsymbol{M}$ is defined in (2.2.8). Put

$$\boldsymbol{K} := \left(\left(\mathbb{1}_{\{\ell\}}(j) - \mathbb{1}_{\{\ell+1\}}(j)\right)\mathbb{1}_{[2,\infty)}(j)\right)_{1\le j,\ell\le k} = \begin{pmatrix} 0 & 0 & 0 & \dots & \dots & 0 \\ -1 & 1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -1 & 1 & 0 \\ 0 & \dots & \dots & 0 & -1 & 1 \end{pmatrix},$$

which is derived from the inverse matrix $\boldsymbol{M}^{-1} = (a_{jk})_{1\le j,\ell\le k}$ of $\boldsymbol{M}$ by replacing the entry a_{11} with 0. Due to Corollary 2.2.9 and Slutsky's theorem, $\mathcal{T}_n$ is asymptotically equivalent to

$$\frac{\sum_{j=1}^{k-1}\left((j+1)\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j+1}\right) - j\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right)\right)^2}{nc_n\,\|\boldsymbol{1}\|_D} = \sum_{j=1}^{k-1}(Y_{n,j+1}-Y_{n,j})^2 = (\boldsymbol{K}\boldsymbol{Y}_n)^\intercal \boldsymbol{K}\boldsymbol{Y}_n.$$

The multivariate mapping theorem implies $\mathcal{T}_n \xrightarrow{D} \boldsymbol{\xi}^\intercal\boldsymbol{M}^\intercal\boldsymbol{K}^\intercal\boldsymbol{K}\boldsymbol{M}\boldsymbol{\xi}$ as $n\to\infty$ where $\boldsymbol{\xi}$ is k-dimensional standard normally distributed. This implies the first assertion since $\boldsymbol{K}\boldsymbol{M} = \operatorname{diag}(0,1,\dots,1)$ is a $k\times k$ diagonal matrix with $k-1$ times the entry 1.

Moreover, τ_n is asymptotically equivalent to

$$\frac{k\, N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{k}\right) - N_{\boldsymbol{U}}^{(n)}(c_n)}{(nc_n\,\|\boldsymbol{1}\|_D\,(k-1))^{\frac{1}{2}}} = \frac{1}{(k-1)^{\frac{1}{2}}}\boldsymbol{1}_k^\intercal\boldsymbol{K}\boldsymbol{Y}_n$$

where $\boldsymbol{1}_k := (1)_{j=1}^k$. This yields the asymptotic normal distribution of τ_n with mean 0 and variance $\frac{1}{k-1}(\boldsymbol{1}_k^\intercal\boldsymbol{K})\boldsymbol{M}\boldsymbol{M}^\intercal(\boldsymbol{1}_k^\intercal\boldsymbol{K})^\intercal = 1$, as asserted. □

Remark 2.2.13 The tests provided by Theorem 2.2.10 and Theorem 2.2.12 are based on $k+1$ estimators of the generator constant $\|\boldsymbol{1}\|_D$ in arbitrary dimension, cf. Corollary 2.2.9.

In contrast, the test of Einmahl et al. (2006), cf. Remark 2.1.9, considers two different estimators of the D-norm on the whole set $(0,1]^2$; recall that this approach is restricted to the bivariate case. However, the cited authors need to consider further technical details. For instance, they assume that some measure Λ on $[0,\infty]^2 \setminus \{(\infty,\infty)^\intercal\}$ — which satisfies $\|\boldsymbol{x}\|_D = \Lambda([\mathbf{0},\boldsymbol{\infty}] \setminus (\boldsymbol{x},\boldsymbol{\infty}])$ for non-negative $\boldsymbol{x} \in \mathbb{R}^2$ — has a continuous density on $[\mathbf{0},\boldsymbol{\infty}) \setminus \{\mathbf{0}\}$. Related tests, which test for multivariate EVCs, can be found in Ghoudi et al. (1998), Kojadinovic et al. (2011), and Berghaus et al. (2013), to name just a few.

Theorem 2.2.12 gives rise to a nice interpretation: The numerator of the test statistic $\mathcal{T}_n$ is essentially the residual sum of squares if the number of exceedances of the threshold $\left(1-\frac{c_n}{j+1}\right)\mathbf{1}$ is predicted by $\frac{j}{j+1}$ times the number of exceedances of the threshold $\left(1-\frac{c_n}{j}\right)\mathbf{1}$, $j=1,\dots,k-1$. Given that we observe i. i. d. random observations originating from a copula, Hypothesis 2.1.8 is rejected if this residual sum of squares is too large. Contrarily, if Hypothesis 2.1.8 is actually true, we would expect $N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j+1}\right) \approx \frac{j}{j+1} N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right)$ for all $j \in \{1,\dots,k-1\}$. This property corresponds to the excursion stability of a GPD, which we considered in Section 1.1: If $\boldsymbol{U}$ follows a d-dimensional GPC with corresponding D-norm $\|\cdot\|_D$, we have

$$\mathrm{P}(\boldsymbol{U} \not\leq \mathbf{1} + r\boldsymbol{x} \mid \boldsymbol{U} \not\leq (1-r)\mathbf{1}) = \frac{\|\boldsymbol{x}\|_D}{\|\mathbf{1}\|_D} \quad \text{for} \quad \boldsymbol{x} \in [-\mathbf{1},\mathbf{0}]$$

whenever $r \in \left(0,\frac{1}{d}\right]$. Thus we obtain

$$\frac{\mathrm{P}\Big(\boldsymbol{U} \not\leq (1-\frac{c}{j+1})\mathbf{1}\Big)}{\mathrm{P}\Big(\boldsymbol{U} \not\leq (1-\frac{c}{j})\mathbf{1}\Big)} = \mathrm{P}\left(\boldsymbol{U} \not\leq \mathbf{1} + \frac{c}{j}\left(-\frac{j}{j+1}\mathbf{1}\right) \,\middle|\, \boldsymbol{U} \not\leq \left(1-\frac{c}{j}\right)\mathbf{1}\right) = \frac{j}{j+1}$$

for $j=1,\dots,k$ if $c \in \left(0,\frac{1}{d}\right]$. Recall that the number of exceedances of the threshold $(1-r)\mathbf{1}$ among n independent copies of $\boldsymbol{U}$ is binomial distributed, which gives

$$\mathrm{E}\left(N_{\boldsymbol{U}}^{(n)}\left(\frac{c}{j+1}\right)\right) = \frac{j}{j+1}\,\mathrm{E}\left(N_{\boldsymbol{U}}^{(n)}\left(\frac{c}{j}\right)\right).$$

Generally, we have the following result for δ-neighborhoods:

Lemma 2.2.14 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \geq 2$. If C is a (finite dimensional) copula which is in the δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$, we obtain*

$$\max_{1\leq j\leq k-1}\left|\frac{1-C\Big((1-\frac{c}{j+1})\mathbf{1}\Big)}{1-C\Big((1-\frac{c}{j})\mathbf{1}\Big)} - \frac{j}{j+1}\right| = \mathcal{O}(c^\delta) \quad \textit{as} \quad c \to 0+.$$

Proof. We only have to consider the case $\delta < \infty$. If $\boldsymbol{U}$ has the distribution function C, there are $K > 0$ and $\varepsilon \in \big(0, K^{-\frac{1}{\delta}}\big)$ such that

$$\left|\mathrm{P}(\boldsymbol{U} \nleq (1-c)\mathbf{1}) - c\,\|\mathbf{1}\|_D\right| \leq Kc^{1+\delta} \quad \text{for} \quad c \in (0, \varepsilon),$$

cf. (2.2.1). This gives

$$\frac{j}{j+1}\frac{\|\mathbf{1}\|_D - K\big(\frac{c}{j+1}\big)^\delta}{\|\mathbf{1}\|_D + K\big(\frac{c}{j}\big)^\delta} \leq \frac{\mathrm{P}\Big(\boldsymbol{U} \nleq \big(1-\frac{c}{j+1}\big)\mathbf{1}\Big)}{\mathrm{P}\Big(\boldsymbol{U} \nleq \big(1-\frac{c}{j}\big)\mathbf{1}\Big)} \leq \frac{j}{j+1}\frac{\|\mathbf{1}\|_D + K\big(\frac{c}{j+1}\big)^\delta}{\|\mathbf{1}\|_D - K\big(\frac{c}{j}\big)^\delta},$$

and thus

$$\begin{aligned}
\max_{1\leq j\leq k-1}\left|\frac{\mathrm{P}\Big(\boldsymbol{U} \nleq \big(1-\frac{c}{j+1}\big)\mathbf{1}\Big)}{\mathrm{P}\Big(\boldsymbol{U} \nleq \big(1-\frac{c}{j}\big)\mathbf{1}\Big)} - \frac{j}{j+1}\right| &\leq \max_{1\leq j\leq k-1}\left(\frac{j}{j+1}\frac{K\big(\frac{c}{j+1}\big)^\delta + K\big(\frac{c}{j}\big)^\delta}{\|\mathbf{1}\|_D - K\big(\frac{c}{j}\big)^\delta}\right)\\
&\leq \frac{1}{2}\max_{1\leq j\leq k-1}\left(\frac{2K\big(\frac{c}{j}\big)^\delta}{1 - K\big(\frac{c}{j}\big)^\delta}\right)\\
&\leq \frac{c^\delta}{\frac{1}{K} - \varepsilon^\delta} \quad \text{for} \quad c \in (0, \varepsilon),
\end{aligned}$$

which implies the assertion. □

Continuously Distributed Data

Observing copula data in practice is rather a special case. The more common one is that we have data with unknown marginal distribution functions, i. e. the copula data are subject to a certain nuisance. However, Sklar's theorem, cf. Theorem 1.1.13, motivates to use the empirical marginal distribution functions to obtain an estimator of the underlying copula. Doing so, we will be able to adapt the test statistics of Theorem 2.2.10 and Theorem 2.2.12 to that kind of nuisance.

Consider a random vector $\boldsymbol{X} = (X_1, \dots, X_d)^\intercal$ whose distribution function F is continuous and has the margins $F_1, \dots, F_d$. Assume that the corresponding copula $C(\boldsymbol{u}) = F\big(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\big)$, $\boldsymbol{u} \in (0,1)^d$, satisfies Hypothesis 2.1.8, i. e. there exist a D-norm $\|\cdot\|_D$ and $K, \delta, \varepsilon > 0$ such that

$$\left|1 - C(\boldsymbol{u}) - \|\boldsymbol{u} - \mathbf{1}\|_D\right| \leq K\,\|\boldsymbol{u} - \mathbf{1}\|_\infty^{1+\delta} \quad \text{for all} \quad \boldsymbol{u} \in \mathcal{B}_+(\mathbf{1}, \varepsilon)$$

and in particular

$$\sup_{\boldsymbol{u}\in[(1-c)\mathbf{1},\mathbf{1}]}\left|1 - C(\boldsymbol{u}) - \|\boldsymbol{u} - \mathbf{1}\|_D\right| \leq Kc^{1+\delta} \quad \text{for all} \quad c \in (0, \varepsilon),$$

cf. (2.2.1). At first we generalize Definition 2.2.3:

Definition 2.2.15 Let $\boldsymbol{X} = (X_1, \dots, X_d)^\intercal$ be a random vector in $\mathbb{R}^d$ with a continuous distribution function F having the margins $F_1, \dots, F_d$. We call

$$S_{\boldsymbol{X}}(c) := \frac{1}{d} \sum_{i=1}^{d} \mathbb{1}_{(1-c,1]}(F_i(X_i)), \quad c \in (0,1),$$

the *sojourn time* of $\boldsymbol{X}$ above the threshold vector $\left(F_i^{-1}(1-c)\right)_{i=1}^d$ and

$$N_{\boldsymbol{X}}^{(n)}(c) := \sum_{i=1}^{n} \mathbb{1}_{(0,1]}(S_{\boldsymbol{X}^{(i)}}(c)), \quad c \in (0,1),$$

the *number of exceedances* among $\boldsymbol{X}^{(1)}, \dots \boldsymbol{X}^{(n)}$ above $\left(F_i^{-1}(1-c)\right)_{i=1}^d$, based on the independent copies $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ of $\boldsymbol{X}$.

Note that this definition complies with Definition 2.2.3 when the distribution function F of $\boldsymbol{X}$ coincides with its copula C. In any case $\boldsymbol{U} := (F_i(X_i))_{i=1}^d$ has distribution function C since F is continuous, and we have $S_{\boldsymbol{X}}(c) = S_{\boldsymbol{U}}(c)$ as well as $N_{\boldsymbol{X}}^{(n)}(c) = N_{\boldsymbol{U}}^{(n)}(c)$. The identity

$$S_{\boldsymbol{X}}(c) = \frac{1}{d} \sum_{i=1}^{d} \mathbb{1}_{\left(F_i^{-1}(1-c), \infty\right)}(X_i) \quad \text{with probability one}$$

shows that $S_{\boldsymbol{X}}(c)$ is the mean number of those components of $\boldsymbol{X}$ that exceed the threshold vector $\left(F_i^{-1}(1-c)\right)_{i=1}^d$, which justifies calling $N_{\boldsymbol{X}}^{(n)}(c)$ the number of exceedances of $\left(F_i^{-1}(1-c)\right)_{i=1}^d$.

Based on an i. i. d. sample $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ with continuous distribution function F, testing whether the underlying copula C satisfies Hypothesis 2.1.8 is particularly easy if the margins $F_1, \dots, F_d$ of F are known; just apply Theorem 2.2.10 and Theorem 2.2.12 to $(F_i(X_i))_{i=1}^d$. However, $F_1, \dots, F_d$ are typically unknown, i. e. we need an estimator of C or, more precisely, of the number of exceedances $N_{\boldsymbol{X}}^{(n)}(c)$. This estimator is obtained by replacing the marginal distribution functions of $\boldsymbol{X}$ with their empirical counterparts. In order to preserve the asymptotic normality of Corollary 2.2.9 we have to assure that the estimators of $S_{\boldsymbol{X}^{(i)}}(c)$, $i = 1, \dots, n$, are sufficiently close to their theoretical values. Precisely, it will turn out necessary to require that the sample size n for the estimation of $S_{\boldsymbol{X}}(c)$ is larger than the sample size m used to compute the corresponding number of exceedances; see (2.2.17) below and the remarks following Lemma 2.2.19.

Definition 2.2.16 Let $\boldsymbol{X} = (X_1, \dots, X_d)^\intercal$ be a random vector in $\mathbb{R}^d$ with a continuous distribution function F having the margins $F_1, \dots, F_d$. If $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ are independent copies of $\boldsymbol{X}$ with $\boldsymbol{X}^{(i)} = \left(X_1^{(i)}, \dots, X_d^{(i)}\right)^\intercal$,

$$\hat{F}_{n,r}(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_r^{(i)}), \qquad x \in \mathbb{R},$$

is the *r-th empirical margin* of $\boldsymbol{X}$ for $r \in \{1,\dots,d\}$. Furthermore,

$$\hat{S}_{\boldsymbol{X}}^{(n,i)}(c) := \frac{1}{d}\sum_{r=1}^{d} \mathbb{1}_{\left(\hat{F}_{n,r}^{-1}(1-c),\infty\right)}(X_r^{(i)}), \quad c \in (0,1),$$

denotes the *empirical sojourn time* of $\boldsymbol{X}^{(i)}$ above the threshold vector $\left(\hat{F}_{n,r}^{-1}(1-c)\right)_{r=1}^{d}$, where the quantile functions $\hat{F}_{n,r}^{-1}$, $r = 1,\dots,d$, are also computed from the sample $\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(n)}$. Finally, we denote by

$$\hat{N}_{\boldsymbol{X}}^{(m,n)}(c) := \sum_{i=1}^{m} \mathbb{1}_{(0,1]}\left(\hat{S}_{\boldsymbol{X}}^{(n,i)}(c)\right), \qquad c \in (0,1),\ m \le n, \tag{2.2.17}$$

the *empirical number of exceedances* among $\boldsymbol{X}^{(1)},\dots \boldsymbol{X}^{(m)}$ above $\left(\hat{F}_{n,r}^{-1}(1-c)\right)_{i=1}^{d}$.

For convenience, we let the empirical number of exceedances depend on the first m copies of $\boldsymbol{X}$. We obviously could also consider an arbitrary subset M of $\{1,\dots,n\}$ satisfying $|M| = m$, and define the empirical number of exceedances among $\boldsymbol{X}^{(i)}$, $i \in M$, accordingly.

Analogously to Theorem 2.2.10 and Theorem 2.2.12, the empirical number of exceedances will be the main component of the test statistics below. Note that

$$\hat{F}_{n,r}^{-1}(1-c) = \inf\left\{x \in \mathbb{R} \;\middle|\; \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_r^{(i)}) \ge n(1-c)\right\} = X_{\lceil n(1-c)\rceil:n,r}, \quad r = 1,\dots,d,$$

where $\lceil x\rceil := \min\{\ell \in \mathbb{N} \mid \ell \ge x\}$, and $X_{1:n,r} \le X_{2:n,r} \le \dots \le X_{n:n,r}$ are the order statistics of $\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(n)}$ in the r-th component, cf. Definition 2.2.18. In particular we have $U_{i:n,r} = F_r(X_{i:n,r})$ and $X_{i:n,r} = F_r^{-1}(U_{i:n,r})$ with probability one for $i = 1,\dots,n$, i. e. the distribution of $\hat{N}_{\boldsymbol{X}}^{(m,n)}(c)$ depends on the copula of $\boldsymbol{X}$ but *not* on its marginal distribution functions $F_1,\dots,F_d$.

Definition 2.2.18 For $n \in \mathbb{N}$ let $X_{1:n} \le X_{2:n} \le \dots \le X_{n:n}$ denote the ordered values of some univariate random variables $X_1,\dots,X_n$. Then $X_{\ell:n}$ is called the *ℓ-th order statistic* of $X_1,\dots,X_n$ for $\ell \in \{1,\dots,n\}$. If $\boldsymbol{X}^{(i)} = \left(X_1^{(i)},\dots,X_d^{(i)}\right)^\intercal$, $i = 1,\dots,n$, are d-dimensional random vectors, then $X_{\ell:n,r}$ denotes the ℓ-th order statistic of $X_r^{(1)},\dots,X_r^{(n)}$, where $\ell \in \{1,\dots,n\}$ and $r \in \{1,\dots,d\}$.

The following auxiliary result enables us to adapt Theorem 2.2.10 and Theorem 2.2.12 to our current setup, assuring that we may consider the empirical number of exceedances in place of its theoretical counterpart.

Lemma 2.2.19 *Suppose that* $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ *with* $\boldsymbol{X}^{(i)} = (X_1^{(i)}, \dots, X_d^{(i)})^\intercal$ *are independent copies of some* $\mathbb{R}^d$*-valued random vector* $\boldsymbol{X}$ *having continuous marginal distribution functions* $F_1, \dots, F_d$*. If* $(m_n)_{n\in\mathbb{N}}$ *and* $(c_n)_{n\in\mathbb{N}}$ *are sequences in* $\mathbb{N}$ *and* $(0,1)$*, respectively, then we obtain*

$$(m_n c_n)^{-\frac{1}{2}} \Big(N_{\boldsymbol{X}}^{(m_n)}(c_n) - \hat{N}_{\boldsymbol{X}}^{(m_n,n)}(c_n) \Big) \to 0 \quad \textit{in probability as } n \to \infty$$

whenever $\frac{m_n}{n} \log(m_n) \to 0$*,* $c_n \to 0$*, and* $m_n c_n \to \infty$ *as* $n \to \infty$*.*

Proof. Define $\boldsymbol{U}^{(i)} = (U_1^{(i)}, \dots, U_d^{(i)})^\intercal$ by $U_r^{(i)} := F_r(X_r^{(i)})$ for $i = 1, \dots, n$ and $r = 1, \dots, d$. Then we have with

$$R_n := \sum_{i=1}^{m_n} \mathbb{1}_{\times_{r=1}^d [0, U_{\lceil n(1-c_n)\rceil:n,r}]}(\boldsymbol{U}^{(i)}) \Big(1 - \mathbb{1}_{[\boldsymbol{0},(1-c_n)\boldsymbol{1}]}(\boldsymbol{U}^{(i)}) \Big)$$

and

$$T_n := \sum_{i=1}^{m_n} \mathbb{1}_{[\boldsymbol{0},(1-c_n)\boldsymbol{1}]}(\boldsymbol{U}^{(i)}) \Big(1 - \mathbb{1}_{\times_{r=1}^d [0, U_{\lceil n(1-c_n)\rceil:n,r}]}(\boldsymbol{U}^{(i)}) \Big)$$

that

$$\begin{aligned} N_{\boldsymbol{X}}^{(m_n)}(c_n) - \hat{N}_{\boldsymbol{X}}^{(m_n,n)}(c_n) &= \sum_{i=1}^{m_n} \Big(\mathbb{1}_{\times_{r=1}^d [0, U_{\lceil n(1-c_n)\rceil:n,r}]}(\boldsymbol{U}^{(i)}) - \mathbb{1}_{[\boldsymbol{0},(1-c_n)\boldsymbol{1}]}(\boldsymbol{U}^{(i)}) \Big) \\ &= R_n - T_n \quad \text{with probability one} \end{aligned}$$

since $F_1, \dots, F_d$ are continuous.

Put $\mu_n := \frac{\lceil n(1-c_n)\rceil}{n+1}$ and observe $\mu_n - (1-c_n) \in \big[-\frac{1-c_n}{n+1}, \frac{c_n}{n+1}\big)$. Then Markov's inequality shows

$$\begin{aligned} P\bigg(\frac{R_n}{(m_n c_n)^{\frac{1}{2}}} \geq \eta \bigg) &\leq \frac{1}{\eta} \Big(\frac{m_n}{c_n} \Big)^{\frac{1}{2}} P\Big(\boldsymbol{U}^{(1)} \leq \Big(U_{\lceil n(1-c_n)\rceil:n,r} \Big)_{r=1}^d, \boldsymbol{U}^{(1)} \not\leq (1-c_n)\boldsymbol{1} \Big) \\ &\leq \frac{d}{\eta} \Big(\frac{m_n}{c_n} \Big)^{\frac{1}{2}} P\Big(1 - c_n < U_1^{(1)} \leq U_{\lceil n(1-c_n)\rceil:n,1} \Big) \\ &\leq \frac{d}{\eta} \Big(\frac{m_n}{c_n} \Big)^{\frac{1}{2}} \bigg[\Big(\frac{c_n}{n+1} + \varepsilon \Big) + P\Big(U_{\lceil n(1-c_n)\rceil:n,1} - \mu_n \geq \varepsilon \Big) \bigg] \end{aligned}$$

for all $\varepsilon, \eta > 0$. We obtain furthermore

$$P\Big(U_{\lceil n(1-c_n)\rceil:n,1} - \mu_n \geq \varepsilon \Big) \leq \exp\left(- \frac{\frac{n\varepsilon^2}{\sigma_n^2}}{3\Big(1 + \frac{\varepsilon}{\sigma_n^2}\Big)} \right)$$

from Reiss (1989, Lemma 3.1.1), where $\sigma_n^2 = \mu_n(1-\mu_n)$. Note that $c_n \to 0$ and $nc_n \to \infty$ as $n \to \infty$ together with

$$\sigma_n^2 \in \left(\frac{n^2(1-c_n)}{(n+1)^2} c_n, \left(1 - \frac{nc_n}{n+1}\right) \frac{1+nc_n}{(n+1)c_n} c_n \right)$$

show $\sigma_n^2 \sim c_n$ as $n \to \infty$. Now put $\varepsilon_n := \delta_n \left(\frac{c_n}{m_n}\right)^{\frac{1}{2}}$ with $\delta_n := 2\left(\frac{m_n}{n} \log(m_n)\right)^{\frac{1}{2}}$ and obtain

$$\frac{\varepsilon_n}{\sigma_n^2} \sim \frac{\delta_n}{(m_n c_n)^{\frac{1}{2}}} \to 0 \quad \text{as well as} \quad \frac{n\varepsilon_n^2}{\sigma_n^2} \sim 4\log(m_n) \to \infty \quad \text{as} \quad n \to \infty.$$

We conclude

$$\begin{aligned} \mathrm{P}\left(\frac{R_n}{(m_n c_n)^{\frac{1}{2}}} \geq \eta \right) &\leq \frac{d}{\eta} \left(\frac{m_n}{c_n} \right)^{\frac{1}{2}} \left[\left(\frac{c_n}{n+1} + \varepsilon_n \right) + \exp\left(- \frac{\frac{n\varepsilon_n^2}{\sigma_n^2}}{3\left(1 + \frac{\varepsilon_n}{\sigma_n^2}\right)} \right) \right] \\ &= \frac{d}{\eta} \left[\left(\frac{m_n}{n+1} \frac{c_n}{n+1} \right)^{\frac{1}{2}} + \delta_n + \left(\frac{m_n}{c_n} \right)^{\frac{1}{2}} \exp\left[-\left(\frac{4}{3} + o(1) \right) \log(m_n) \right] \right] \\ &\leq \frac{d}{\eta} \left[\left(\frac{m_n}{n+1} \frac{c_n}{n+1} \right)^{\frac{1}{2}} + \delta_n + \frac{1}{(m_n c_n)^{\frac{1}{2}}} \right] \end{aligned}$$

whenever n is sufficiently large, and thus $(m_n c_n)^{-\frac{1}{2}} R_n \to 0$ in probability as $n \to \infty$.

Similarly, Markov's inequality and $1 - c_n - \mu_n \leq \frac{1-c_n}{n+1}$ show

$$\begin{aligned} \mathrm{P}\left(\frac{T_n}{(m_n c_n)^{\frac{1}{2}}} \geq \eta \right) &\leq \frac{d}{\eta} \left(\frac{m_n}{c_n} \right)^{\frac{1}{2}} \mathrm{P}\left(U_{\lceil n(1-c_n) \rceil:n,1} < U_1^{(1)} \leq 1 - c_n \right) \\ &\leq \frac{d}{\eta} \left(\frac{m_n}{c_n} \right)^{\frac{1}{2}} \left[\mathrm{P}\left(U_{\lceil n(1-c_n) \rceil:n,1} - \mu_n \leq -\varepsilon \right) + \left(\frac{1-c_n}{n+1} + \varepsilon \right) \right] \end{aligned}$$

for all $\varepsilon, \eta > 0$, and Reiss (1989, Lemma 3.1.1) gives

$$\mathrm{P}\left(U_{\lceil n(1-c_n) \rceil:n,1} - \mu_n \leq -\varepsilon \right) \leq \exp\left(- \frac{\frac{n\varepsilon^2}{\sigma_n^2}}{3\left(1 + \frac{\varepsilon}{\sigma_n^2}\right)} \right).$$

As before, we also obtain $(m_n c_n)^{-\frac{1}{2}} T_n \to 0$ in probability as $n \to \infty$, and the proof is complete. $\square$

At a first sight it appears somewhat unnatural to require $m < n$ when the number of exceedances $N_{\boldsymbol{X}}^{(n)}(c)$ above the threshold $\big(F_i^{-1}(1-c)\big)_{i=1}^d$ is estimated by

$$\hat{N}_{\boldsymbol{X}}^{(m,n)}(c) = \sum_{i=1}^{m} \mathbb{1}_{(0,1]}\Big(\hat{S}_{\boldsymbol{X}}^{(n,i)}(c)\Big).$$

However, the crucial tool in the proof of Lemma 2.2.19 was to find a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ in $(0,\infty)$ such that

$$\frac{m_n}{c_n}\,\varepsilon_n^2 \to 0 \quad \text{and} \quad \frac{m_n}{c_n}\exp\Big(-\frac{2}{3}\frac{n}{c_n}\,\varepsilon_n^2\Big) \to 0 \quad \text{as} \quad n\to\infty. \tag{2.2.20}$$

Note that (2.2.20) implies the existence of a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $(0,\infty)$ with limit 0 satisfying

$$\frac{3}{2}\frac{m_n}{n}\log\Big(\frac{m_n}{c_n\alpha_n}\Big) \le \frac{m_n}{c_n}\,\varepsilon_n^2 \to 0.$$

Since $c_n\alpha_n \to 0$, this gives $\frac{m_n}{n}\log(m_n) \to 0$, which turns out to be a condition in Lemma 2.2.19 that cannot be dropped.

Analogously to Theorem 2.2.10 and Theorem 2.2.12, the test statistics for Hypothesis 2.1.8 will be computed by considering various threshold levels and their corresponding number of exceedances simultaneously. So fix $k \in \mathbb{N}$ with $k \ge 2$ and assume that the data consist of independent realizations of an $\mathbb{R}^d$-valued random vector $\boldsymbol{X} = (X_1,\dots,X_d)^\intercal$ with continuous distribution function F. Lemma 2.2.19 justifies to estimate $\big(N_{\boldsymbol{X}}^{(m)}(\frac{c}{1}),\dots,N_{\boldsymbol{X}}^{(m)}(\frac{c}{k})\big)^\intercal$ by $\big(\hat{N}_{\boldsymbol{X}}^{(m,n)}(\frac{c}{1}),\dots,\hat{N}_{\boldsymbol{X}}^{(m,n)}(\frac{c}{k})\big)^\intercal$, where $c \in (0,1)$ is close to zero.

Theorem 2.2.21 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \ge 2$. Furthermore, let $\boldsymbol{X}$ be a random vector with continuous distribution function F such that the corresponding copula is in the δ-neighborhood of a GPC. Consider sequences $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ and $(0,1)$, respectively, satisfying $\frac{m_n}{n}\log(m_n) \to 0$, $c_n \to 0$, $m_n c_n \to \infty$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$. Then we have*

$$\hat{T}_n := \frac{\sum_{j=1}^k \Big(j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}(\frac{c_n}{j}) - \frac{1}{k}\sum_{\ell=1}^k \ell\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}(\frac{c_n}{\ell})\Big)^2}{\frac{1}{k}\sum_{\ell=1}^k \ell\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}(\frac{c_n}{\ell})} \xrightarrow{D} \sum_{j=1}^{k-1}\lambda_j\xi_j^2 \quad \textit{as} \quad n\to\infty$$

as well as

$$\hat{\mathcal{T}}_n := \frac{\sum_{j=1}^{k-1}\Big((j+1)\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}(\frac{c_n}{j+1}) - j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}(\frac{c_n}{j})\Big)^2}{\frac{1}{k}\sum_{j=1}^k j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}(\frac{c_n}{j})} \xrightarrow{D} \chi_{k-1}^2 \quad \textit{as} \quad n\to\infty$$

and

$$\hat{\tau}_n := \frac{k\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{k}\big) - \hat{N}_{\boldsymbol{X}}^{(m_n,n)}(c_n)}{\Big(\frac{k-1}{k}\sum_{j=1}^{k} j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{j}\big)\Big)^{\frac{1}{2}}} \xrightarrow{D} \boldsymbol{N}(0,1) \quad \text{as} \quad n \to \infty,$$

where

$$\lambda_j = \frac{1}{4\sin^2\Big(\frac{j}{k}\frac{\pi}{2}\Big)}, \qquad j = 1, \dots, k-1,$$

and $\xi_1, \dots, \xi_{k-1}$ *are independent and standard normally distributed random variables.*

Proof. Denote the margins of F by $F_1, \dots, F_d$. Furthermore let $\boldsymbol{X}^{(i)} = \big(X_1^{(i)}, \dots, X_d^{(i)}\big)^\intercal$, $i = 1, \dots, n$, be independent copies of $\boldsymbol{X}$ and define $\boldsymbol{U}^{(i)} = \big(U_1^{(i)}, \dots, U_d^{(i)}\big)^\intercal$ by $U_r^{(i)} := F_r\big(X_r^{(i)}\big)$ for $i = 1, \dots, n$ and $r = 1, \dots, d$. Then we have $N_{\boldsymbol{X}}^{(m_n)}(c_n) = N_{\boldsymbol{U}}^{(m_n)}(c_n)$ with probability one and the proofs of Theorem 2.2.10 and Theorem 2.2.12 carry over by considering Corollary 2.2.9 and Lemma 2.2.19. □

2.3 Testing for Functional δ-Neighborhoods

The previous section has shown how finite dimensional data can be tested whether the underlying copula is in a δ-neighborhood of a GPC. The crucial tool was Corollary 2.2.9, yielding that, roughly speaking, the observed number of exceedances over a high threshold is asymptotically normally distributed. Now we focus on the functional part of Hypothesis 2.1.8, which is restated below for convenience, and aim at generalizing the above tests to data in $C[0,1]$.

Hypothesis 2.1.8 There exist $\delta \in (0, \infty]$ and a GPC such that the copula underlying the observed data is in the δ-neighborhood of this GPC.

As before, the tests below will be based on the number of functional data exceeding some high functional threshold. The sojourn time of a stochastic process is defined analogously to Section 2.2, cf. Falk and Hofmann (2011). Due to the motivation at the beginning of Section 2.1, we restrict ourselves to processes having continuous marginal distributions.

Definition 2.3.1 Let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ with distribution function F such that all margins F_t, $t \in [0,1]$, are continuous. For $c \in (0,1)$ we call

$$S_{\boldsymbol{X}}(c) := \int_0^1 \mathbb{1}_{(1-c,1]}(F_t(X_t))\,\mathrm{d}t$$

the *sojourn time* of $\boldsymbol{X}$ above the threshold function $\left(F_t^{-1}(1-c)\right)_{t\in[0,1]}$ and

$$N_{\boldsymbol{X}}^{(n)}(c) := \sum_{i=1}^{n} \mathbb{1}_{(0,1]}(S_{\boldsymbol{X}^{(i)}}(c))$$

the *number of exceedances* among independent copies $\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(n)}$ of $\boldsymbol{X}$ above the threshold $\left(F_t^{-1}(1-c)\right)_{t\in[0,1]}$.

We have obviously $S_{\boldsymbol{X}}(c) = S_{\boldsymbol{U}}(c)$ for all $c \in (0,1)$ where $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ denotes the copula process of $\boldsymbol{X}$, cf. Definition 1.2.10. Imitating the procedure of Section 2.2, we first focus on the simple case where we observe copula processes directly. It will turn out that the results for multivariate copula data are carried over easily. In a second step we consider an extension of Lemma 2.2.19 in order to generalize the test for copula processes to more general data.

Copula Processes

Assume that our data $\boldsymbol{U}^{(1)},\dots,\boldsymbol{U}^{(n)}$ consist of independent copies of a copula process $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ in $\boldsymbol{C}[0,1]$ with distribution function C. For $c \in (0,1)$ the sojourn time of $\boldsymbol{U}$ above the threshold function $(1-c)\,\mathbb{1}_{[0,1]}$ simplifies to

$$S_{\boldsymbol{U}}(c) = \int_0^1 \mathbb{1}_{(1-c,1]}(U_t)\,\mathrm{d}t.$$

If Hypothesis 2.1.8 is true, there exist a D-norm $\|\cdot\|_D$ and $\delta \in (0,\infty]$, $\varepsilon \in (0,1)$, $K > 0$ such that

$$\left|\mathrm{P}(\boldsymbol{U} \not\leq f) - \left\|f - \mathbb{1}_{[0,1]}\right\|_D\right| \leq K \left\|f - \mathbb{1}_{[0,1]}\right\|_\infty^{1+\delta} \quad \text{for all} \quad f \in \mathcal{B}_+(\mathbb{1}_{[0,1]},\varepsilon), \tag{2.3.2}$$

and in particular

$$\left|\frac{1 - C\left((1-c)\,\mathbb{1}_{[0,1]}\right)}{c} - \left\|\mathbb{1}_{[0,1]}\right\|_D\right| \leq Kc^\delta \quad \text{for all} \quad c \in (0,\varepsilon).$$

Again, $\frac{1}{nc} N_{\boldsymbol{U}}^{(n)}(c)$ is a natural estimator of the typically unknown generator constant $\left\|\mathbb{1}_{[0,1]}\right\|_D$, with the tradeoff situation that c must be small enough to detect the δ-neighborhood, but large enough to obtain a stable estimate. It is therefore natural to let $c = c_n$ depend on the sample size and to require $c_n \to 0$ as $n \to \infty$. Later on we will also have to replace n with a sequence $(m_n)_{n\in\mathbb{N}}$ tending to infinity, cf. Section 2.2.

The various tools of Section 2.2 make it easy to obtain a functional version of Theorem 2.2.10 and Theorem 2.2.12.

Theorem 2.3.3 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \geq 2$. Furthermore let C be a (functional) copula which is in the δ-neighborhood of some GPC with D-norm $\|\cdot\|_D$. Consider a stochastic process $\boldsymbol{U}$ in $\boldsymbol{C}[0,1]$ with distribution function C, and a sequence $(c_n)_{n\in\mathbb{N}}$ in $(0,1)$ satisfying $c_n \to 0$, $nc_n \to \infty$, and $nc_n^{1+2\delta} \to 0$ as $n \to \infty$. Then we obtain*

$$T_n = \frac{\sum_{j=1}^{k}\Big(j\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{j}\big) - \frac{1}{k}\sum_{\ell=1}^{k}\ell\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{\ell}\big)\Big)^2}{\frac{1}{k}\sum_{\ell=1}^{k}\ell\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{\ell}\big)} \xrightarrow{D} \sum_{j=1}^{k-1}\lambda_j\xi_j^2 \quad as \quad n\to\infty$$

as well as

$$\mathcal{T}_n = \frac{\sum_{j=1}^{k-1}\Big((j+1)\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{j+1}\big) - j\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{j}\big)\Big)^2}{\frac{1}{k}\sum_{j=1}^{k} j\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{j}\big)} \xrightarrow{D} \chi^2_{k-1} \quad as \quad n\to\infty$$

and

$$\tau_n = \frac{k\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{k}\big) - N_{\boldsymbol{U}}^{(n)}(c_n)}{\Big(\frac{k-1}{k}\sum_{j=1}^{k} j\,N_{\boldsymbol{U}}^{(n)}\big(\frac{c_n}{j}\big)\Big)^{\frac{1}{2}}} \xrightarrow{D} \mathcal{N}(0,1) \quad as \quad n\to\infty$$

where

$$\lambda_j = \frac{1}{4\sin^2\Big(\frac{j}{k}\frac{\pi}{2}\Big)}, \qquad j = 1,\dots,k-1,$$

and $\xi_1,\dots,\xi_{k-1}$ are independent and standard normally distributed random variables.

Proof. Due to Lemma 2.2.7, the proofs of Theorem 2.2.10 and Theorem 2.2.12 carry over to stochastic processes in $\boldsymbol{C}[0,1]$. □

More General Processes

Now we aim at generalizing Theorem 2.3.3 to the case where a copula process cannot be observed directly but is subject to a certain kind of nuisance. Precisely, let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $\boldsymbol{C}[0,1]$ such that all its margins F_t, $t \in [0,1]$, are continuous but unknown. As in Section 2.2, we replace the margins with their empirical counterparts. However, in order to apply the arguments of Section 2.2, we will assume additionally that all margins of $\boldsymbol{X}$ are identical, cf. Lemma 2.3.8 below. This assumption will be dropped again in Section 2.4, where the results of this current section and those of Section 2.2 are linked with one another.

Definition 2.3.4 Let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ with continuous margins F_t, $t \in [0,1]$. If $\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(n)}$ are independent copies of $\boldsymbol{X}$ with $\boldsymbol{X}^{(i)} = \big(X_t^{(i)}\big)_{t\in[0,1]}$, we call

$$\boldsymbol{X}_{\ell:n} = (X_{\ell:n,t})_{t\in[0,1]} \quad \text{for} \quad \ell \in \{1,\dots,n\}$$

the corresponding *ℓ-th order statistic*, where $X_{\ell:n,t}$ is the ordinary ℓ-th order statistic of $X_t^{(1)},\dots,X_t^{(n)}$ for all $t \in [0,1]$.

Assume moreover $F_t = F_0$ for all $t \in [0,1]$, and denote by $\hat{F}_{n,0}$ the empirical distribution function of $X_0^{(1)},\dots,X_0^{(n)}$. Then we call for $c \in (0,1)$

$$\hat{S}_{\boldsymbol{X}}^{(n,i)}(c) := \int_0^1 \mathbb{1}_{\left(\hat{F}_{n,0}^{-1}(1-c),\infty\right)}\big(X_t^{(i)}\big)\,\mathrm{d}t \tag{2.3.5}$$

the *empirical sojourn time* of $\boldsymbol{X}^{(i)}$ above the threshold vector $\big(\hat{F}_{n,0}^{-1}(1-c)\big)_{t\in[0,1]}$, and

$$\hat{N}_{\boldsymbol{X}}^{(m,n)}(c) := \sum_{i=1}^{m} \mathbb{1}_{(0,1]}\Big(\hat{S}_{\boldsymbol{X}}^{(n,i)}(c)\Big), \qquad m \in \{1,\dots,n\}, \tag{2.3.6}$$

the *empirical number of exceedances* among $\boldsymbol{X}^{(1)},\dots\boldsymbol{X}^{(m)}$ above $\big(\hat{F}_{n,0}^{-1}(1-c)\big)_{t\in[0,1]}$.

In Section 2.2 we carried the test for copula data over to general data by exploiting the fact that the empirical number of exceedances does almost surely not depend on the margins of the data. Now we establish a functional version of this fact: The continuity of all margins F_t, $t \in [0,1]$, of $\boldsymbol{X}$ implies

$$\gamma_t(u) := \sup\{x \in \mathbb{R} \mid F_t(x) \le u\} \ge F_t^{-1}(u) \quad \text{for all} \quad t \in [0,1],\, u \in (0,1)$$

as well as

$$F_t(X_t) > 1-c \iff X_t > \gamma_t(1-c) \qquad \text{for all} \qquad t \in [0,1],\, c \in (0,1).$$

In accordance with Section 2.2 we obtain

$$S_{\boldsymbol{X}}(c) = \int_0^1 \mathbb{1}_{\left(\gamma_t(1-c),\infty\right)}(X_t)\,\mathrm{d}t = \int_0^1 \mathbb{1}_{\left(F_t^{-1}(1-c),\infty\right)}(X_t)\,\mathrm{d}t \quad \text{with probability one} \tag{2.3.7}$$

if $\mathrm{P}\Big(\bigcup_{t\in[0,1]}\{F_t^{-1}(1-c) < X_t \le \gamma_t(1-c)\}\Big) = 0$ or if F_t^{-1} is continuous at $1-c$ for λ_1-almost all $t \in [0,1]$. In particular, the following auxiliary result shows that (2.3.7) is true if all margins of $\boldsymbol{X}$ coincide.

Lemma 2.3.8 *Let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ such that all X_t, $t \in [0,1]$, are identically distributed with continuous distribution function F_0. Then we have*

$$\mathrm{P}\Big(\boldsymbol{X} \le F_0^{-1}(u)\,\mathbb{1}_{[0,1]}\Big) + \mathrm{P}\Big(\boldsymbol{X} \ge \gamma_0(u)\,\mathbb{1}_{[0,1]}\Big) = 1 \quad \text{for all} \quad u \in M_0,$$

where M_0 denotes the (at most countable) set of all discontinuities of F_0^{-1}.

Proof. Let $u \in M_0$ and $t_1, t_2 \in [0,1]$ such that $t_1 \neq t_2$. Since $\boldsymbol{X}$ has continuous sample paths, we obtain

$$\begin{aligned}\mathrm{P}\Big(X_{t_1} \le F_0^{-1}(u),\, X_{t_2} \ge \gamma_0(u)\Big) &\le \mathrm{P}\left(\bigcup_{q\in\mathbb{Q}\cap[0,1]} \Big\{X_q \in \Big(F_0^{-1}(u), \gamma_0(u)\Big)\Big\}\right)\\ &\le \sum_{q\in\mathbb{Q}\cap[0,1]} \mathrm{P}\Big(F_0^{-1}(u) < X_q \le \gamma_0(u)\Big)\\ &= 0.\end{aligned}$$

This yields

$$\begin{aligned}1 &= \mathrm{P}\left(\bigcap_{q\in\mathbb{Q}\cap[0,1]} \Big[\Big\{X_q \le F_0^{-1}(u)\Big\} \cup \Big\{X_q \ge \gamma_0(u)\Big\}\Big]\right)\\ &= \mathrm{P}\left(\left[\bigcap_{q\in\mathbb{Q}\cap[0,1]} \Big\{X_q \le F_0^{-1}(u)\Big\}\right] \cup \left[\bigcap_{q\in\mathbb{Q}\cap[0,1]} \Big\{X_q \ge \gamma_0(u)\Big\}\right]\right)\\ &= \mathrm{P}\left(\bigcap_{q\in\mathbb{Q}\cap[0,1]} \Big\{X_q \le F_0^{-1}(u)\Big\}\right) + \mathrm{P}\left(\bigcap_{q\in\mathbb{Q}\cap[0,1]} \Big\{X_q \ge \gamma_0(u)\Big\}\right)\\ &= \mathrm{P}\Big(\boldsymbol{X} \le F_0^{-1}(u)\,\mathbb{1}_{[0,1]}\Big) + \mathrm{P}\Big(\boldsymbol{X} \ge \gamma_0(u)\,\mathbb{1}_{[0,1]}\Big)\end{aligned}$$

and the proof is complete. □

In other words, we have $\big[\inf_{t\in[0,1]} X_t, \sup_{t\in[0,1]} X_t\big] \cap M_0 = \emptyset$ almost surely if all margins of $\boldsymbol{X}$ are identical, i. e. F_0 is with probability one strictly increasing on the random domain $\big[\inf_{t\in[0,1]} X_t, \sup_{t\in[0,1]} X_t\big]$. Since we obtain

$$\hat{F}_{n,0}^{-1}(1-c) = X_{\lceil n(1-c)\rceil:n,0} \quad \text{for all} \quad c \in (0,1)$$

as in Section 2.2, this proves

$$\boldsymbol{U}_{\lceil n(1-c)\rceil:n} = \Big(F_0\Big(X_{\lceil n(1-c)\rceil:n,t}\Big)\Big)_{t\in[0,1]} \quad \text{and} \quad \boldsymbol{X}_{\lceil n(1-c)\rceil:n} = \Big(F_0^{-1}\Big(U_{\lceil n(1-c)\rceil:n,t}\Big)\Big)_{t\in[0,1]}$$

with probability one, as well as

(2.3.9)
$$\hat{S}_{\boldsymbol{X}}^{(n,i)}(c) = \int_0^1 \mathbb{1}_{(X_{\lceil n(1-c)\rceil:n,0},\infty)}\big(X_t^{(i)}\big)\,\mathrm{d}t = \int_0^1 \mathbb{1}_{(U_{\lceil n(1-c)\rceil:n,0},\infty)}\big(U_t^{(i)}\big)\,\mathrm{d}t = \hat{S}_{\boldsymbol{U}}^{(n,i)}(c)$$

almost surely, where $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$ denote the copula processes of the independent copies $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ of $\boldsymbol{X}$. This yields in particular $\hat{N}_{\boldsymbol{X}}^{(m,n)}(c) = \hat{N}_{\boldsymbol{U}}^{(m,n)}(c)$ almost surely, i. e. the empirical number of exceedances only depends on the copula of $\boldsymbol{X}$ — as in the multivariate framework of Section 2.2.

We are now ready to provide a functional version of Lemma 2.2.19. In contrast to Section 2.2, this version has to incorporate Hypothesis 2.1.8 since we consider an infinite number of dimensions. As before, we need to require $m < n$ in (2.3.6) in order to justify the replacement of the numbers of exceedances in Theorem 2.3.3 with their empirical counterparts.

Lemma 2.3.10 *Let $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ be independent copies of some stochastic process $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ in $\boldsymbol{C}[0,1]$ such that X_t, $t \in [0,1]$, are identically distributed with continuous distribution function F_0. Suppose furthermore that the copula of $\boldsymbol{X}$ is in the δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$ for some $\delta \in (0,\infty]$. If $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{N}$ and $(0,1)$, respectively, then we obtain*

$$(m_n c_n)^{-\frac{1}{2}} \Big(N_{\boldsymbol{X}}^{(m_n)}(c_n) - \hat{N}_{\boldsymbol{X}}^{(m_n,n)}(c_n) \Big) \to 0 \quad \textit{in probability as } n \to \infty$$

whenever $\frac{m_n}{n} \log(m_n) \to 0$, $c_n \to 0$, $m_n c_n \to \infty$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$.

Proof. Define $\boldsymbol{U}^{(i)} = (U_t^{(i)})_{t\in[0,1]}$ by $U_t^{(i)} := F_0(X_t^{(i)})$ for $i = 1, \dots, n$ and $t \in [0,1]$, and denote the distribution function of $\boldsymbol{U}^{(1)}$ by C. Then we have with

$$R_n := \sum_{i=1}^{m_n} \mathbb{1}_{[0, U_{\lceil n(1-c_n)\rceil:n,0}]^{[0,1]}}\big(\boldsymbol{U}^{(i)}\big) \Big(1 - \mathbb{1}_{[0,1-c_n]^{[0,1]}}\big(\boldsymbol{U}^{(i)}\big) \Big)$$

and

$$T_n := \sum_{i=1}^{m_n} \mathbb{1}_{[0,1-c_n]^{[0,1]}}\big(\boldsymbol{U}^{(i)}\big) \Big(1 - \mathbb{1}_{[0, U_{\lceil n(1-c_n)\rceil:n,0}]^{[0,1]}}\big(\boldsymbol{U}^{(i)}\big) \Big)$$

that

$$\begin{aligned} N_{\boldsymbol{X}}^{(m_n)}(c_n) - \hat{N}_{\boldsymbol{X}}^{(m_n,n)}(c_n) &= \sum_{i=1}^{m_n} \Big(\mathbb{1}_{[0, U_{\lceil n(1-c_n)\rceil:n,0}]^{[0,1]}}\big(\boldsymbol{U}^{(i)}\big) - \mathbb{1}_{[0,1-c_n]^{[0,1]}}\big(\boldsymbol{U}^{(i)}\big) \Big) \\ &= R_n - T_n \quad \text{with probability one.} \end{aligned}$$

Put $\mu_n := \frac{\lceil n(1-c_n)\rceil}{n+1}$ and observe $\mu_n - (1-c_n) \in \big[-\frac{1-c_n}{n+1}, \frac{c_n}{n+1}\big)$. Markov's inequality shows

$$\mathrm{P}\bigg(\frac{R_n}{(m_n c_n)^{\frac{1}{2}}} \geq \eta \bigg) \leq \frac{1}{\eta} \Big(\frac{m_n}{c_n} \Big)^{\frac{1}{2}} \mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq U_{\lceil n(1-c_n)\rceil:n,0}\, \mathbb{1}_{[0,1]},\ \boldsymbol{U}^{(1)} \not\leq (1-c_n)\, \mathbb{1}_{[0,1]} \Big)$$

for all $\eta > 0$, and we obtain for $\varepsilon > 0$

$$\mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq U_{\lceil n(1-c_n)\rceil:n,0}\,\mathbb{1}_{[0,1]}\Big) \leq C\Big((\mu_n+\varepsilon)\,\mathbb{1}_{[0,1]}\Big) + \mathrm{P}\Big(U_{\lceil n(1-c_n)\rceil:n,0} > \mu_n + \varepsilon\Big)$$

as well as

$$\begin{aligned}
&\mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq \min\Big\{U_{\lceil n(1-c_n)\rceil:n,0}, 1-c_n\Big\}\,\mathbb{1}_{[0,1]}\Big)\\
&\quad\geq \mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq \min\{\mu_n-\varepsilon, 1-c_n\}\,\mathbb{1}_{[0,1]}\Big)\\
&\qquad - \mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq \min\{\mu_n-\varepsilon, 1-c_n\}\,\mathbb{1}_{[0,1]},\ U_{\lceil n(1-c_n)\rceil:n,0} < \mu_n - \varepsilon\Big)\\
&\qquad + \mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq \min\Big\{U_{\lceil n(1-c_n)\rceil:n,0}, 1-c_n\Big\}\,\mathbb{1}_{[0,1]},\ U_{\lceil n(1-c_n)\rceil:n,0} < \mu_n - \varepsilon\Big)\\
&\quad\geq C\Big(\min\{\mu_n-\varepsilon, 1-c_n\}\,\mathbb{1}_{[0,1]}\Big) - \mathrm{P}\Big(U_{\lceil n(1-c_n)\rceil:n,0} < \mu_n - \varepsilon\Big).
\end{aligned}$$

This yields

$$\begin{aligned}
&\mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq U_{\lceil n(1-c_n)\rceil:n,0}\,\mathbb{1}_{[0,1]},\ \boldsymbol{U}^{(1)} \not\leq (1-c_n)\,\mathbb{1}_{[0,1]}\Big)\\
&\quad= \mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq U_{\lceil n(1-c_n)\rceil:n,0}\,\mathbb{1}_{[0,1]}\Big) - \mathrm{P}\Big(\boldsymbol{U}^{(1)} \leq \min\Big\{U_{\lceil n(1-c_n)\rceil:n,0}, 1-c_n\Big\}\,\mathbb{1}_{[0,1]}\Big)\\
&\quad\leq C\Big((\mu_n+\varepsilon)\,\mathbb{1}_{[0,1]}\Big) - C\Big(\min\{\mu_n-\varepsilon, 1-c_n\}\,\mathbb{1}_{[0,1]}\Big) + \mathrm{P}\Big(\Big|U_{\lceil n(1-c_n)\rceil:n,0} - \mu_n\Big| > \varepsilon\Big).
\end{aligned}$$

If $\varepsilon \leq \frac{nc_n}{n+1}$ and n is sufficiently large, then $\mu_n + \varepsilon - 1 < 0 < 1 - \mu_n + \varepsilon$ and (2.3.2) imply

$$\begin{aligned}
&C\Big((\mu_n+\varepsilon)\,\mathbb{1}_{[0,1]}\Big) - C\Big(\min\{\mu_n-\varepsilon, 1-c_n\}\,\mathbb{1}_{[0,1]}\Big)\\
&\quad= |\min\{\mu_n-\varepsilon-1, -c_n\}|\,\big\|\mathbb{1}_{[0,1]}\big\|_D + \boldsymbol{O}\Big(|\min\{\mu_n-\varepsilon-1, -c_n\}|^{1+\delta}\Big)\\
&\qquad - |\mu_n+\varepsilon-1|\,\big\|\mathbb{1}_{[0,1]}\big\|_D + \boldsymbol{O}\Big(|\mu_n+\varepsilon-1|^{1+\delta}\Big)\\
&\quad= (\max\{1-\mu_n+\varepsilon, c_n\} + \mu_n + \varepsilon - 1)\,\big\|\mathbb{1}_{[0,1]}\big\|_D + \boldsymbol{O}\Big(\max\{1-\mu_n+\varepsilon, c_n\}^{1+\delta}\Big)\\
&\quad\leq \max\Big\{2\varepsilon, \frac{c_n}{n+1} + \varepsilon\Big\}\,\big\|\mathbb{1}_{[0,1]}\big\|_D + \boldsymbol{O}\Big(\max\{1-\mu_n+\varepsilon, c_n\}^{1+\delta}\Big),
\end{aligned}$$

and we obtain altogether

$$\begin{aligned}
\mathrm{P}\left(\frac{R_n}{(m_nc_n)^{\frac{1}{2}}} \geq \eta\right) \leq \frac{1}{\eta}\Bigg[&\left(\frac{m_n}{c_n}\right)^{\frac{1}{2}} \max\Big\{2\varepsilon, \frac{c_n}{n+1} + \varepsilon\Big\}\,\big\|\mathbb{1}_{[0,1]}\big\|_D\\
&+ \left(\frac{m_n}{c_n}\right)^{\frac{1}{2}} \mathrm{P}\Big(\Big|U_{\lceil n(1-c_n)\rceil:n,0} - \mu_n\Big| > \varepsilon\Big)\\
&+ \boldsymbol{O}\left(\max\left\{\left(\frac{m_n}{c_n}\,(1-\mu_n+\varepsilon)^{2+2\delta}\right)^{\frac{1}{2}}, \Big(m_nc_n^{1+2\delta}\Big)^{\frac{1}{2}}\right\}\right)\Bigg]
\end{aligned}$$

for all $\eta > 0$.

As in the proof of Lemma 2.2.19, Reiss (1989, Lemma 3.1.1) gives

$$\left(\frac{m_n}{c_n}\right)^{\frac{1}{2}} \mathrm{P}\left(\left|U_{\lceil n(1-c_n)\rceil:n,0} - \mu_n\right| > \varepsilon_n\right) \to 0 \quad \text{as} \quad n \to \infty$$

with $\varepsilon_n := 2\big(\frac{c_n}{n}\log(m_n)\big)^{\frac{1}{2}}$. In particular, we have for large n

$$\left(\frac{m_n}{c_n}\right)^{\frac{1}{2}} \max\left\{2\varepsilon_n, \frac{c_n}{n+1} + \varepsilon_n\right\} = 2\left(\frac{m_n}{c_n}\,\varepsilon_n^2\right)^{\frac{1}{2}} \to 0$$

as well as

$$\begin{aligned} \frac{m_n}{c_n}\,(1-\mu_n+\varepsilon_n)^{2+2\delta} &\le \frac{m_n}{c_n}\left(\left(1+\frac{1-c_n}{(n+1)\,c_n}\right)c_n+\varepsilon_n\right)^{2+2\delta} \\ &= m_n c_n^{1+2\delta}\left(1+\frac{1-c_n}{(n+1)\,c_n}+\frac{\varepsilon_n}{c_n}\right)^{2+2\delta} \to 0, \end{aligned}$$

and thus $(m_n c_n)^{-\frac{1}{2}} R_n \to 0$ in probability as $n \to \infty$. Similar arguments also show $(m_n c_n)^{-\frac{1}{2}} T_n \to 0$ in probability as $n \to \infty$, which completes the proof. $\square$

Again, we now consider various thresholds simultaneously in order to obtain an estimator of $\|\mathbb{1}_{[0,1]}\|_D$, cf. Corollary 2.2.9. The following result is implied by Lemma 2.2.7, Lemma 2.3.10, (2.3.9), and the arguments in the proofs of Theorem 2.2.10 and Theorem 2.2.12.

Theorem 2.3.11 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \ge 2$. Furthermore let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ such that $\mathrm{P}(X_t \le \cdot) = F_0$, $t \in [0,1]$, for some continuous distribution function F_0. Suppose that the copula of $\boldsymbol{X}$ is in the δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$. If $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{N}$ and $(0,1)$, respectively, satisfying $\frac{m_n}{n}\log(m_n) \to 0$, $c_n \to 0$, $m_n c_n \to \infty$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$, then we have*

$$\hat{T}_n = \frac{\sum_{j=1}^{k}\Big(j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{j}\big) - \frac{1}{k}\sum_{\ell=1}^{k}\ell\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{\ell}\big)\Big)^2}{\frac{1}{k}\sum_{\ell=1}^{k}\ell\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{\ell}\big)} \xrightarrow{D} \sum_{j=1}^{k-1}\lambda_j\xi_j^2 \quad \textit{as} \quad n\to\infty$$

as well as

$$\hat{\mathcal{T}}_n = \frac{\sum_{j=1}^{k-1}\Big((j+1)\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{j+1}\big) - j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{j}\big)\Big)^2}{\frac{1}{k}\sum_{j=1}^{k} j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\big(\frac{c_n}{j}\big)} \xrightarrow{D} \chi_{k-1}^2 \quad \textit{as} \quad n\to\infty$$

and

$$\hat{\tau}_n = \frac{k\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\left(\frac{c_n}{k}\right) - \hat{N}_{\boldsymbol{X}}^{(m_n,n)}(c_n)}{\left(\frac{k-1}{k}\sum_{j=1}^{k} j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\left(\frac{c_n}{j}\right)\right)^{\frac{1}{2}}} \xrightarrow{D} \mathcal{N}(0,1) \quad as \quad n\to\infty,$$

where

$$\lambda_j = \frac{1}{4\sin^2\left(\frac{j}{k}\frac{\pi}{2}\right)}, \qquad j = 1,\dots,k-1,$$

and $\xi_1,\dots,\xi_{k-1}$ *are independent and standard normally distributed random variables.*

2.4 Testing for Functional δ-Neighborhoods via a Grid of Points

Observing a complete process on $[0,1]$ as in the preceding section might be too strong of an assumption. For instance, measuring the height of a tide at every point of the length of a dike is presently hardly achievable. Although the data are driven by a continuous process, it is more realistic that this process is measured only at a finite set of points. This gives rise to the question whether the finite dimensional tests of Section 2.2 lead asymptotically to the same test decisions as the functional versions of Section 2.3, if the number of observation points tends to infinity in a certain manner. This current section will show that this is actually true. In particular, we will be able to drop the assumption of Section 2.3 that all margins of the underlying processes are identical. We restate the null hypothesis again for better reference:

Hypothesis 2.1.8 There exist $\delta\in(0,\infty]$ and a GPC such that the copula underlying the observed data is in the δ-neighborhood of this GPC.

Consider a functional D-norm $\|\cdot\|_D$, a grid of points $0 = t_1^{(d)} < t_2^{(d)} < \dots < t_d^{(d)} = 1$ for $d\in\mathbb{N}$, $d\geq 2$, and a stochastic process $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ in $C[0,1]$ with continuous margins F_t, $t\in[0,1]$. Assume that our data consist of independent copies $\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(n)}$ of $\boldsymbol{X}$, and that Hypothesis 2.1.8 is satisfied, i.e. the copula C of $\boldsymbol{X}$ is in the δ-neighborhood of a GPC for some $\delta\in(0,\infty]$ and some D-norm $\|\cdot\|_D$. Observe that

$$(2.4.1)\quad C_d(\boldsymbol{u}) := C\left(\sum_{r=1}^{d} u_i\,\mathbb{1}_{\left\{t_r^{(d)}\right\}} + \mathbb{1}_{[0,1]\setminus\left\{t_1^{(d)},\dots,t_d^{(d)}\right\}}\right) \quad \text{for} \quad \boldsymbol{u} = (u_1,\dots,u_d)^\intercal\in[0,1]^d$$

is the copula of the random vector $\left(X_{t_r^{(d)}}\right)_{r=1}^{d}$, and C_d is obviously in the δ-neighborhood a *finite dimensional* GPC with D-norm given by

$$\|\boldsymbol{x}\|_{D,d} := \left\| \sum_{r=1}^{d} x_i \, \mathbb{1}_{\{t_r^{(d)}\}} \right\|_D \qquad \text{for} \qquad \boldsymbol{x} = (x_1, \dots, x_d)^\intercal \in \mathbb{R}^d.$$

Consequently, $\big(Z_{t_r^{(d)}}\big)_{r=1}^{d}$ is a generator of $\|\cdot\|_{D,d}$ if $\boldsymbol{Z} = (Z_t)_{t\in[0,1]}$ is a generator of $\|\cdot\|_D$.

In order to obtain convergence of the finite dimensional projection $\|\cdot\|_{D,d}$ to the underlying functional D-norm $\|\cdot\|_D$, it is natural to let $d = d_n$ depend on the sample size, and to require

$$(2.4.2) \qquad d_n \to \infty \quad \text{and} \quad \max_{1\le r\le d_n-1} \left| t_{r+1}^{(d_n)} - t_r^{(d_n)} \right| \to 0 \quad \text{as} \quad n \to \infty,$$

i. e. the grid gets finer and finer as the number of observation points increases. Then

$$\max_{1\le r\le d_n-1} \left| Z_{t_{r+1}^{(d_n)}} - Z_{t_r^{(d_n)}} \right| \to 0 \quad \text{and} \quad \max_{1\le r\le d_n} Z_{t_r^{(d_n)}} \to \sup_{t\in[0,1]} Z_t \quad \text{as} \quad n \to \infty$$

with probability one since $\boldsymbol{Z}$ is sample continuous. The sequence of generator constants converges therefore as well:

$$(2.4.3) \qquad \|\mathbf{1}_{d_n}\|_{D,d_n} = \mathrm{E}\Big(\max_{1\le r\le d_n} Z_{t_r^{(d_n)}} \Big) \to \mathrm{E}\Big(\sup_{t\in[0,1]} Z_t \Big) = \big\|\mathbb{1}_{[0,1]}\big\|_D \quad \text{as} \quad n \to \infty$$

where the index of the vector $\mathbf{1}$ emphasizes its dimension. Recall that all of the test statistics of the previous sections highly depended on a certain estimator of the generator constant. Thus, (2.4.2) and (2.4.3) are necessary conditions for the desired asymptotic equivalence of the multivariate tests with the functional ones.

Copula Data

Suppose that the data actually consist of continuous copula processes $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$ with distribution function C. We know from (2.3.2) that there are $\varepsilon \in (0,1)$ and $K > 0$ satisfying

$$\left| 1 - C(f) - \big\|f - \mathbb{1}_{[0,1]}\big\|_D \right| \le K \,\big\|f - \mathbb{1}_{[0,1]}\big\|_\infty^{1+\delta} \quad \text{for all} \quad f \in \mathcal{B}_+(\mathbb{1}_{[0,1]}, \varepsilon)$$

if Hypothesis 2.1.8 is true, and in particular

$$\left| 1 - C_{d_n}(\boldsymbol{u}) - \|\boldsymbol{u} - \mathbf{1}_{d_n}\|_{D,d_n} \right| \le K \,\|\boldsymbol{u} - \mathbf{1}_{d_n}\|_\infty^{1+\delta} \quad \text{for all } \boldsymbol{u} \in \mathcal{B}_+(\mathbf{1}_{d_n}, \varepsilon) \text{ and } n \in \mathbb{N}.$$

This yields

$$1 - C_{d_n}((1-c)\mathbf{1}_{d_n}) = c\,\|\mathbf{1}_{d_n}\|_{D,d_n}\left(1 + \boldsymbol{O}(c^\delta)\right) \quad \text{for all} \quad c \in (0,\varepsilon). \tag{2.4.4}$$

Lemma 2.2.7 and (2.4.3) imply therefore the following version of Corollary 2.2.9.

Lemma 2.4.5 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \geq 2$. Consider sequences $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ and $(0,1)$, respectively, satisfying $c_n \to 0$, $m_n c_n \to \infty$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$. Moreover, let $(d_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{N}$, and $0 = t_1^{(d_n)} < \dots < t_{d_n}^{(d_n)} = 1$ such that (2.4.2) is true. Suppose we have independent copies $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(m_n)}$ of a copula process $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ in $\boldsymbol{C}[0,1]$ with distribution function C, and denote by $\boldsymbol{U}(n), \boldsymbol{U}^{(1)}(n), \dots, \boldsymbol{U}^{(m_n)}(n)$ the corresponding projections onto this grid, i. e. $\boldsymbol{U}(n) = \big(U_{t_r^{(d_n)}}\big)_{r=1}^{d_n}$. If C is in the δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$, then*

$$\frac{1}{k}\sum_{j=1}^{k} \frac{j}{m_n c_n}\, N_{\boldsymbol{U}(n)}^{(m_n)}\!\left(\frac{c_n}{j}\right) \to \|\mathbb{1}_{[0,1]}\|_D \quad \textit{in probability as } n \to \infty \tag{2.4.6}$$

and

$$\left((m_n c_n)^{\frac{1}{2}}\left(\frac{j}{m_n c_n}\, N_{\boldsymbol{U}(n)}^{(m_n)}\!\left(\frac{c_n}{j}\right) - \|\mathbb{1}_{[0,1]}\|_D\right)\right)_{j=1}^{k} \xrightarrow{D} \boldsymbol{N}\!\left(\boldsymbol{0}, \|\mathbb{1}_{[0,1]}\|_D\, \boldsymbol{M}\boldsymbol{M}^\intercal\right) \quad \textit{as} \quad n \to \infty$$

where $\boldsymbol{M} = \left(\mathbb{1}_{[\ell,\infty)}(j)\right)_{1\leq j,\ell\leq k}$, cf. (2.2.8).

As before, the preceding result incorporated a sequence $(m_n)_{n\in\mathbb{N}}$ having the general case in mind, where it will be necessary to require $m_n < n$. For copula data, however, it is sufficient to choose $m_n = n$.

Lemma 2.4.5 shows in particular that the finite dimensional and the functional versions of our tests are consistently linked with one another: The left side of (2.4.6) is based on the number of exceedances of *finite dimensional* projections of the underlying copula processes, whereas the limit denotes the *functional* generator constant, cf. Section 2.2 and Section 2.3. Of course, we required that the dimension of these projections tends to infinity, but we did *not* make any assumption about the speed of convergence: The increasing fineness of the projection grid together with the continuity of the underlying processes turned out to be sufficient for the desired asymptotic normality our test statistics are based upon. The theorems 2.2.10, 2.2.12, and 2.3.3 carry over:

Theorem 2.4.7 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \geq 2$. Furthermore let C be a (functional) copula which is in the δ-neighborhood of some GPC with D-norm $\|\cdot\|_D$. Consider a stochastic process $\boldsymbol{U}$ in $\boldsymbol{C}[0,1]$ with distribution function C, and its projection $\boldsymbol{U}(n)$ onto a grid $0 = t_1^{(d_n)} < \dots < t_{d_n}^{(d_n)} = 1$ satisfying (2.4.2). If $(c_n)_{n\in\mathbb{N}}$ is a sequence in $(0,1)$ such* *that* $c_n \to 0$, $nc_n \to \infty$, *and* $nc_n^{1+2\delta} \to 0$ *as* $n \to \infty$, *then we obtain*

$$T_n^* := \frac{\sum_{j=1}^k \Big(j\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{j}\big) - \frac{1}{k}\sum_{\ell=1}^k \ell\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{\ell}\big)\Big)^2}{\frac{1}{k}\sum_{\ell=1}^k \ell\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{\ell}\big)} \xrightarrow{D} \sum_{j=1}^{k-1} \lambda_j \xi_j^2 \quad as \quad n \to \infty$$

as well as

$$\mathcal{T}_n^* := \frac{\sum_{j=1}^{k-1} \Big((j+1)\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{j+1}\big) - j\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{j}\big)\Big)^2}{\frac{1}{k}\sum_{j=1}^k j\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{j}\big)} \xrightarrow{D} \chi_{k-1}^2 \quad as \quad n \to \infty$$

and

$$\tau_n^* := \frac{k\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{k}\big) - N_{\boldsymbol{U}(n)}^{(n)}(c_n)}{\Big(\frac{k-1}{k}\sum_{j=1}^k j\, N_{\boldsymbol{U}(n)}^{(n)}\big(\frac{c_n}{j}\big)\Big)^{\frac{1}{2}}} \xrightarrow{D} \mathcal{N}(0,1) \quad as \quad n \to \infty$$

where

$$\lambda_j = \frac{1}{4\sin^2\Big(\frac{j}{k}\frac{\pi}{2}\Big)}, \qquad j = 1, \dots, k-1,$$

and $\xi_1, \dots, \xi_{k-1}$ *are independent and standard normally distributed random variables.*

Continuously Distributed Data

Now we aim at an extension of Theorem 2.4.7 to processes in $C[0,1]$ having continuous margins. This will in particular overcome the disadvantage of Section 2.3, where processes with identical margins were considered exclusively. As several times before, we need to prove that the number of exceedances above a certain high threshold can be approximated by its empirical counterpart reasonably well. Here we consider the finite dimensional empirical number of exceedances, in the sense of Definition 2.2.16, of the projections to a given grid and combine the arguments needed to derive Lemma 2.2.19 and Lemma 2.3.10.

Lemma 2.4.8 *Let* $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ *be a stochastic process in* $C[0,1]$ *with continuous margins* F_t, $t \in [0,1]$, *such that its copula is in the* δ*-neighborhood of a GPC for some* $\delta \in (0,\infty]$. *Consider a grid* $0 = t_1^{(d_n)} < \dots < t_{d_n}^{(d_n)} = 1$ *satisfying* (2.4.2), *and the projection* $\boldsymbol{X}(n)$ *of* $\boldsymbol{X}$ *onto this grid. If* $(m_n)_{n\in\mathbb{N}}$ *and* $(c_n)_{n\in\mathbb{N}}$ *are sequences in* $\mathbb{N}$ *and* $(0,1)$, *respectively, and if* $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ *are independent copies of* $\boldsymbol{X}$, *then*

$$(m_n c_n)^{-\frac{1}{2}} \Big(N_{\boldsymbol{X}(n)}^{(m_n)}(c_n) - \hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}(c_n)\Big) \to 0 \quad \textit{in probability as } n \to \infty$$

whenever $\frac{m_n}{n}\log(m_n) \to 0$, $c_n \to 0$, $m_n c_n \to \infty$, $\frac{d_n^2}{m_n c_n} \to 0$, *and* $m_n c_n^{1+2\delta} \to 0$ *as* $n \to \infty$.

Proof. Define $\boldsymbol{U}^{(i)} = (U_t)_{t\in[0,1]}$ by $U_t^{(i)} := F_t\big(X_t^{(i)}\big)$ for $t \in [0,1]$ and $i = 1,\dots,n$. Furthermore, denote by $\boldsymbol{U}(n)$ the projection of $\boldsymbol{U}$ onto the grid $0 = t_1^{(d_n)} < \dots < t_{d_n}^{(d_n)} = 1$. The vectors of $\lceil n(1-c_n)\rceil$-th order statistics of $\boldsymbol{X}^{(1)}(n),\dots,\boldsymbol{X}^{(n)}(n)$ and $\boldsymbol{U}^{(1)}(n),\dots,\boldsymbol{U}^{(n)}(n)$, namely

$$\boldsymbol{Y}_{\lceil n(1-c_n)\rceil:n} = \Big(Y_{\lceil n(1-c_n)\rceil:n,r}\Big)_{r=1}^{d_n} := \Big(X_{\lceil n(1-c_n)\rceil:n,t_r^{(d_n)}}\Big)_{r=1}^{d_n}$$

and

$$\boldsymbol{V}_{\lceil n(1-c_n)\rceil:n} = \Big(V_{\lceil n(1-c_n)\rceil:n,r}\Big)_{r=1}^{d_n} := \Big(U_{\lceil n(1-c_n)\rceil:n,t_r^{(d_n)}}\Big)_{r=1}^{d_n},$$

satisfy

$$\boldsymbol{V}_{\lceil n(1-c_n)\rceil:n} = \Big(F_{t_r^{(d_n)}}\Big(Y_{\lceil n(1-c_n)\rceil:n,r}\Big)\Big)_{r=1}^{d_n} \text{ and } \boldsymbol{Y}_{\lceil n(1-c_n)\rceil:n} = \Big(F^{-1}_{t_r^{(d_n)}}\Big(V_{\lceil n(1-c_n)\rceil:n,r}\Big)\Big)_{r=1}^{d_n}$$

with probability one since the distribution function of $\boldsymbol{X}(n)$ is continuous. Put

$$R_n := \sum_{i=1}^{m_n} \mathbb{1}_{[\mathbf{0}_{d_n},\boldsymbol{V}_{\lceil n(1-c_n)\rceil:n}]}\Big(\boldsymbol{U}^{(i)}(n)\Big)\Big[1 - \mathbb{1}_{[\mathbf{0}_{d_n},(1-c_n)\mathbf{1}_{d_n}]}\Big(\boldsymbol{U}^{(i)}(n)\Big)\Big]$$

and

$$T_n := \sum_{i=1}^{m_n} \mathbb{1}_{[\mathbf{0}_{d_n},(1-c_n)\mathbf{1}_{d_n}]}\Big(\boldsymbol{U}^{(i)}(n)\Big)\Big[1 - \mathbb{1}_{[\mathbf{0}_{d_n},\boldsymbol{V}_{\lceil n(1-c_n)\rceil:n}]}\Big(\boldsymbol{U}^{(i)}(n)\Big)\Big],$$

where the subscripts of the vectors $\mathbf{0}$ and $\mathbf{1}$ emphasize their dimensions. Then we obtain

$$\begin{aligned} N^{(m_n)}_{\boldsymbol{X}(n)}(c_n) - \hat{N}^{(m_n,n)}_{\boldsymbol{X}(n)}(c_n) &= \sum_{i=1}^{m_n}\Big[\mathbb{1}_{[\mathbf{0}_{d_n},\boldsymbol{V}_{\lceil n(1-c_n)\rceil:n}]}\Big(\boldsymbol{U}^{(i)}(n)\Big) - \mathbb{1}_{[\mathbf{0}_{d_n},(1-c_n)\mathbf{1}_{d_n}]}\Big(\boldsymbol{U}^{(i)}(n)\Big)\Big] \\ &= R_n - T_n \quad \text{with probability one.} \end{aligned}$$

Put $\mu_n := \frac{\lceil n(1-c_n)\rceil}{n+1}$ and observe $\mu_n - (1-c_n) \in \big[-\frac{1-c_n}{n+1}, \frac{c_n}{n+1}\big)$. Markov's inequality shows

$$\mathrm{P}\left(\frac{R_n}{(m_n c_n)^{\frac{1}{2}}} \ge \eta\right) \le \frac{1}{\eta}\left(\frac{m_n}{c_n}\right)^{\frac{1}{2}} \mathrm{P}\Big(\boldsymbol{U}^{(1)}(n) \le \boldsymbol{V}_{\lceil n(1-c_n)\rceil:n},\ \boldsymbol{U}^{(1)}(n) \not\le (1-c_n)\mathbf{1}_{d_n}\Big)$$

for all $\eta > 0$. We have furthermore

$$\mathrm{P}\Big(\boldsymbol{U}^{(1)}(n) \le \boldsymbol{V}_{\lceil n(1-c_n)\rceil:n}\Big) \le C_{d_n}\big((\mu_n+\varepsilon)\mathbf{1}_{d_n}\big) + \mathrm{P}\Big(\max_{1\le r\le d_n} V_{\lceil n(1-c_n)\rceil:n,r} > \mu_n + \varepsilon\Big)$$

as well as

$$\begin{aligned}
&\mathrm{P}\Big(\boldsymbol{U}^{(1)}(n) \le \min\Big\{\boldsymbol{V}_{\lceil n(1-c_n)\rceil:n}, (1-c_n)\mathbf{1}_{d_n}\Big\}\Big)\\
&\quad \ge \mathrm{P}\Big(\boldsymbol{U}^{(1)}(n) \le \min\{\mu_n-\varepsilon, 1-c_n\}\mathbf{1}_{d_n}\Big)\\
&\qquad - \mathrm{P}\Big(\boldsymbol{U}^{(1)}(n) \le \min\{\mu_n-\varepsilon, 1-c_n\}\mathbf{1}_{d_n},\ \min_{1\le r\le d_n} V_{\lceil n(1-c_n)\rceil:n,r} < \mu_n-\varepsilon\Big)\\
&\quad \ge C_{d_n}(\min\{\mu_n-\varepsilon,1-c_n\}\mathbf{1}_{d_n}) - \mathrm{P}\Big(\min_{1\le r\le d_n} V_{\lceil n(1-c_n)\rceil:n,r} < \mu_n-\varepsilon\Big).
\end{aligned}$$

for $\varepsilon>0$, where C denotes the distribution function of $\boldsymbol{U}^{(1)}$ and C_{d_n} is its projection to the grid, cf. (2.4.1). This and (2.3.2) yield

$$\begin{aligned}
&\mathrm{P}\Big(\boldsymbol{U}^{(1)}(n) \le \boldsymbol{V}_{\lceil n(1-c_n)\rceil:n},\ \boldsymbol{U}^{(1)}(n) \not\le (1-c_n)\mathbf{1}_{d_n}\Big)\\
&\quad = \mathrm{P}\Big(\boldsymbol{U}^{(1)}(n) \le \boldsymbol{V}_{\lceil n(1-c_n)\rceil:n}\Big) - \mathrm{P}\Big(\boldsymbol{U}^{(1)} \le \min\Big\{\boldsymbol{V}_{\lceil n(1-c_n)\rceil:n}, (1-c_n)\mathbf{1}_{d_n}\Big\}\Big)\\
&\quad \le C_{d_n}((\mu_n+\varepsilon)\mathbf{1}_{d_n}) - C_{d_n}(\min\{\mu_n-\varepsilon,1-c_n\}\mathbf{1}_{d_n})\\
&\qquad + \sum_{r=1}^{d_n}\Big[\mathrm{P}\Big(V_{\lceil n(1-c_n)\rceil:n,r} > \mu_n+\varepsilon\Big) + \mathrm{P}\Big(V_{\lceil n(1-c_n)\rceil:n,r} < \mu_n-\varepsilon\Big)\Big]\\
&\quad \le \max\Big\{2\varepsilon, \frac{c_n}{n+1}+\varepsilon\Big\}\,\|\mathbf{1}_{d_n}\|_{D,d_n} + \boldsymbol{O}\Big(\max\{1-\mu_n+\varepsilon, c_n\}^{1+\delta}\Big)\\
&\qquad + d_n\,\mathrm{P}\Big(\Big|U_{\lceil n(1-c_n)\rceil:n,0}-\mu_n\Big| > \varepsilon\Big)
\end{aligned}$$

if $\varepsilon \le \frac{nc_n}{n+1}$ and n is sufficiently large, cf. Lemma 2.3.10. For these ε, n and for all $\eta>0$ we obtain altogether

$$\begin{aligned}
\mathrm{P}\bigg(\frac{R_n}{(m_nc_n)^{\frac12}} \ge \eta\bigg) \le \frac{1}{\eta}\bigg[&\Big(\frac{m_n}{c_n}\Big)^{\frac12}\max\Big\{2\varepsilon, \frac{c_n}{n+1}+\varepsilon\Big\}\,\|\mathbf{1}_{d_n}\|_{D,d_n}\\
&+\Big(\frac{m_n}{c_n}\Big)^{\frac12} d_n\,\mathrm{P}\Big(\Big|U_{\lceil n(1-c_n)\rceil:n,0}-\mu_n\Big|>\varepsilon\Big)\\
&+\boldsymbol{O}\bigg(\max\bigg\{\Big(\frac{m_n}{c_n}(1-\mu_n+\varepsilon)^{2+2\delta}\Big)^{\frac12}, \Big(m_nc_n^{1+2\delta}\Big)^{\frac12}\bigg\}\bigg)\bigg].
\end{aligned}$$

Now put $\varepsilon_n := 2(\frac{c_n}{n}\log(m_n))^{\frac12}$ and obtain

$$\Big(\frac{m_n}{c_n}\Big)^{\frac12} d_n\,\mathrm{P}\Big(\Big|U_{\lceil n(1-c_n)\rceil:n,0}-\mu_n\Big| > \varepsilon_n\Big) \le \frac{2d_n}{(m_nc_n)^{\frac12}} \to 0 \quad \text{as} \quad n\to\infty$$

from Reiss (1989, Lemma 3.1.1), as in Lemma 2.2.19. By considering (2.4.3) we conclude analogously to Lemma 2.3.10 that $(m_nc_n)^{-\frac12}R_n \to 0$ in probability as $n\to\infty$. Since similar arguments also show $(m_nc_n)^{-\frac12}T_n \to 0$ in probability as $n\to\infty$, the proof is complete. $\square$

Lemma 2.4.8 above shows that the empirical number of exceedances approximates the true number of exceedances, even if we observe a process only at a discrete number of points *and* the copula data are subject to a certain nuisance. Note that the requirements essentially coincide with those of Section 2.2 and Section 2.3. The main difference here is that the number of observation points must not tend too quickly to infinity. Precisely, the rate of convergence of d_n to infinity is less than the one of $(m_n c_n)^{\frac{1}{2}}$. In presence of (2.4.4), this means that the ratio of d_n^2 and the expected number of exceedances above the threshold $(1-c_n)\mathbf{1}_{d_n}$, among the first m_n projections of the underlying copula data, tends to zero as $n \to \infty$.

Due to Lemma 2.4.8 it is easy to transfer the results of Section 2.2 and Section 2.3 to the observed projections. This is done by considering various thresholds simultaneously and by applying the arguments used in the derivation of the theorems 2.2.12, 2.2.21, and 2.3.11.

Theorem 2.4.9 *Let $\delta \in (0,\infty]$ and $k \in \mathbb{N}$, $k \geq 2$. Furthermore, let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ be a stochastic process in $C[0,1]$ with continuous margins such that the copula of $\boldsymbol{X}$ is in the δ-neighborhood of a GPC. Consider a grid $0 = t_1^{(d_n)} < \cdots < t_{d_n}^{(d_n)} = 1$ satisfying (2.4.2), and the projection $\boldsymbol{X}(n)$ of $\boldsymbol{X}$ onto this grid. If $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{N}$ and $(0,1)$, respectively, satisfying $\frac{m_n}{n}\log(m_n) \to 0$, $c_n \to 0$, $m_n c_n \to \infty$, $\frac{d_n^2}{m_n c_n} \to 0$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$, then we have*

$$\hat{T}_n^* = \frac{\sum_{j=1}^k \Big(j\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{j}\big) - \frac{1}{k}\sum_{\ell=1}^k \ell\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{\ell}\big)\Big)^2}{\frac{1}{k}\sum_{\ell=1}^k \ell\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{\ell}\big)} \xrightarrow{D} \sum_{j=1}^{k-1} \lambda_j \xi_j^2 \quad \text{as} \quad n\to\infty$$

as well as

$$\hat{\mathcal{T}}_n^* = \frac{\sum_{j=1}^{k-1} \Big((j+1)\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{j+1}\big) - j\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{j}\big)\Big)^2}{\frac{1}{k}\sum_{j=1}^k j\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{j}\big)} \xrightarrow{D} \chi_{k-1}^2 \quad \text{as} \quad n\to\infty$$

and

$$\hat{\tau}_n^* = \frac{k\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{k}\big) - \hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}(c_n)}{\Big(\frac{k-1}{k}\sum_{j=1}^k j\,\hat{N}_{\boldsymbol{X}(n)}^{(m_n,n)}\big(\frac{c_n}{j}\big)\Big)^{\frac{1}{2}}} \xrightarrow{D} \mathcal{N}(0,1) \quad \text{as} \quad n\to\infty,$$

where

$$\lambda_j = \frac{1}{4\sin^2\Big(\frac{j}{k}\frac{\pi}{2}\Big)}, \qquad j = 1,\ldots,k-1,$$

and $\xi_1,\ldots,\xi_{k-1}$ are independent and standard normally distributed random variables.

3 Supplements and Related Topics

In Chapter 2, we considered essentially three different test statistics for the null hypothesis that the observed data underlies a copula of a δ-neighborhood of a GPD. We saw that all these tests are based on a rather general tool, namely Lemma 2.2.7. Thus, the asymptotic behavior of each of the test statistics under consideration is invariant under Hypothesis 2.1.8 — no matter whether the data are finite dimensional, functional and observed everywhere, or functional and observed only at a finite grid.

This current chapter is dedicated to supplement and extend the previous results. In particular, we will consider their technical prerequisites. Recall that the restriction to a certain subsample when copula data cannot be observed directly was due to the application of Reiss' inequality, as discussed following Lemma 2.2.19. In Section 3.1, we start with some prominent examples of copulas that are in a δ-neighborhood of a GPC. In contrast to that, Section 3.2 considers copulas that are not in the domain of attraction of an MSD, and thus do *not* satisfy Hypothesis 2.1.8. Both these sections are then supplemented by Section 3.3 which, on the one hand, discusses the assumption that the data emerge from a distribution with a continuous copula process and, on the other hand, deals with an approach of extending a finite dimensional copula to a functional one. Since the test statistics of Chapter 2 highly depend on the choice of additional parameters, such as suitable thresholds and, where applicable, the size of a subsample, Section 3.4 gives some practical advice. The subsequent chapter will, in addition, compare the three different test statistics and a test for similar hypothesis found in the literature by means of a simulation study.

3.1 Some Examples

The tests of Chapter 2 check whether the copula of the data generating distribution is in a δ-neighborhood of a GPC. We know so far that for each $\delta \in (0, \infty]$ and each GPC the δ-neighborhood of this GPC is non-empty and collects, roughly speaking, all distribution functions with a certain polynomial rate of convergence towards the corresponding standard MSD, cf. Lemma 2.1.6 and Lemma 2.1.10. Recall that a distribution function belongs to the ∞-neighborhood of a GPC if it coincides in its upper tail with this GPC, cf. Definition 2.1.7.

The aim of this section is to provide some prominent and non-trivial examples of copulas belonging to a certain δ-neighborhood. We begin with a rather general result on finite dimensional *and* functional copulas. Afterwards, we will focus on popular finite dimensional copula models, namely the Archimedean copulas and the normal copulas. The latter results will also be revisited in Section 3.3.

EVCs

Since a GPD is derived from the corresponding MSD, cf. Definition 1.1.34 and Definition 1.2.21, it appears natural that their copulas are linked as well. Recall that any (finite dimensional of functional) EVC C with D-norm $\|\cdot\|_D$ has the representation $C = \exp(-\|\log(\cdot)\|_D)$, cf. (1.1.31) and Definition 1.2.11. A Taylor expansion yields the following rather general result.

Lemma 3.1.1 *For any D-norm, the corresponding EVC is in the 1-neighborhood of any GPC with the same D-norm.*

Proof. We only proof the functional part of the assertion; the finite dimensional one follows from the same arguments. Since any D-norm $\|\cdot\|_D$ is monotonically increasing, i. e. $\|f\|_D \le \|g\|_D$ for $f, g \in \mathcal{E}[0,1] \cap [0,\infty)^{[0,1]}$ with $f \le g$, a Taylor expansion yields the existence of some constants $c > 0$ and $\varepsilon \in (0,1)$ satisfying

$$0 \le \|\log(f)\|_D - \|f - \mathbb{1}_{[0,1]}\|_D \le \|\log(f) - (f - \mathbb{1}_{[0,1]})\|_D \le c \left\| \left((f(t)-1)^2 \right)_{t\in[0,1]} \right\|_D$$

whenever $f \in \mathcal{B}_+(\mathbb{1}_{[0,1]}, \varepsilon)$. If C denotes the EVC with D-norm $\|\cdot\|_D$, another Taylor expansion implies thus

$$\begin{aligned}
&\left| \frac{1 - C(f) - \|f - \mathbb{1}_{[0,1]}\|_D}{\|f - \mathbb{1}_{[0,1]}\|_\infty} \right| \\
&\quad \le \left| -\sum_{k=2}^{\infty} \frac{(-\|\log(f)\|_D)^k}{k!\, \|f - \mathbb{1}_{[0,1]}\|_\infty} \right| + c \left\| \left(\frac{|f(t)-1|}{\|f - \mathbb{1}_{[0,1]}\|_\infty} \, |f(t)-1| \right)_{t\in[0,1]} \right\|_D \\
&\quad \le \frac{\|\log(f)\|_D^2}{\|f - \mathbb{1}_{[0,1]}\|_\infty} \sum_{k=2}^{\infty} \frac{\|\log(f)\|_D^{k-2}}{k!} + c \, \|f - \mathbb{1}_{[0,1]}\|_D \quad \text{for} \quad f \in \mathcal{B}_+(\mathbb{1}_{[0,1]}, \varepsilon).
\end{aligned}$$

Because of

$$(f(t)-1)^2 \le \frac{(f(t)-1)^2}{\|f - \mathbb{1}_{[0,1]}\|_\infty} \le |f(t)-1|$$

for all $f \in \mathcal{B}_+(\mathbb{1}_{[0,1]}, \varepsilon)$ and $t \in [0,1]$, we conclude

$$\begin{aligned}
\frac{\|\log(f)\|_D^2}{\|f-\mathbb{1}_{[0,1]}\|_\infty} &\le \frac{1}{\|f-\mathbb{1}_{[0,1]}\|_\infty}\left(c\left\|\left((f(t)-1)^2\right)_{t\in[0,1]}\right\|_D + \|f-\mathbb{1}_{[0,1]}\|_D\right)^2 \\
&= \left(\frac{c}{\|f-\mathbb{1}_{[0,1]}\|_\infty}\left\|\left((f(t)-1)^2\right)_{t\in[0,1]}\right\|_D + \frac{\|f-\mathbb{1}_{[0,1]}\|_D}{\|f-\mathbb{1}_{[0,1]}\|_\infty}\right) \\
&\quad \cdot \left(\frac{c}{\|f-\mathbb{1}_{[0,1]}\|_D}\left\|\left((f(t)-1)^2\right)_{t\in[0,1]}\right\|_D + 1\right)\|f-\mathbb{1}_{[0,1]}\|_D \\
&\le \left(c\|f-\mathbb{1}_{[0,1]}\|_D + \frac{\|f-\mathbb{1}_{[0,1]}\|_D}{\|f-\mathbb{1}_{[0,1]}\|_\infty}\right)(c+1)\|f-\mathbb{1}_{[0,1]}\|_D .
\end{aligned}$$

This gives

$$\left|\frac{1-C(f)-\|f-\mathbb{1}_{[0,1]}\|_D}{\|f-\mathbb{1}_{[0,1]}\|_\infty}\right| = O\left(\|f-\mathbb{1}_{[0,1]}\|_\infty\right)$$

as $\|f-\mathbb{1}_{[0,1]}\|_\infty \to 0$, since $\|\cdot\|_D$ and $\|\cdot\|_\infty$ are equivalent. □

Recall that any member of a δ-neighborhood of a GPC is in the domain of attraction of the corresponding MSD. Lemma 3.1.1 now reverses this well-known implication on the copula-level.

Archimedean Copulas

Due to their simple method of construction, the following class of finite dimensional copulas is quite popular in applications:

Definition 3.1.2 Let $\varphi : [0,1] \to [0,\infty]$ be a continuous and strictly decreasing function satisfying $\varphi(1) = 0$. Put $\varphi^{[-1]}(y) := \inf\{x \in [0,1] \mid \varphi(x) \le y\}$ for $y \in [0,\infty]$. For an integer $d \ge 2$ let $\varphi^{[-1]}$ be $d-2$ times differentiable on $(0,\infty)$ with the both properties that

$$(-1)^i \left(\varphi^{[-1]}\right)^{(i)}(y) \ge 0 \quad \text{for all } y \in (0,\infty) \text{ and } i \in \{0,\dots,d-2\}$$

and that $(-1)^{d-2}\left(\varphi^{[-1]}\right)^{(d-2)}$ is monotonically decreasing and convex. Then we call φ an *Archimedean generator* and

$$C_\varphi(\boldsymbol{u}) := \varphi^{[-1]}\left(\sum_{i=1}^d \varphi(u_i)\right) \quad \text{for} \quad \boldsymbol{u} = (u_1,\dots,u_d)^\intercal \in [0,1]^d \tag{3.1.3}$$

an *Archimedean copula.*

We have the following characterization, which is taken from McNeil and Nešlehová (2009, Section 2) and translated to our notation.

Lemma 3.1.4 (McNeil and Nešlehová, 2009) *Any Archimedean copula is a copula. Moreover, if a copula has representation* (3.1.3) *with a continuous and strictly decreasing function* $\varphi : [0,1] \to [0,\infty]$ *satisfying* $\varphi(1) = 0$, *then* φ *is an Archimedean generator.*

It is a rather mild assumption that the first derivative φ' of an Archimedean generator φ exists close to 1: If $\varphi^{[-1]}$ is differentiable in some neighborhood of 0, which is necessarily the case for $d \geq 3$, then $(\varphi^{[-1]})'$ attains only non-positive values and is monotonically increasing and continuous. This yields in particular $(\varphi^{[-1]})'(0) = \lim_{h\to 0+}(\varphi^{[-1]})'(h) \in [-\infty, 0)$. Since $\varphi^{[-1]}(\varphi(x)) = x$ for $x \in [0,1]$ as well as $\varphi(\varphi^{[-1]}(y)) = y$ for $y \in [0, \varphi(0)]$, it follows that φ is differentiable in a neighborhood of 1 with $\varphi'(1) = \frac{1}{(\varphi^{[-1]})'(0)} \in (-\infty, 0]$. If in particular $\varphi'(1) < 0$ then the corresponding Archimedean copula is in the domain of attraction of the standard MSD with D-norm $\|\cdot\|_1$, as can be verified easily by considering Theorem 1.1.39 and the Taylor expansions of $\varphi^{[-1]}$ and φ at 0 and 1, respectively. Assuming a certain shape of φ' in a neighborhood of 1 yields moreover that this Archimedean copula is actually in a δ-neighborhood of a GPC with D-norm $\|\cdot\|_1$:

Lemma 3.1.5 (Archimedean copulas) *Suppose that* φ *is an Archimedean generator that is differentiable on* $(\varepsilon, 1]$ *for some* $\varepsilon \in (0,1)$ *such that*

$$\varphi'(1) < 0 \quad \textit{and} \quad \varphi'(1-h) = \varphi'(1) + O(h^\delta) \textit{ for some } \delta > 0 \textit{ as } h \to 0+. \tag{3.1.6}$$

Then the corresponding Archimedean copula is in the δ*-neighborhood of any GPC with* D*-norm* $\|\cdot\|_1$.

Proof. The existence of φ' in a neighborhood of 1 and (3.1.6) show that $\varphi^{[-1]}$ is differentiable in some neighborhood of 0 as well as

$$\lim_{h\to 0+}(\varphi^{[-1]})'(h) = (\varphi^{[-1]})'(0) = \frac{1}{\varphi'(1)} \in (-\infty, 0).$$

Since $\varphi^{[-1]}$ is convex, we have for $h > 0$ close to 0 and $y \in (0, h)$

$$\varphi^{[-1]}(y) \leq \frac{y}{h}\,\varphi^{[-1]}(h) + \left(1 - \frac{y}{h}\right)\varphi^{[-1]}(0) = \left(1 - \frac{h-y}{h}\right)\varphi^{[-1]}(h) + \frac{h-y}{h}\,\varphi^{[-1]}(0)$$

and thus

$$(\varphi^{[-1]})'(0) \leq \frac{\varphi^{[-1]}(y) - \varphi^{[-1]}(0)}{y}$$

$$
\begin{aligned}
&\le \frac{\varphi^{[-1]}(h) - \varphi^{[-1]}(0)}{h} \\
&\le \frac{\varphi^{[-1]}(h) - \varphi^{[-1]}(y)}{h - y} \le (\varphi^{[-1]})'(h).
\end{aligned}
$$

This gives $0 \le 1 - \varphi^{[-1]}(h) = \varphi^{[-1]}(0) - \varphi^{[-1]}(h) \le \frac{-h}{\varphi'(1)}$ as well as

$$
\begin{aligned}
(\varphi^{[-1]})'(h) - (\varphi^{[-1]})'(0) &= \frac{\varphi'(1) - \varphi'(\varphi^{[-1]}(h))}{\varphi'(1)\,\varphi'(\varphi^{[-1]}(h))} \\
&= \frac{O\Big(\big|1-\varphi^{[-1]}(h)\big|^{\delta}\Big)}{\varphi'(1)\Big(\varphi'(1) + O\Big(\big|1-\varphi^{[-1]}(h)\big|^{\delta}\Big)\Big)} = O(h^{\delta}) \quad \text{as} \quad h \to 0+.
\end{aligned}
$$

Altogether we conclude

$$
\begin{aligned}
\frac{1}{|\varphi'(1)|}\sum_{i=1}^{d}\varphi(u_i) &\ge 1 - \varphi^{[-1]}\left(\sum_{i=1}^{d}\varphi(u_i)\right) \\
&\ge \left|(\varphi^{[-1]})'\left(\sum_{i=1}^{d}\varphi(u_i)\right)\right|\sum_{i=1}^{d}\varphi(u_i) \\
&= \left[\frac{1}{|\varphi'(1)|} + O\left(\left|\sum_{i=1}^{d}\varphi(u_i)\right|^{\delta}\right)\right]\sum_{i=1}^{d}\varphi(u_i) \\
&= \frac{1}{|\varphi'(1)|}\sum_{i=1}^{d}\varphi(u_i) + O\left(\left|\sum_{i=1}^{d}\varphi(u_i)\right|^{1+\delta}\right)
\end{aligned} \tag{3.1.7}
$$

as $\boldsymbol{u} = (u_1, \dots, u_d)^\intercal \to \mathbf{1}-$. Since $\varphi^{[-1]}$ is convex and φ is decreasing, we obtain

$$
\varphi((1-\lambda)x_1 + \lambda x_2) \le \varphi\Big(\varphi^{[-1]}\big((1-\lambda)\,\varphi(x_1) + \lambda\,\varphi(x_2)\big)\Big) = (1-\lambda)\,\varphi(x_1) + \lambda\,\varphi(x_2)
$$

for $x_1, x_2 \in (0,1]$ and $\lambda \in [0,1]$, and thus

$$
|x-1|\,|\varphi'(1)| \le \varphi(x) \le -|x-1|\,\varphi'(x) = |x-1|\,|\varphi'(1)| + O(|x-1|^{1+\delta}) \quad \text{as} \quad x \to 1-.
$$

This gives

$$
0 \le \frac{1}{|\varphi'(1)|}\sum_{i=1}^{d}\varphi(u_i) - \|\boldsymbol{u} - \mathbf{1}\|_1 = O\left(\sum_{i=1}^{d}|u_i - 1|^{1+\delta}\right) = O\Big(\|\boldsymbol{u} - \mathbf{1}\|_{\infty}^{1+\delta}\Big)
$$

as $\|\boldsymbol{u} - \mathbf{1}\|_{\infty} \to 0$, which implies the assertion, cf. (3.1.7). □

For general results on limit distributions of Archimedean copulas we refer to Charpentier and Segers (2009) and Larsson and Nešlehová (2011). Lemma 3.1.5 can easily be applied to the Clayton family of Archimedean copulas:

Example 3.1.8 (Clayton copula) The function $\varphi_p : [0,1] \to [0,\infty]$ defined by

$$\varphi_p(x) := \begin{cases} \frac{1}{p}\left(\frac{1}{x^p} - 1\right) & \text{if } p \in [-1,\infty) \setminus \{0\} \\ -\log(x) & \text{if } p = 0 \end{cases}$$

generates a d-dimensional Archimedean copula C_p, called *Clayton copula* with parameter p, whenever $p > \frac{-1}{d-2}$, which becomes $p > -\infty$ for $d = 2$. Lemma 3.1.5 shows that C_p is in the 1-neighborhood of a GPC with D-norm $\|\cdot\|_1$. If in particular $d = 2$ then C_{-1} is a GPC itself.

Next we consider a subclass of Archimedean copulas which, in general, do not satisfy (3.1.6) but are nevertheless in a δ-neighborhood of a GPC.

Example 3.1.9 (Gumbel-Hougaard copula) The function $\varphi_p : [0,1] \to [0,\infty]$, $\varphi_p(x) := (-\log(x))^p$ is for $p \in [1,\infty)$ an Archimedean generator in arbitrary dimension. This is due to the fact that φ_p is invertible and the n-th derivative of φ_p^{-1}, $n \in \mathbb{N}$, has the expansion

$$(\varphi_p^{-1})^{(n)}(y) = \left(-\frac{1}{p}\right)^n y^{\frac{1}{p}-n}\, q_{n-1}\left(y^{\frac{1}{p}}\right) \exp\left(-y^{\frac{1}{p}}\right)$$

where $q_{n-1}(y) = \sum_{i=0}^{n-1} a_i^{(n-1)} y^i$ is a polynomial with coefficients $a_0^{(n-1)}, \dots, a_{n-1}^{(n-1)} \geq 0$. In particular we have for $n \in \mathbb{N}$ the recursion $a_{n-1}^{(n-1)} = 1$, $a_0^{(n)} = (np-1)a_0^{(n-1)}$ and $a_i^{(n)} = a_{i-1}^{(n-1)} + (np-(i+1))\, a_i^{(n-1)}$ for $i = 1, \dots, n-1$. Although φ_p does not satisfy (3.1.6) for $p > 1$,

$$C_p(\boldsymbol{u}) = \exp\left(-\|\log(\boldsymbol{u})\|_p\right) = C_p^n\left(\boldsymbol{u}^{\frac{1}{n}}\right)$$

defines an Archimedean copula for $p \in [1,\infty)$ which is, due to Lemma 3.1.1, in the 1-neighborhood of a GPC with D-norm $\|\cdot\|_p$.

Normal Copula

We close this section with a result on the copula of a normally distributed random vector.

Definition 3.1.10 Let $\boldsymbol{X} = (X_1, \dots, X_d)^\intercal$ be d-dimensional normally distributed with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma} = (\varrho_{ij})_{1 \leq i,j \leq d}$ where $\varrho_{ii} = 1$ for $i = 1, \dots, d$. Then the distribution function of $(\Phi(X_i))_{i=1}^d$, Φ denoting the standard normal distribution function, is called the *normal copula* with correlation matrix $\boldsymbol{\Sigma}$.

Whenever all entries of a correlation matrix, apart from the main diagonal, are non-positive and greater than -1, the corresponding normal copula is in the 1-neighborhood of a GPC with D-norm $\|\cdot\|_1$. This assertion can be sharpened if the non-main-diagonal entries are strictly negative:

Lemma 3.1.11 (Normal copula) *If C is a normal copula with correlation matrix $\boldsymbol{\Sigma} = (\varrho_{ij})_{1\le i,j\le d}$ such that $-1 < \varrho_{ij} \le 0$ for $i \ne j$, then the following assertions hold:*

(i) C is in the 1-neighborhood of a GPC with D-norm $\|\cdot\|_1$.

(ii) If $\max\{\varrho_{ij} \mid 1 \le i < j \le d\} < 0$ then C is in the $(1+h)$-neighborhood of a GPC with D-norm $\|\cdot\|_1$ for all $h \in \left(0, \min_{1\le i<j\le d} \frac{2|\varrho_{ij}|}{1+\varrho_{ij}}\right)$.

Proof. If $\boldsymbol{X} = (X_1,\dots,X_d)^\intercal$ is normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, the inclusion-exclusion formula gives

$$\begin{aligned}\left|1 - C(\boldsymbol{u}) - \|\boldsymbol{u}-\mathbf{1}\|_1\right| &= \sum_{i=1}^d \mathrm{P}\Big(X_i > \Phi^{-1}(u_i)\Big) - \mathrm{P}\left(\bigcup_{i=1}^d \Big\{X_i > \Phi^{-1}(u_i)\Big\}\right)\\ &= \sum_{\substack{T\subset\{1,\dots,d\}\\ |T|\ge 2}} (-1)^{|T|}\, \mathrm{P}\left(\bigcap_{i\in T}\Big\{X_i > \Phi^{-1}(u_i)\Big\}\right)\end{aligned}$$

for $\boldsymbol{u} = (u_1,\dots,u_d)^\intercal \in (\mathbf{0},\mathbf{1})$.

Moreover, if $(\Omega,\mathcal{A})$ is a measurable space and $d \ge 2$ an integer, we have

$$\sum_{\substack{T\subset\{1,\dots,d\}\\ |T|\ge 3}} (-1)^{|T|}\, \mathrm{Q}\left(\bigcap_{i\in T} A_i\right) \le 0, \qquad A_1,\dots,A_d \in \mathcal{A}, \tag{3.1.12}$$

for any probability measure Q on $(\Omega,\mathcal{A})$. This is obvious for $d \in \{2,3\}$. If (3.1.12) holds for some $d \ge 2$ and all probability measures on $(\Omega,\mathcal{A})$, we obtain

$$\begin{aligned}&\sum_{\substack{T\subset\{1,\dots,d+1\}\\ |T|\ge 3}} (-1)^{|T|}\, \mathrm{Q}\left(\bigcap_{i\in T} A_i\right)\\ &= \sum_{\substack{T\subset\{1,\dots,d\}\\ |T|\ge 3}} (-1)^{|T|}\, \mathrm{Q}\left(\bigcap_{i\in T} A_i\right) + \sum_{\substack{T\subset\{1,\dots,d\}\\ |T|\ge 2}} (-1)^{|T|-1}\, \mathrm{Q}\left(\bigcap_{i\in T\cup\{d+1\}} A_i\right)\\ &= \sum_{\substack{T\subset\{1,\dots,d\}\\ |T|\ge 3}} (-1)^{|T|}\, \mathrm{Q}\left(A_{d+1}^{\mathrm{c}} \cap \bigcap_{i\in T} A_i\right) + \sum_{\substack{T\subset\{1,\dots,d\}\\ |T|= 2}} (-1)^{|T|-1}\, \mathrm{Q}\left(\bigcap_{i\in T\cup\{d+1\}} A_i\right)\end{aligned}$$

$$\leq \mathrm{P}\big(A_{d+1}^{\mathrm{c}}\big) \sum_{\substack{T\subset\{1,\dots,d\}\\ |T|\geq 3}} (-1)^{|T|}\,\mathrm{Q}\Bigg(\bigcap_{i\in T} A_i \,\Bigg|\, A_{d+1}^{\mathrm{c}}\Bigg) \leq 0$$

where $A_{d+1}\in\mathcal{A}$, $A_{d+1}^{\mathrm{c}} = \Omega\setminus A_{d+1}$, and $\mathrm{Q}\big(A_{d+1}^{\mathrm{c}}\big) > 0$.

This gives altogether

$$|1 - C(\boldsymbol{u}) - \|\boldsymbol{u}-\mathbf{1}\|_1| \leq \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} \mathrm{P}\Big(X_i > \Phi^{-1}(u_i),\, X_j > \Phi^{-1}(u_j)\Big). \tag{3.1.13}$$

Note that the random vector $(X_i, X_j)^\intercal$ is for $1\leq i<j\leq d$ normally distributed with mean vector $(0,0)^\intercal$ and covariance matrix $\boldsymbol{\Sigma}_{ij} := \left(\begin{smallmatrix}1 & \varrho_{ij}\\ \varrho_{ij} & 1\end{smallmatrix}\right)$. Since $\varrho_{ij}\in(-1,0]$, we obtain for $\max\{|u_i-1|, |u_j-1|\} < \frac{1}{2}$ and $x_i := \Phi^{-1}(u_i)$, $x_j := \Phi^{-1}(u_j)$ that

$$\boldsymbol{\Sigma}_{ij}^{-1}\begin{pmatrix}x_i\\ x_j\end{pmatrix} = \frac{1}{1-\varrho_{ij}^2}\begin{pmatrix}1 & -\varrho_{ij}\\ -\varrho_{ij} & 1\end{pmatrix}\begin{pmatrix}x_i\\ x_j\end{pmatrix} = \frac{1}{1-\varrho_{ij}^2}\begin{pmatrix}x_i - \varrho_{ij}x_j\\ x_j - \varrho_{ij}x_i\end{pmatrix} > \begin{pmatrix}0\\ 0\end{pmatrix}.$$

Put

$$c := \max_{1\leq k<\ell\leq d}\left((2\pi)^{\frac{|\varrho_{k\ell}|}{1+\varrho_{k\ell}}}\sqrt{\frac{1+\varrho_{k\ell}}{(1-\varrho_{k\ell})^3}}\right).$$

Then Savage (1962) — cf. Tong (1990), Hashorva and Hüsler (2003), and Hashorva (2005) — shows

$$\begin{aligned}
&\mathrm{P}\Big(X_i > \Phi^{-1}(u_i),\, X_j > \Phi^{-1}(u_j)\Big)\\
&< \frac{1}{2\pi\sqrt{1-\varrho_{ij}^2}}\exp\left(-\frac{x_i^2 - 2\,\varrho_{ij}\,x_ix_j + x_j^2}{2\left(1-\varrho_{ij}^2\right)}\right)\frac{1-\varrho_{ij}^2}{(x_i-\varrho_{ij}x_j)(x_j-\varrho_{ij}x_i)}\\
&\leq \frac{1}{2\pi}\sqrt{1-\varrho_{ij}^2}\,\exp\left(-\frac{2\,(1-\varrho_{ij})\,(\min\{x_i,x_j\})^2}{2\left(1-\varrho_{ij}^2\right)}\right)\frac{1}{(1-\varrho_{ij})^2(\min\{x_i,x_j\})^2}\\
&= \frac{1}{2\pi}\sqrt{\frac{1+\varrho_{ij}}{(1-\varrho_{ij})^3}}\,\frac{1}{(\min\{x_i,x_j\})^2}\left[\exp\left(-\frac{(\min\{x_i,x_j\})^2}{2}\right)\right]^{2-\frac{2\varrho_{ij}}{1+\varrho_{ij}}}\\
&\leq c\left(\frac{\Phi'(\min\{x_i,x_j\})}{\min\{x_i,x_j\}}\right)^2\Big(\Phi'(\min\{x_i,x_j\})\Big)^{\frac{2|\varrho_{ij}|}{1+\varrho_{ij}}}.
\end{aligned}$$

Consequently, $1-\Phi(x)\sim\frac{\Phi'(x)}{x}$ as $x\to\infty$ and $0<\Phi'(x)\leq 1$ for large x give

$$\frac{\mathrm{P}(X_i > \Phi^{-1}(u_i),\, X_j > \Phi^{-1}(u_j))}{(\max\{|u_i-1|, |u_j-1|\})^2} \leq \frac{3c}{2}\left(\frac{1-\Phi(\min\{x_i,x_j\})}{\max\{|u_i-1|, |u_j-1|\}}\right)^2 = \frac{3c}{2}$$

whenever $\max\{|u_i - 1|, |u_j - 1|\}$ is sufficiently close to 0. Equation (3.1.13) implies part (i). If $\max\{\varrho_{k\ell} \mid 1 \le k < \ell \le d\} < 0$, we obtain for $h \in \left(0, \min_{1\le k<\ell\le d} \frac{2|\varrho_{k\ell}|}{1+\varrho_{k\ell}}\right)$

$$\frac{\left(\Phi'(x)\right)^{\frac{2|\varrho_{ij}|}{(1+\varrho_{ij})h}}}{1-\Phi(x)} \sim x\left(\Phi'(x)\right)^{\frac{2|\varrho_{ij}|}{(1+\varrho_{ij})h}-1} = \frac{(2\pi)^{\frac{1}{2}+\frac{\varrho_{ij}}{(1+\varrho_{ij})h}}\, x}{\exp\left(\left(\frac{|\varrho_{ij}|}{(1+\varrho_{ij})h} - \frac{1}{2}\right)x^2\right)} \to 0 \quad \text{as} \quad x \to \infty,$$

which completes the proof. □

3.2 Some Copulas not in the Domain of Attraction of an MSD

Until now we have mainly dealt with GPD approximations of (finite dimensional and functional) copulas. Chapter 2 provided in particular some tests for a δ-neighborhood of a GPC. While the previous section provided examples of copulas satisfying this hypothesis, the question arises whether there actually are copulas that do *not* belong to a δ-neighborhood of a GPC. Due to Lemma 2.1.6 (iii), it suffices to find a copula that is not in the domain of attraction of a standard MSD. Note that constructing such a copula is by no means obvious; see Kortschak and Albrecher (2009) for a finite dimensional example. However, it turns out that modifying the approach of constructing a GPD via $-U\frac{\mathbf{1}}{\mathbf{Z}}$, cf. Theorem 1.1.41, provides parametric families of random vectors, whose copulas do not satisfy the extreme value condition (1.1.21) unless the parameter is zero. Note that these copulas get arbitrarily close to a standard GPD, which itself is in the domain of attraction of an MSD, as the parameter tends to zero. We will see in Section 3.3 how to extend these finite dimensional copulas to whole copula processes in $\mathcal{C}[0,1]$. Since the obtained parametric models are easy to simulate, they will serve as a benchmark for the simulation study in Chapter 4.

Lemma 3.2.1 *Let the random variable V have distribution function*

$$H_\lambda(u) := u\left(1 + \lambda \sin(\log(u))\right), \quad u \in [0,1], \tag{3.2.2}$$

with parameter $\lambda \in \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$. Furthermore, let the random variable $U \sim \mathcal{U}[0,1]$ be independent of V. Then the copula C_λ of the bivariate random vector

$$\boldsymbol{X} := -\frac{V}{2}\left(\frac{1}{U}, \frac{1}{1-U}\right)^{\intercal} \tag{3.2.3}$$

is for $\lambda \neq 0$ not in the domain of attraction of a multivariate MSD, whereas C_0 is a GPC *whose D-norm is given by*

$$\|\boldsymbol{x}\|_D = \|\boldsymbol{x}\|_1 - \frac{|x_1|\,|x_2|}{\|\boldsymbol{x}\|_1} \quad \textit{for} \quad \boldsymbol{x} = (x_1, x_2)^\intercal \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\}.$$

Note that $H_\lambda(0) = 0$, $H_\lambda(1) = 1$ and $H'_\lambda(u) \geq 0$ for $u \in (0,1)$. Furthermore, we obtain from $\boldsymbol{X} \in (-\infty, 0]^2$ with probability one and elementary computations that the distribution function F_λ of $-\frac{V}{U}$ is

$$F_\lambda(x) = \begin{cases} \frac{1}{|x|}\left(\frac{1}{2} + \frac{\lambda}{5}\right) & \text{for} \quad x \leq -1, \\ 1 - |x|\left(\frac{1}{2} + \frac{\lambda}{5}(2\sin(\log|x|) - \cos(\log|x|))\right) & \text{for} \quad x \in (-1, 0). \end{cases} \tag{3.2.4}$$

Thus F_λ is continuous and strictly increasing on $(-\infty, 0]$.

Proof of Lemma 3.2.1. If C_λ is in the domain of attraction of some MSD, Lemma 1.1.20 shows that the limit

$$\lim_{s \to 0+} \frac{1 - C_\lambda(1-s, 1-s)}{s}$$

exists. We prove that this is not the case for $\lambda \in [-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}] \setminus \{0\}$. Since C_λ coincides with the copula of $2\boldsymbol{X}$, we obtain for $t \in (-1, 0)$

$$\begin{aligned} \frac{1 - C_\lambda(F_\lambda(t), F_\lambda(t))}{1 - F_\lambda(t)} &= \frac{1 - \mathrm{P}\left(-\frac{V}{U} \leq t, -\frac{V}{1-U} \leq t\right)}{1 - \mathrm{P}\left(-\frac{V}{U} \leq t\right)} \\ &= \frac{1 - \mathrm{P}(V \geq |t| \max\{U, 1-U\})}{1 - \mathrm{P}(V \geq |t|\, U)} \\ &= \frac{\int_0^1 \mathrm{P}(V \leq |t| \max\{u, 1-u\})\,\mathrm{d}u}{\int_0^1 \mathrm{P}(V \leq |t|\, u)\,\mathrm{d}u} = 2\left(1 - \frac{\int_0^{\frac{1}{2}} H_\lambda(|t|\, u)\,\mathrm{d}u}{\int_0^1 H_\lambda(|t|\, u)\,\mathrm{d}u}\right). \end{aligned}$$

Since, on the one hand,

$$\int_0^c H_\lambda(|t|\, u)\,\mathrm{d}u = \frac{1}{|t|}\int_0^{|t|c} H_\lambda(u)\,\mathrm{d}u = \frac{1}{|t|}\left(\frac{(|t|\, c)^2}{2} + \lambda \int_0^{|t|c} u \sin(\log(u))\,\mathrm{d}u\right) \tag{3.2.5}$$

for $c \in [0,1]$ and, on the other hand, applying the rule of integration by parts twice gives

$$\int_0^{|t|c} u \sin(\log(u))\,\mathrm{d}u = \frac{(|t|\, c)^2}{5}\left(2\sin(\log(|t|\, c)) - \cos(\log(|t|\, c))\right), \tag{3.2.6}$$

we conclude

$$\frac{1-C_\lambda(F_\lambda(t),F_\lambda(t))}{1-F_\lambda(t)}=2\left(1-\frac{1}{4}\frac{\frac{1}{2}+\frac{\lambda}{5}\Big(2\sin\big(\log\big(\frac{|t|}{2}\big)\big)-\cos\big(\log\big(\frac{|t|}{2}\big)\big)\Big)}{\frac{1}{2}+\frac{\lambda}{5}\Big(2\sin\big(\log|t|\big)-\cos\big(\log|t|\big)\Big)}\right).$$

Considering the sequences $t_n^{(1)}=-\exp((1-2n)\pi)$, $t_n^{(2)}=-\exp((1/2-2n)\pi)$ and $s_n^{(i)}=1-F_\lambda(t_n^{(i)})$, $i\in\{1,2\}$, $n\in\mathbb{N}$, yields

$$\frac{1-C_\lambda\big(1-s_n^{(1)},1-s_n^{(1)}\big)}{s_n^{(1)}}=2-\frac{\frac{1}{2}+\frac{\lambda}{5}\Big(2\sin\big(\pi-\log(2)\big)-\cos\big(\pi-\log(2)\big)\Big)}{1+\frac{2}{5}\lambda}$$

as well as

$$\frac{1-C_\lambda\big(1-s_n^{(2)},1-s_n^{(2)}\big)}{s_n^{(2)}}=2-\frac{\frac{1}{2}+\frac{\lambda}{5}\Big(2\sin\big(\frac{\pi}{2}-\log(2)\big)-\cos\big(\frac{\pi}{2}-\log(2)\big)\Big)}{1+\frac{4}{5}\lambda}$$

and both values are distinct for $\lambda\in\big[-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\big]\setminus\{0\}$.

If $\lambda=0$, Theorem 1.1.39 and Theorem 1.1.41 show that C_0 is a GPC with D-norm

$$\|\boldsymbol{x}\|_D=\lim_{t\to 0+}\frac{1-C_0(1+t\boldsymbol{x})}{t}=2\,\mathrm{E}\big(\max\{|x_1|\,U,|x_2|\,(1-U)\}\big)=\|\boldsymbol{x}\|_1-\frac{|x_1|\,|x_2|}{\|\boldsymbol{x}\|_1}$$

for $\boldsymbol{x}=(x_1,x_2)^\intercal\in(-\infty,0]^2\setminus\{\boldsymbol{0}\}$. □

Similar results can be obtained when the denominator in (3.2.3) is exchanged:

Lemma 3.2.7 *If V is as in Lemma 3.2.1 and the random variables $U_1,U_2\sim\mathcal{U}[0,1]$ are chosen such that U_1,U_2,V are independent, then the copula C_λ of the random vector*

$$-\frac{V}{2}\left(\frac{1}{U_1},\frac{1}{U_2}\right)^\intercal$$

is not in the domain of attraction of an MSD unless $\lambda=0$. If $\lambda=0$, the corresponding D-norm is given by

$$\|\boldsymbol{x}\|_D=\|\boldsymbol{x}\|_\infty+\frac{(\|\boldsymbol{x}\|_1-\|\boldsymbol{x}\|_\infty)^2}{3\,\|\boldsymbol{x}\|_\infty}\quad\textit{for}\quad\boldsymbol{x}\neq\boldsymbol{0}.$$

Proof. We obtain

$$\frac{1-C_\lambda(F_\lambda(t),F_\lambda(t))}{1-F_\lambda(t)}=\frac{1-\mathrm{P}(V\geq|t|\max\{U_1,U_2\})}{1-\mathrm{P}(V\geq|t|\,U_1)}=\frac{\int_0^1\int_0^1 H_\lambda(|t|\max\{u_1,u_2\})\,\mathrm{d}u_2\,\mathrm{d}u_1}{\int_0^1 H_\lambda(|t|\,u)\,\mathrm{d}u}$$

for $t \in (-1, 0)$, where

$$\begin{aligned}
(3.2.8) \qquad & \int_0^1 \int_0^1 H_\lambda(|t| \max\{u_1, u_2\}) \, \mathrm{d}u_2 \, \mathrm{d}u_1 \\
&= \int_0^1 \int_0^{u_1} H_\lambda(|t| \, u_1) \, \mathrm{d}u_2 + \int_{u_1}^1 H_\lambda(|t| \, u_2) \, \mathrm{d}u_2 \, \mathrm{d}u_1 \\
&= \int_0^1 u_1 \, H_\lambda(|t| \, u_1) \, \mathrm{d}u_1 + \int_0^1 H_\lambda(|t| \, u_2) \, \mathrm{d}u_2 - \int_0^1 \int_0^{u_1} H_\lambda(|t| \, u_2) \, \mathrm{d}u_2 \, \mathrm{d}u_1.
\end{aligned}$$

The rule of integration by parts implies

$$(3.2.9) \qquad \int_0^1 u_1 \, H_\lambda(|t| \, u_1) \, \mathrm{d}u_1 = \int_0^1 H_\lambda(|t| \, u_1) \, \mathrm{d}u_1 - \int_0^1 \int_0^{u_1} H_\lambda(|t| \, u_2) \, \mathrm{d}u_2 \, \mathrm{d}u_1$$

and (3.2.5), (3.2.6) show

$$\begin{aligned}
& \int_0^1 \int_0^{u_1} H_\lambda(|t| \, u_2) \, \mathrm{d}u_2 \, \mathrm{d}u_1 \\
&= \frac{1}{|t|} \int_0^1 \frac{(|t| \, u_1)^2}{2} + \lambda \frac{(|t| \, u_1)^2}{5} \big(2 \sin(\log(|t| \, u_1)) - \cos(\log(|t| \, u_1))\big) \, \mathrm{d}u_1 \\
&= \frac{1}{|t|^2} \int_0^{|t|} \frac{u^2}{2} + \lambda \frac{u^2}{5} \big(2 \sin(\log(u)) - \cos(\log(u))\big) \, \mathrm{d}u \\
&= \frac{1}{2 \, |t|^2} \int_0^{|t|} u^2 \big(1 + \lambda \sin(\log(u))\big) \, \mathrm{d}u - \frac{\lambda}{10 \, |t|^2} \int_0^{|t|} u^2 \big(\sin(\log(u)) + 2 \cos(\log(u))\big) \, \mathrm{d}u \\
&= \frac{1}{2} \int_0^1 u \, H_\lambda(|t| \, u) \, \mathrm{d}u - \frac{\lambda \, |t|}{20} \big(\sin(\log |t|) + \cos(\log |t|)\big)
\end{aligned}$$

since

$$\begin{aligned}
& \int_0^{|t|} u^2 \big(\sin(\log(u)) + 2 \cos(\log(u))\big) \, \mathrm{d}u \\
&= \frac{|t|^3}{3} \big(\sin(\log |t|) + 2 \cos(\log |t|)\big) - \frac{1}{3} \int_0^{|t|} u^2 \big(\cos(\log(u)) - 2 \sin(\log(u))\big) \, \mathrm{d}u \\
&= \frac{|t|^3}{3} \big(\sin(\log |t|) + 2 \cos(\log |t|)\big) \\
&\quad - \frac{|t|^3}{9} \big(\cos(\log |t|) - 2 \sin(\log |t|)\big) - \frac{1}{9} \int_0^{|t|} u^2 \big(\sin(\log(u)) + 2 \cos(\log(u))\big) \, \mathrm{d}u.
\end{aligned}$$

Now we conclude from (3.2.5), (3.2.6), (3.2.8) and (3.2.9) that

$$\int_0^1 u \, H_\lambda(|t| \, u) \, \mathrm{d}u = \frac{2}{3} \left(\int_0^1 H_\lambda(|t| \, u) \, \mathrm{d}u + \frac{\lambda \, |t|}{20} \big(\sin(\log |t|) + \cos(\log |t|)\big) \right)$$

and

$$\frac{1-C_\lambda(F_\lambda(t),F_\lambda(t))}{1-F_\lambda(t)}=\frac{4}{3}\left(1+\lambda\frac{\sin(\log|t|)+\cos(\log|t|)}{10+4\lambda\,(2\sin(\log|t|)-\cos(\log|t|))}\right).$$

Considering the both sequences $(t_n^{(1)})_{n\in\mathbb{N}}$ and $(t_n^{(2)})_{n\in\mathbb{N}}$ in the proof of Lemma 3.2.1 yields

$$\frac{1-C_\lambda(F_\lambda(t_n^{(1)}),F_\lambda(t_n^{(1)}))}{1-F_\lambda(t_n^{(1)})}=\frac{4}{3}\left(1-\frac{\lambda}{10+4\lambda}\right)$$

and

$$\frac{1-C_\lambda(F_\lambda(t_n^{(2)}),F_\lambda(t_n^{(2)}))}{1-F_\lambda(t_n^{(2)})}=\frac{4}{3}\left(1+\frac{\lambda}{10+8\lambda}\right),$$

i.e. $\frac{1-C_\lambda(1-s,1-s)}{s}$ has for $\lambda\in[-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}]\setminus\{0\}$ at least two different accumulation points as $s\to 0+$.

Furthermore, we have for $\lambda=0$ and $\boldsymbol{x}=(x_1,x_2)^\intercal<\boldsymbol{0}$

$$\begin{aligned}\|\boldsymbol{x}\|_D&=2\,\mathrm{E}(\max\{|x_1|\,U_1,|x_2|\,U_2\})\\&=2\int_0^1\int_0^1\max\{|x_1|\,u_1,|x_2|\,u_2\}\,\mathrm{d}u_2\,\mathrm{d}u_1\\&=2\int_0^1\int_0^{1-\left(1-\frac{|x_1|}{|x_2|}u_1\right)_+}|x_1|\,u_1\,\mathrm{d}u_2+\int_{1-\left(1-\frac{|x_1|}{|x_2|}u_1\right)_+}^1|x_2|\,u_2\,\mathrm{d}u_2\,\mathrm{d}u_1\end{aligned}$$

since $1-\left(1-\frac{|x_1|}{|x_2|}u_1\right)_+=\min\{1,\frac{|x_1|}{|x_2|}u_1\}$. This gives

$$\begin{aligned}\|\boldsymbol{x}\|_D&=2\int_0^1\left(1-\left(1-\frac{|x_1|}{|x_2|}u_1\right)_+\right)|x_1|\,u_1+\frac{|x_2|}{2}\left[1-\left(1-\left(1-\frac{|x_1|}{|x_2|}u_1\right)_+\right)^2\right]\mathrm{d}u_1\\&=|x_2|+2\int_0^1\left(1-\left(1-\frac{|x_1|}{|x_2|}u_1\right)_+\right)\left[|x_1|\,u_1-\frac{|x_2|}{2}\left(1-\left(1-\frac{|x_1|}{|x_2|}u_1\right)_+\right)\right]\mathrm{d}u_1,\end{aligned}$$

and thus

$$\begin{aligned}\frac{\|\boldsymbol{x}\|_D-|x_2|}{2}&=\int_0^{1-\left(1-\frac{|x_2|}{|x_1|}\right)_+}\frac{|x_1|^2}{2\,|x_2|}u^2\,\mathrm{d}u+\int_{1-\left(1-\frac{|x_2|}{|x_1|}\right)_+}^1|x_1|\,u-\frac{|x_2|}{2}\,\mathrm{d}u\\&=\frac{|x_1|^2}{6\,|x_2|}\left(1-\left(1-\frac{|x_2|}{|x_1|}\right)_+\right)^3\\&\quad+\frac{|x_1|}{2}\left[1-\left(1-\left(1-\frac{|x_2|}{|x_1|}\right)_+\right)^2\right]-\frac{|x_2|}{2}\left(1-\frac{|x_2|}{|x_1|}\right)_+.\end{aligned}$$

Since

$$1 - \left(1 - \frac{|x_2|}{|x_1|}\right)_+ = 1 - \frac{1}{|x_1|}(|x_1| - |x_2|)_+ = 1 - \frac{1}{|x_1|}(\|\boldsymbol{x}\|_\infty - |x_2|) = \frac{\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty}{|x_1|}$$

and $\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty = \min\{|x_1|, |x_2|\}$, we get

$$\begin{aligned}
\|\boldsymbol{x}\|_D &= \frac{(\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty)^3}{3\,|x_1|\,|x_2|} + |x_1| - \frac{(\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty)^2}{|x_1|} + |x_2|\,\frac{\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty}{|x_1|} \\
&= \frac{(\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty)^2}{3\,\|\boldsymbol{x}\|_\infty} + |x_1| + (\|\boldsymbol{x}\|_\infty - |x_1|)\frac{\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty}{|x_1|} \\
&= \frac{(\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty)^2}{3\,\|\boldsymbol{x}\|_\infty} + |x_1| + \frac{|x_1|\,|x_2|}{|x_1|} - (\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty) \\
&= \|\boldsymbol{x}\|_\infty + \frac{(\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty)^2}{3\,\|\boldsymbol{x}\|_\infty},
\end{aligned}$$

which completes the proof. □

Lemma 3.2.1 and Lemma 3.2.7 give rise to two one-parametric models of bivariate random vectors that are not in the domain of attraction of any MSD unless the parameter is zero. This was achieved by replacing the nominator of a GPD random vector $-U\frac{\mathbf{1}}{\boldsymbol{Z}}$ — cf. Theorem 1.1.41 and Theorem 1.2.26 — with a random variable V that has an oscillating density close to its lower endpoint, as illustrated in Figure 3.2.10. Thus, the density of $-\frac{U}{Z_i}$, $i \in \{1, 2\}$, is oscillating close to its upper endpoint, cf. (3.2.4). One might expect that the copula of $-V\frac{\mathbf{1}}{\boldsymbol{Z}}$ is not in a domain of attraction for any generator $\boldsymbol{Z}$. However, if U and V are as in Lemma 3.2.1, the random vector $-\frac{V}{2}\left(\frac{1}{U}, \frac{1}{U}\right)^\intercal$ has the copula

$$\begin{aligned}
C_\lambda(\boldsymbol{u}) &= \mathrm{P}\left(-\frac{V}{U} \le F_\lambda^{-1}(u_1), -\frac{V}{U} \le F_\lambda^{-1}(u_2)\right) \\
&= F_\lambda\left(\min\left\{F_\lambda^{-1}(u_1), F_\lambda^{-1}(u_2)\right\}\right) \\
&= \min\{u_1, u_2\} \qquad \text{for} \qquad \boldsymbol{u} = (u_1, u_2)^\intercal \in (0,1)^2
\end{aligned}$$

since F_λ is strictly increasing. Thus, the copula does not depend on λ and Theorem 1.2.28 shows that C_λ is in the domain of attraction of the standard MSD with D-norm $\|\cdot\|_\infty$. In fact, C_λ is an EVC since

$$\min\{u_1, u_2\} = \left(\min\left\{u_1^{1/n}, u_2^{1/n}\right\}\right)^n \quad \text{for all} \quad n \in \mathbb{N},$$

cf. Definition 1.1.17 and Lemma 1.1.32.

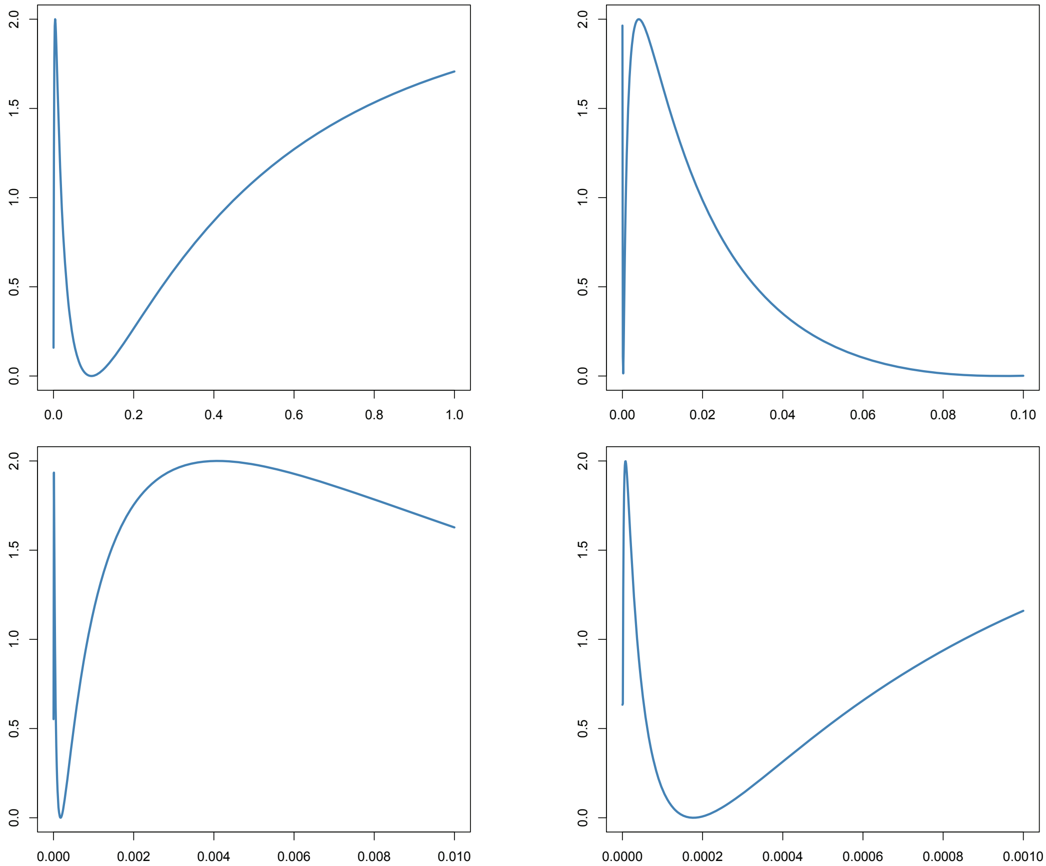

Figure 3.2.10 Plots of the density of H_λ in Lemma 3.2.1 for $\lambda = \frac{\sqrt{2}}{2}$.

3.3 Continuous Copula Processes

Copulas are quite powerful tools to characterize the domain of attraction of an MSD, which motivated to focus on standard MSDs. Recall that Sklar's Theorem (Theorem 1.1.13) assures that any finite dimensional distribution function has a copula. An extension to $C[0,1]$ that is in full accordance with the finite dimensional setting would require a stochastic process $\boldsymbol{X}$ in $C[0,1]$ to have a *continuous* copula process, cf. Definition 1.2.8. Indeed we noted in Lemma 1.2.9 that a continuous copula process exists if all margins of $\boldsymbol{X}$ are continuous. Chapter 2 therefore considered these kind of processes only.

In the first part of this section, we will see that there are stochastic processes in $C[0,1]$ that do not have a continuous copula process. However, we show that those processes can be approximated reasonably by another process that does have a copula process in $C[0,1]$. This yields that the requirement of Chapter 2, that there is a copula process having continuous sample paths, is not too strong. The second part deals with an interpolation method that transforms a copula random vector into a whole copula process in $C[0,1]$. This implies in particular that the parametric families in Section 3.2 can be generalized to the space $C[0,1]$, i. e. there are stochastic processes in $C[0,1]$ which are not in the domain of attraction of an MSP.

A Continuous Process that has No Continuous Copula Process

Inspired by Hofmann (2012, Section 2.4), consider $p \in (0,1)$ and let $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ have the distribution defined by $\mathrm{P}(\boldsymbol{X} = \mathrm{id}_{[0,1]}) = p$ and $\mathrm{P}(\boldsymbol{X} = \mathbb{1}_{[0,1]} - \mathrm{id}_{[0,1]}) = 1-p$, where $\mathrm{id}_{[0,1]}$ denotes the identity function of the interval $[0,1]$, cf. Figure 3.3.1. Then X_t has the distribution function $F_t(x) = p\,\mathbb{1}_{[t,\infty)}(x) + (1-p)\,\mathbb{1}_{[1-t,\infty)}(x)$, $x \in \mathbb{R}$, and the quantile function

$$F_t^{-1}(u) = \begin{cases} t\,\mathbb{1}_{(0,p]}(u) + (1-t)\,\mathbb{1}_{(p,1)}(u), & t \in [0,\frac{1}{2}], \\ (1-t)\,\mathbb{1}_{(0,1-p]}(u) + t\,\mathbb{1}_{(1-p,1)}(u), & t \in (\frac{1}{2},1], \end{cases} \qquad u \in (0,1).$$

Let $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ be an arbitrary copula process in $C[0,1]$. Then we have, on the one hand,

$$\mathrm{P}\Big(\big(F_t^{-1}(U_t)\big)_{t\in[0,1]} = \mathrm{id}_{[0,1]}\Big) \le \mathrm{P}\Big(\boldsymbol{U}\,\mathbb{1}_{[0,\frac{1}{2})} \le p\,\mathbb{1}_{[0,\frac{1}{2})},\, \boldsymbol{U}\,\mathbb{1}_{(\frac{1}{2},1]} > (1-p)\,\mathbb{1}_{(\frac{1}{2},1]}\Big) = 0$$

for $p < \frac{1}{2}$ and, on the other hand,

$$\mathrm{P}\Big(\big(F_t^{-1}(U_t)\big)_{t\in[0,1]} = \mathrm{id}_{[0,1]}\Big) \le \mathrm{P}\Big(1-p \le U_{\frac{1}{2}} \le p\Big) = 2p-1$$

for $p \ge \frac{1}{2}$. Thus $\mathrm{P}\big((F_t^{-1}(U_t))_{t\in[0,1]} = \mathrm{id}_{[0,1]}\big) \ne \mathrm{P}(\boldsymbol{X} = \mathrm{id}_{[0,1]})$, i. e. the stochastic process $\boldsymbol{X}$ does not have a continuous copula process.

However, according to Rüschendorf (2009), $\boldsymbol{X}$ does have a copula process in $\mathcal{E}[0,1]$; just define $\boldsymbol{U} = (U_t)_{t\in[0,1]}$ by

$$U_t = p\Big(\mathbb{1}_{(t,\infty)}(X_t) + V\,\mathbb{1}_{\{t\}}(X_t)\Big) + (1-p)\Big(\mathbb{1}_{(1-t,\infty)}(X_t) + V\,\mathbb{1}_{\{1-t\}}(X_t)\Big) \tag{3.3.2}$$

where $V \sim \mathcal{U}[0,1]$ is independent of $\boldsymbol{X}$. In this case one has $U_t \sim \mathcal{U}[0,1]$, $t \in [0,1]$, and

$$\begin{aligned} 1 &= \mathrm{P}\Big(\boldsymbol{X} = \mathrm{id}_{[0,1]}\Big) + \mathrm{P}\Big(\boldsymbol{X} = \mathbb{1}_{[0,1]} - \mathrm{id}_{[0,1]}\Big) \\ &= \mathrm{P}\Big((pV)_{t\in[0,\frac{1}{2})} \le p\,\mathbb{1}_{[0,\frac{1}{2})},\, (pV + (1-p))_{t\in(\frac{1}{2},1]} > (1-p)\,\mathbb{1}_{(\frac{1}{2},1]},\, \boldsymbol{X} = \mathrm{id}_{[0,1]}\Big) \\ &\quad + \mathrm{P}\Big((p+(1-p)V)_{t\in[0,\frac{1}{2})} > p\,\mathbb{1}_{[0,\frac{1}{2})},\, ((1-p)V)_{t\in(\frac{1}{2},1]} \le (1-p)\,\mathbb{1}_{(\frac{1}{2},1]}, \\ &\qquad\quad \boldsymbol{X} = \mathbb{1}_{[0,1]} - \mathrm{id}_{[0,1]}\Big) \end{aligned}$$

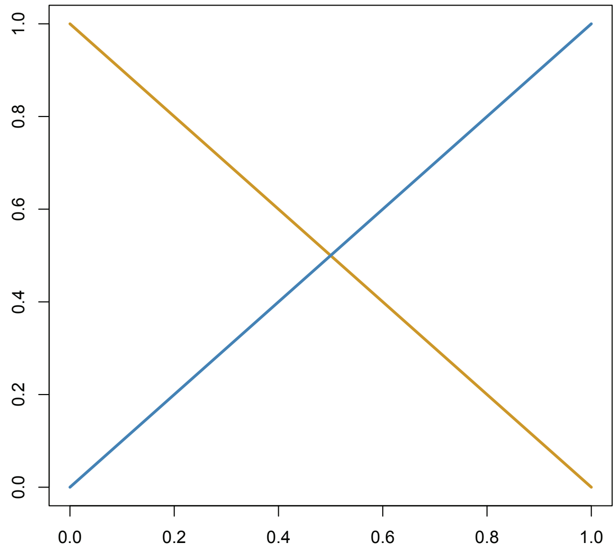

Figure 3.3.1 The both sample paths of the process $\boldsymbol{X}$.

$$
\begin{aligned}
&= \mathrm{P}\Big((U_t)_{t\in[0,\frac{1}{2})} \le p\,\mathbb{1}_{[0,\frac{1}{2})},\ (U_t)_{t\in(\frac{1}{2},1]} > (1-p)\,\mathbb{1}_{(\frac{1}{2},1]},\ \boldsymbol{X} = \mathrm{id}_{[0,1]}\Big) \\
&\quad + \mathrm{P}\Big((U_t)_{t\in[0,\frac{1}{2})} > p\,\mathbb{1}_{[0,\frac{1}{2})},\ (U_t)_{t\in(\frac{1}{2},1]} \le (1-p)\,\mathbb{1}_{(\frac{1}{2},1]},\ \boldsymbol{X} = \mathbb{1}_{[0,1]} - \mathrm{id}_{[0,1]}\Big) \\
&= \mathrm{P}\Big(\big(F_t^{-1}(U_t)\big)_{t\in[0,1]} = \mathrm{id}_{[0,1]},\ \boldsymbol{X} = \mathrm{id}_{[0,1]}\Big) \\
&\quad + \mathrm{P}\Big(\big(F_t^{-1}(U_t)\big)_{t\in[0,1]} = \mathbb{1}_{[0,1]} - \mathrm{id}_{[0,1]},\ \boldsymbol{X} = \mathbb{1}_{[0,1]} - \mathrm{id}_{[0,1]}\Big) \\
&= \mathrm{P}\Big(\big(F_t^{-1}(U_t)\big)_{t\in[0,1]} = \boldsymbol{X}\Big).
\end{aligned}
$$

Note that this proof of $\mathrm{P}\big((F_t^{-1}(U_t))_{t\in[0,1]} = \boldsymbol{X}\big) = 1$ would still hold if we would replace V in (3.3.2) with $V_t \sim \mathcal{U}[0,1]$. Nevertheless, even the space $\mathcal{E}[0,1]$ is not large enough, in the sense that there are stochastic processes in $C[0,1]$ that do not have a copula process in $\mathcal{E}[0,1]$:

Example 3.3.3 Define a stochastic process $\boldsymbol{Y} = (Y_t)_{t\in[0,1]}$ by $Y_1 = \frac{1}{2}$ and

$$
Y_t = \frac{1}{2} + (-1)^{n-1}\Big(B - \frac{1}{2}\Big)\Big(2n - 1 - (2n+1)t\Big) \quad \text{for} \quad t \in [t_n, t_{n+1})
$$

where $t_n = 1 - \frac{1}{n}$, $n \in \mathbb{N}$, and B is $\mathcal{B}(1,p)$-distributed with parameter $p \in (0,1)$. Thus $\boldsymbol{Y}$ is in $C[0,1]$ and we have

$$
Y_{t_n} = \frac{1}{2} + \frac{(-1)^{n-1}}{n}\Big(B - \frac{1}{2}\Big) \quad \text{as well as} \quad Y_{\frac{2n-1}{2n+1}} = \frac{1}{2}
$$

with probability one. The sample paths are illustrated in Figure 3.3.4. As this process is essentially a sequence of scaled versions of the process $\boldsymbol{X}$ from above, the same reasoning as before shows that any copula process of $\boldsymbol{Y}$ is not continuous at the points $\frac{2n-1}{2n+1}$, $n \in \mathbb{N}$, and $\boldsymbol{Y}$ has therefore no copula process in $\mathcal{E}[0,1]$.

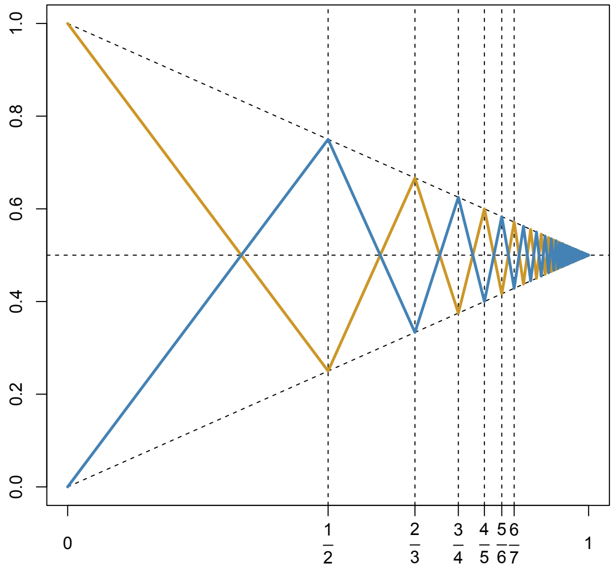

Figure 3.3.4 The both sample paths of the process $\boldsymbol{X}$ in Example 3.3.3.

Although the above considerations suggest that one of the main assumptions of Chapter 2, namely that the data emerged from a distribution with a continuous copula process, is rather restrictive, the following simple result hints how to overcome this disadvantage.

Lemma 3.3.5 *Let V be a non-negative random variable with a continuous distribution function. If $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ is a stochastic process in $C[0,1] \cap (-\infty,0)^{[0,1]}$ that is independent of V, then $V\boldsymbol{X}$ has a continuous copula process.*

Proof. We have obviously that $V\boldsymbol{X}$ is in $C[0,1]$. Moreover, $t \in [0,1]$ and $y < 0$ imply

$$\begin{aligned} \mathrm{P}(VX_t \le y) &= \int_{(-\infty,0)} \mathrm{P}\left(V \le \frac{y}{x}\right) (\mathrm{P} * X_t)(\mathrm{d}x) \\ &\to \int_{(-\infty,0)} \mathrm{P}\left(V \le \frac{y_0}{x}\right) (\mathrm{P} * X_t)(\mathrm{d}x) = \mathrm{P}(VX_t \le y_0) \quad \text{as} \quad y \to y_0-, \end{aligned}$$

i. e. the assertion follows from the dominated convergence theorem and Lemma 1.2.9. $\square$

Consider a stochastic process $\boldsymbol{X}$ in $C[0,1] \cap (-\infty,0)^{[0,1]}$ that has no continuous copula process. Lemma 3.3.5 shows in particular that $\boldsymbol{X}$ can be approximated by another process that does have a copula process in $C[0,1]$: For $n \in \mathbb{N}$ let V_n be a random variable with Lebesgue-density $g_n = n\,\mathbb{1}_{\left(1-\frac{1}{2n},1+\frac{1}{2n}\right)}$. We obtain

$$\frac{2n+1}{2n}\boldsymbol{X} < V_n\boldsymbol{X} < \frac{2n-1}{2n}\boldsymbol{X}$$

with probability one.

Copula Processes from Copula Random Vectors

We have seen that there are finite dimensional copulas which are not in a domain of attraction of an MSD. The question is whether the same is true for continuous copula processes. In fact, the both parametric families given in Lemma 3.2.1 and Lemma 3.2.7 can be extended to stochastic processes in $C[0,1]$ by a linear interpolation of the corresponding generator, which is another immediate consequence of Lemma 3.3.5.

Corollary 3.3.6 *Let $\widetilde{\boldsymbol{Z}} = (\widetilde{Z}_1, \dots, \widetilde{Z}_d)^\intercal$ be a d-variate generator with $d \geq 2$, and consider a grid $\{t_1, \dots, t_d\}$ such that $t_1 = 0$, $t_d = 1$, and $t_i < t_{i+1}$ for $i = 1, \dots, d-1$. Define a continuous generator process $\boldsymbol{Z} = (Z_t)_{t\in[0,1]}$ by*

$$Z_t := \begin{cases} \widetilde{Z}_i & \text{if } t = t_i,\ i = 1,\dots,d, \\ \frac{t_{i+1}-t}{t_{i+1}-t_i} Z_{t_i} + \frac{t-t_i}{t_{i+1}-t_i} Z_{t_{i+1}} & \text{if } t \in (t_i, t_{i+1}),\ i = 1,\dots,d-1. \end{cases} \tag{3.3.7}$$

Furthermore, let $h \in C[0,1] \cap (0,\infty)^{[0,1]}$ and choose a non-negative and continuously distributed random variable V that is independent of $\widetilde{\boldsymbol{Z}}$. Then the stochastic process

$$\boldsymbol{W} = \begin{cases} -V\frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}} & \text{if} \quad \mathrm{P}(\widetilde{\boldsymbol{Z}} > \boldsymbol{0}) = 1 \\ -V\min\left\{h, \frac{\mathbb{1}_{[0,1]}}{\boldsymbol{Z}}\right\} & \text{if} \quad \mathrm{P}(\widetilde{\boldsymbol{Z}} > \boldsymbol{0}) < 1 \end{cases} \tag{3.3.8}$$

is in $C[0,1]$ and has a continuous copula process.

The interpolation method (3.3.7) was introduced by Hofmann (2012) and shows that any finite dimensional MSD (or GPD) can be extended to a functional MSD (or GPD), where the original distribution is preserved as a finite dimensional margin of the functional version. Due to Lemma 3.3.5 we now also know that the copula of the original distribution is by (3.3.7) extended to a copula process in $C[0,1]$.

The approach discussed in Corollary 3.3.6 extends the finite dimensional copula of $\widetilde{\boldsymbol{W}} := -V\,(\boldsymbol{1}/\widetilde{\boldsymbol{Z}})$ to a functional one, where the lower end of $\widetilde{\boldsymbol{W}}$ is cut off as in (3.3.8) if $\mathrm{P}(\widetilde{\boldsymbol{Z}} > \boldsymbol{0}) < 1$. Since neither the finite dimensional nor the functional copula has to be known explicitly, this interpolation method might be considered as an *indirect* one. However, if a random vector $\widetilde{\boldsymbol{U}}$ follows a finite dimensional EVC, then $\widetilde{\boldsymbol{U}}$ can be interpolated *directly* by means of generalized max-linear models, which were defined by Falk et al. (2014) as an extension of the max-linear models introduced by Wang and Stoev (2011). We also refer to Dombry et al. (2015) for a recent account of simulation techniques for MSPs.

Lemma 3.3.9 *Consider an integer $d \geq 2$, a random vector $\widetilde{\boldsymbol{U}} = (\widetilde{U}_1, \dots, \widetilde{U}_d)^\intercal$ which follows a d-variate EVC C, and a grid $\{t_1, \dots, t_d\}$ such that $t_1 = 0$, $t_d = 1$, and $t_i < t_{i+1}$ for $i = 1, \dots, d-1$. Then*

$$U_t := \begin{cases} \widetilde{U}_i & \text{if } t = t_i,\ i = 1, \dots, d, \\ \max\left\{ \widetilde{U}_i^{\left\|\left(1, \frac{t-t_i}{t_{i+1}-t}\right)^\intercal\right\|_{D,i}}, \widetilde{U}_{i+1}^{\left\|\left(\frac{t_{i+1}-t}{t-t_i}, 1\right)^\intercal\right\|_{D,i}} \right\} & \text{if } t \in (t_i, t_{i+1}),\ i = 1, \dots, d-1, \end{cases}$$

defines a copula process $\boldsymbol{U} = (U_t)_{t \in [0,1]}$ in $\boldsymbol{C}[0,1]$ which follows a functional EVC, where

$$\left\|(y,z)^\intercal\right\|_{D,i} := -\log\Big[C\Big(\Big(\exp(-|y|)\, \mathbb{1}_{\{i\}}(\ell) + \exp(-|z|)\, \mathbb{1}_{\{i+1\}}(\ell) + \mathbb{1}_{\{1,\dots,d\}\setminus\{i,i+1\}}(\ell) \Big)_{\ell=1}^d\Big)\Big]$$

for all $y, z \in \mathbb{R}$ and $i \in \{1, \dots, d-1\}$.

Proof. According to (1.1.31) the D-norm $\|\cdot\|_D$ of C is given by

$$\|\boldsymbol{x}\|_D = -\log\Big(C\big(\exp(-|\boldsymbol{x}|)\big)\Big) \quad \text{for} \quad \boldsymbol{x} \in \mathbb{R}^d,$$

which yields in particular

$$\left\|(y,z)^\intercal\right\|_{D,i} = \left\|\Big(y\, \mathbb{1}_{\{i\}}(\ell) + z\, \mathbb{1}_{\{i+1\}}(\ell)\Big)_{\ell=1}^d\right\|_D \quad \text{for all } y, z \in \mathbb{R} \text{ and } i \in \{1, \dots, d-1\}.$$

Since $\log(\widetilde{\boldsymbol{U}})$ follows the standard MSD with D-norm $\|\cdot\|_D$, Falk et al. (2014, Corollary 3.2) proves that

$$\eta_t := \begin{cases} \log(\widetilde{U}_i) & \text{if } t = t_i,\ i = 1, \dots, d, \\ \left\|(t_{i+1}-t, t-t_i)^\intercal\right\|_{D,i} \max\left\{ \frac{\log(\widetilde{U}_i)}{t_{i+1}-t}, \frac{\log(\widetilde{U}_{i+1})}{t-t_i} \right\} & \text{if } t \in (t_i, t_{i+1}),\ i = 1, \dots, d-1, \end{cases}$$

defines a standard MSP $\boldsymbol{\eta} = (\eta_t)_{t \in [0,1]}$ in $\boldsymbol{C}[0,1]$. Now the assertion follows from the identity $\boldsymbol{U} = \exp(\boldsymbol{\eta})$. $\square$

The D-norms $\|\cdot\|_D$ and $\|\cdot\|_{D,*}$ of $\widetilde{\boldsymbol{U}}$ and $\boldsymbol{U}$, respectively, satisfy by construction the equation

$$\|\boldsymbol{x}\|_D = \left\| \sum_{i=1}^d x_i \, \mathbb{1}_{\{t_i\}} \right\|_{D,*} \qquad \text{for} \quad \boldsymbol{x} = (x_1, \dots, x_d)^\intercal \in \mathbb{R}^d$$

and Falk et al. (2014, Corollary 3.2) shows that $\|\cdot\|_{D,*}$ is generated by $\boldsymbol{Z} = (Z_t)_{t\in[0,1]}$ with

$$Z_t := \begin{cases} \widetilde{Z}_i & \text{if } t = t_i,\ i = 1, \dots, d, \\ \max\left\{ \left\| \left(1, \frac{t-t_i}{t_{i+1}-t}\right)^\intercal \right\|_{D,i}^{-1} \widetilde{Z}_i,\ \left\| \left(\frac{t_{i+1}-t}{t-t_i}, 1\right)^\intercal \right\|_{D,i}^{-1} \widetilde{Z}_{i+1} \right\} \\ & \text{if } t \in (t_i, t_{i+1}),\ i = 1, \dots, d-1, \end{cases}$$

where $\widetilde{\boldsymbol{Z}} = (\widetilde{Z}_1, \dots, \widetilde{Z}_d)^\intercal$ is a generator of $\|\cdot\|_D$. Falk et al. (2014, Corollary 4.2) implies moreover that

$$V_t := \begin{cases} \widetilde{V}_i & \text{if } t = t_i,\ i = 1, \dots, d, \\ \max\left\{ \left\| \left(1, \frac{t-t_i}{t_{i+1}-t}\right)^\intercal \right\|_{D,i} \widetilde{V}_i,\ \left\| \left(\frac{t_{i+1}-t}{t-t_i}, 1\right)^\intercal \right\|_{D,i} \widetilde{V}_{i+1} \right\} \\ & \text{if } t \in (t_i, t_{i+1}),\ i = 1, \dots, d-1, \end{cases}$$

defines a stochastic process $\boldsymbol{V} = (V_t)_{t\in[0,1]}$ in $\boldsymbol{C}[0,1]$ which follows a standard GPD with D-norm $\|\cdot\|_{D,*}$ if the distribution of the random vector $\left(\widetilde{V}_1, \dots, \widetilde{V}_d\right)^\intercal$ is a standard GPD with D-norm $\|\cdot\|_D$.

For the sake of completeness, we also note that any copula random vector may be interpolated directly: Consider a random vector $(U_1, \dots, U_d)^\intercal$ that is distributed according to some d-variate copula, $d \geq 2$. The procedure consists of two steps:

(i) Interpolate the values $U_1, \dots, U_d$ using a modified version of (3.3.7), see below. The result is a stochastic process $\widetilde{\boldsymbol{U}} = (\widetilde{U}_t)_{t\in[0,1]}$ in $\boldsymbol{C}[0,1]$.

(ii) For each $t \in [0,1]$, apply the distribution function of $\widetilde{U}_t$ to $\widetilde{U}_t$ and obtain a continuous copula process $\overline{\boldsymbol{U}} = (\overline{U}_t)_{t\in[0,1]}$, satisfying $\overline{U}_{\frac{i-1}{d-1}} = U_i$ for each $i \in \{1, \dots, d\}$.

To assure that the result of the second step is a continuous copula process, we modify the interpolation method in (3.3.7) such that the distribution function of $\widetilde{U}_t$ is continuous for all $t \in [0,1]$. Therefore, take $d-1$ random variables $W_1, \dots, W_{d-1} \sim \boldsymbol{\mathcal{U}}[0,1]$ such that W_i is independent of U_i and U_{i+1}, $i \in \{1, \dots, d-1\}$, and define

$$\widetilde{U}_t := \begin{cases} U_i & \text{if } t = \frac{i-1}{d-1},\ i = 1, \dots, d, \\ W_i & \text{if } t = \frac{i-1}{d-1} + \frac{1}{2(d-1)},\ i = 1, \dots, d-1, \\ \text{linearly interpolated} & \text{elsewhere.} \end{cases}$$

This means we apply the interpolation method in (3.3.7) to the $(2d-1)$-dimensional random vector $(U_1, W_1, U_2, W_2, U_3, \ldots, W_{d-1}, U_d)^\mathsf{T}$ and the grid $\{\frac{\ell-1}{2(d-1)} \mid \ell \in \{1, \ldots, 2d-1\}\}$. If there is $i_0 \in \{1, \ldots, d-1\}$ such that the random variable $(1-s)\,U_{i_0} + s\,U_{i_0+1}$ has a continuous distribution function for all $s \in (0,1)$, then W_{i_0} can be dropped and U_{i_0}, U_{i_0+1} can be interpolated directly. This is, e. g., the case for U_1 and U_2 in Figure 3.3.10 which are independent of one another. However, W_2 in the same figure *cannot* be dropped since otherwise $\widetilde{\boldsymbol{U}}$ would have no continuous copula process, cf. Example 3.3.3.

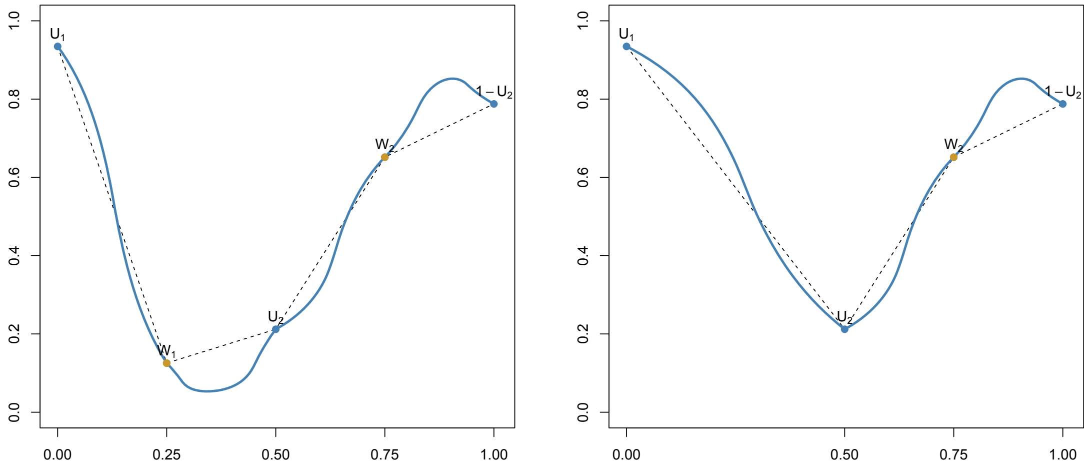

Figure 3.3.10 Two interpolations of the random vector $(U_1, U_2, 1-U_2)^\mathsf{T}$ with independent random variables $U_1, U_2 \sim \mathcal{U}[0,1]$.

Figure 3.3.10 was obtained by considering the following result, which can be seen by elementary calculations:

Lemma 3.3.11 *If $U_1, U_2 \sim \mathcal{U}[0,1]$ are independent, one gets*

$$\mathrm{P}\big((1-s)\,U_1 + s\,U_2 \le x\big) = \begin{cases} \frac{x^2}{2s(1-s)}, & x \in \big(0, \min\{s, 1-s\}\big) \\ \frac{2x-\min\{s,1-s\}}{2\max\{s,1-s\}}, & x \in \big[\min\{s,1-s\}, \max\{s,1-s\}\big) \\ 1 - \frac{(1-x)^2}{2s(1-s)}, & x \in \big[\max\{s,1-s\}, 1\big) \end{cases}$$

for all $s \in (0,1)$.

3.4 Selection of the Parameters

The preceding sections provided (finite dimensional) copulas that do or do not satisfy Hypothesis 2.1.8, and considered how some of these examples can be extended to continuous copula processes. Before we apply the tests of Chapter 2 to random samples of these copulas, which will be the content of Chapter 4, our aim is now to review the technical requirements; cf. Theorem 2.2.10, Theorem 2.2.12 and Theorem 2.2.21 for finite dimensional data, Theorem 2.3.3 and Theorem 2.3.11 for functional data, and Theorem 2.4.7 and Theorem 2.4.9 for functional data that are observed at a finite set of points.

Recall that the test statistics under consideration depend on several sequences, which are required to have certain asymptotic properties as the sample size grows to infinity. For a finite sample size, however, Chapter 4 will show that the test results are highly sensitive to a proper choice of the corresponding elements of these sequences. These elements will be referred to as the parameters of the test statistics. In order to obtain a first impression of how to obtain reasonable values for the parameters, we start with an exploratory approach for copula data, which will motivate the theoretical considerations that follow afterwards.

An Exploratory Approach for Copula Data

For convenience, we assume for a moment that our data consist of independent copies $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$ of a random element $\boldsymbol{U}$ that emerged from a (finite dimensional or functional) copula. If we want to test whether this copula is in the δ-neighborhood of a GPC, the test statistics depend on the parameters k and c_n, where $k \geq 2$ is an integer and $c_n \in (0,1)$ has the asymptotic properties $c_n \to 0$, $nc_n \to \infty$, and $nc_n^{1+2\delta} \to 0$ as $n \to \infty$. We assume $\delta > 0$ to be given; e. g., if we observe finite dimensional data and want to check whether the Gumbel-Hougaard family in Example 3.1.9 is a candidate for modeling the copula of the data, we would choose $\delta = 1$. Moreover, since parts (i) and (ii) of Lemma 2.1.6 show that δ-neighborhoods are nested, we assume $\delta < \infty$.

A graphically based approach for finding reasonable values for the parameters is as follows: In a preliminary step, we consider an estimator of the generator constant which depends on c_n but not on k. Based on its asymptotic properties, we obtain a range of suitable values for c_n, which is used in step two for choosing c_n and, afterwards, k. The last step is to check the goodness of these values by considering an estimator of the generator constant that depends on c_n *and* k. Precisely, we proceed as follows:

(i) Plot the function $\gamma_*(c) := \frac{1}{nc} N_{\boldsymbol{U}}^{(n)}(c)$ for $c \in (0,1]$ and recall that $N_{\boldsymbol{U}}^{(n)}(c)$ is the number of observations which do not belong to the interval $\big[\mathbf{0}, (1-c)\mathbf{1}\big]$ or $\big[\mathbb{0}_{[0,1]}, (1-c)\, \mathbb{1}_{[0,1]}\big]$, respectively. If Hypothesis 2.1.8 is true and c_n is chosen properly, then $\gamma_*(\frac{c_n}{1}),\dots,\gamma_*(\frac{c_n}{k})$ estimate the generator constant of the underlying GPC consistently as $n\to\infty$; this is due to Lemma 2.2.7, cf. Corollary 2.2.9 and Lemma 2.4.5. We expect therefore that there is an interval $I\subset(0,1]$ with the both properties that the range of the restriction $\gamma_*|_I$ is a subset of $[1,d]$ or $[1,\infty)$ — depending on whether the data are points in $\mathbb{R}^d$ or functions in $\boldsymbol{C}[0,1]$ —, and that γ_* is constant on I, apart from random fluctuations. Moreover, we can compute asymptotic confidence intervals for the generator constant based on the asymptotic normality

$$\left(\frac{nc_n}{\gamma_*(c_n)}\right)^{\frac{1}{2}}\left(\gamma_*(c_n)-\|\mathbf{1}\|_D\right)\xrightarrow{D}\boldsymbol{N}(0,1)\quad\text{as}\quad n\to\infty,\tag{3.4.1}$$

cf. Corollary 2.2.9 and Lemma 2.4.5, where $\mathbf{1}$ has to be interpreted as $\mathbb{1}_{[0,1]}$ for functional data.

(ii) Given that Hypothesis 2.1.8 is true, c_n has, on the one hand, to be chosen small enough such that the threshold level $1-c_n$ is sufficiently close to one in order to detect the δ-neighborhood, cf. Definition 2.1.5. On the other hand, c_n must be large enough in order to guarantee that there are sufficiently many observations in the extremal region $[\mathbf{0},\mathbf{1}]\setminus[\mathbf{0},(1-c_n)\mathbf{1}]$ or $[\mathbb{0}_{[0,1]},\mathbb{1}_{[0,1]}]\setminus[\mathbb{0}_{[0,1]},(1-c_n)\,\mathbb{1}_{[0,1]}]$ such that the asymptotic normality in Corollary 2.2.9 and Lemma 2.4.5 is justified. Altogether, the selection of c_n is a typical tradeoff situation, similar to the problem of choosing a threshold for the adaption of a GPD to univariate data, see e. g. Embrechts et al. (1997, Section 6.5). If we consider the interval I derived in (i), a reasonable strategy is to choose c_n from I such that c_n is close to $\sup I$. Similarly, the integer $k\ge 2$ should be chosen such that $\frac{c_n}{k}\in I$ and such that there are sufficiently many exceedances above the threshold level $1-\frac{c_n}{k}$, where it is reasonable to put $k=2$.

(iii) If Hypothesis 2.1.8 is true and the parameters c_n and k are derived as in (ii), then the differences

$$\frac{(j+1)\,N_{\boldsymbol{U}}^{(n)}(\frac{c_n}{j+1})}{j\,N_{\boldsymbol{U}}^{(n)}(\frac{c_n}{j})}-1,\quad j=1,\dots,k-1,$$

should be close to zero, cf. Lemma 2.2.14. Furthermore, the same reasoning as in (i) shows that the function $\gamma(c):=\frac{1}{k}\sum_{j=1}^{k}\frac{j}{nc}\,N_{\boldsymbol{U}}^{(n)}(\frac{c}{j})=\frac{1}{k}\sum_{j=1}^{k}\gamma_*(\frac{c}{j})$, $c\in(0,1]$, should be almost constant on some interval $J\subset(0,1]$, where we expect $c_n\in J$. However, $c_n>\sup J$ would indicate that step (ii) overestimated this parameter, which suggests to repeat steps (ii) and (iii) with $c_n:=\sup J$. Moreover, the

k-dimensional asymptotic normality as in Corollary 2.2.9 and Lemma 2.4.5 shows

$$
\begin{aligned}
(3.4.2)\qquad &\left(\frac{6knc_n}{(k+1)(2k+1)\,\gamma(c_n)}\right)^{\frac{1}{2}}\left(\gamma(c_n)-\|\mathbf{1}\|_D\right)\\
&=\left(\frac{nc_n}{\frac{k(k+1)(2k+1)}{6}\cdot\frac{1}{k}\sum_{j=1}^{k}\frac{j}{nc_n}N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right)}\right)^{\frac{1}{2}}\sum_{j=1}^{k}\left(\frac{j}{nc_n}N_{\boldsymbol{U}}^{(n)}\left(\frac{c_n}{j}\right)-\|\mathbf{1}\|_D\right)\\
&\xrightarrow{D}\boldsymbol{N}(0,1)\quad\text{as}\quad n\to\infty,
\end{aligned}
$$

which can be used to obtain another asymptotic confidence interval for the generator constant. Again, $\mathbf{1}$ has to be interpreted as $\mathbb{1}_{[0,1]}$ for functional observations.

Note that the asymptotic confidence intervals in (i) and (iii) have to be interpreted with care, since they are only reliable if Hypothesis 2.1.8 is true, the sample size n is large, and c_n is chosen properly. If all these conditions are true, then we obtain the $k+1$ consistent estimates $\gamma_*\left(\frac{c_n}{1}\right),\dots,\gamma_*\left(\frac{c_n}{k}\right)$, and $\gamma(c_n)$ for the underlying generator constant.

An Analytical Approach for Copula Data

Now we turn over to a technical analysis of the exploratory procedure from above. If the distribution function C of our copula data $\boldsymbol{U}^{(1)},\dots,\boldsymbol{U}^{(n)}$ is in the domain of attraction of a standard MSD with corresponding D-norm $\|\cdot\|_D$, the expected number of exceedances above the threshold $(1-c)\mathbf{1}$ has the expansion

$$
(3.4.3)\qquad \mathrm{E}\left(N_{\boldsymbol{U}}^{(n)}(c_n)\right)=nc_n\cdot\frac{1-C((1-c_n)\mathbf{1})}{c_n}\sim nc_n\,\|\mathbf{1}\|_D\quad\text{as}\quad n\to\infty
$$

since $N_{\boldsymbol{U}}^{(n)}(c_n)$ is $\mathcal{B}(n,p_n)$-distributed with $p_n=1-C((1-c_n)\mathbf{1})$, cf. Definition 2.2.3 and Definition 2.3.1. For convenience, we temporarily restrict ourselves to multivariate copula data.

Due to (3.4.3), the conditions $c_n\to 0$, $nc_n\to\infty$, and $nc_n^{1+2\delta}\to 0$ as $n\to\infty$ specify at which rate the expected number of exceedances approaches infinity when the sample size increases. As outlined previously, we aim at testing the data for a *certain* δ-neighborhood of a GPC with D-norm $\|\cdot\|_D$, i. e. we assume that $\delta>0$ is given. In particular, if C is also in the δ_0-neighborhood of a GPC for some $\delta_0>\delta$, then we obtain

$$
\begin{aligned}
1-C\left((1-c_n)\mathbf{1}\right)&=c_n\,\|\mathbf{1}\|_D+c_n^{1+\delta}\cdot c_n^{\delta_0-\delta}\,\frac{1-C((1-c_n)\mathbf{1})-c_n\,\|\mathbf{1}\|_D}{c_n^{1+\delta_0}}\\
&=c_n\,\|\mathbf{1}\|_D+o\left(c_n^{1+\delta}\right)\quad\text{as}\quad n\to\infty,
\end{aligned}
$$

cf. Lemma 2.1.6 (iv). Thus, the same arguments that proved Theorem 2.2.10 and Theorem 2.2.12 also show:

Corollary 3.4.4 *Let $\delta > 0$ and $k \in \mathbb{N}$, $k \geq 2$. If a (finite dimensional) copula C is in the δ_0-neighborhood of a GPC for some $\delta_0 > \delta$, and $(c_n)_{n\in\mathbb{N}}$ is a sequence in $(0,1)$ satisfying $c_n \to 0$, $nc_n \to \infty$, and $nc_n^{1+2\delta} \to s \geq 0$, then the conclusions of Theorem 2.2.10 and Theorem 2.2.12 remain valid.*

Now we return to the task of specifying c_n reasonably for a finite sample size n. On the one hand, (3.4.3) and Corollary 2.2.9 show that the condition $nc_n \to \infty$ is crucial in order to assure that we observe sufficiently many data in the extremal regions, which leads to the desired normal approximation. On the other hand, Corollary 3.4.4 suggests that the condition $nc_n^{1+2\delta} \to 0$ is rather a mild one. Therefore, it is reasonable to choose c_n rather large, i. e. close to $n^{-\frac{1}{1+2\delta}}$; this corresponds to step (ii) in the exploratory approach, where we motivated to choose c_n close to the upper endpoint of the interval I.

The same reasoning as above leads to analogous versions of Corollary 3.4.4 for functional copula data and for functional copula data that are observed at finitely many points only. Thus, the simulations in Chapter 4 will choose the parameter c_n as follows:

Example 3.4.5 Consider the sequence given by $c_n := (n\log(n))^{-\frac{1}{1+2\delta}}$ for $n \geq 2$. Then we obviously have $c_n \to 0$, $nc_n = \big(\frac{n^{2\delta}}{\log(n)}\big)^{\frac{1}{1+2\delta}} \to \infty$, and $nc_n^{1+2\delta} = \frac{1}{\log(n)} \to 0$ as $n \to \infty$.

For the remaining parameter k, the same arguments as in the exploratory approach apply. Since there will typically be very few observations in the extremal regions, we usually choose $k = 2$.

An Approach for More General Data

Now we consider that the observed data $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ are independent copies of a random element $\boldsymbol{X}$ with continuous but unknown margins. If $\boldsymbol{X}$ is a stochastic process in $C[0,1]$, we assume moreover that all margins coincide, cf. Section 2.3. In this more general framework, the whole sample is used for the estimation of the margins, whereas the computation of the test statistics is based on a subsample of size m_n. This led to the conditions $\frac{m_n}{n}\log(m_n) \to 0$, $c_n \to 0$, $m_n c_n \to \infty$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$.

Since we now have to specify the three parameters k, c_n, and m_n, applying an adaption of the exploratory approach from above would be a very time consuming task: The function γ_* in step (i) would have to be replaced with

$$\hat{\gamma}_*^{(m)}(c) := \frac{1}{mc}\,\hat{N}_{\boldsymbol{X}}^{(m,n)}(c) \quad \text{for } c \in (0,1] \text{ and } m \in \{1,\dots,n\}; \tag{3.4.6}$$

i. e., instead of plotting a single function, we would need to consider a whole family of functions, indexed by m. Provided that Hypothesis 2.1.8 is true, we would at first have to figure out an m such that $\hat{\gamma}_*^{(m)}$ is — apart from random fluctuations — constant on some interval $I_m \subset (0,1]$, and then choose c_n as in step (ii). Despite of the high computational effort, analyzing a large number of function plots in order to find a suitable tuple (m_n, c_n) would be a time consuming and subjective judgement.

Instead, we will extend the analytical approach from above and use the graphical tools as a visual goodness of fit check. The following result will be quite useful in order to link the selection of the both parameters m_n and c_n.

Lemma 3.4.7

(i) Let $(\beta_n)_{n\in\mathbb{N}}$ be a sequence in $(0,1)$ with $\beta_n \to 0$ and $n\beta_n \to \infty$ as $n \to \infty$. Then the sequence $(m_n)_{n\in\mathbb{N}}$ with $m_n := \min\Big\{\Big\lceil \frac{n\beta_n}{\log(n\beta_n)} \Big\rceil, n\Big\}$ satisfies $\frac{m_n}{n}\log(m_n) \sim \beta_n$ as $n \to \infty$.

(ii) Let $(m_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{N}$ such that $m_n \to \infty$ and $\frac{m_n}{n}\log(m_n) \to 0$ as $n \to \infty$. If a sequence $(\beta_n)_{n\in\mathbb{N}}$ satisfies $\beta_n \sim \frac{m_n}{n}\log(m_n)$ as $n \to \infty$, we obtain $m_n \sim \frac{n\beta_n}{\log(n\beta_n)}$ as $n \to \infty$.

Proof. In order to prove the first assertion, note that $m_n = \Big\lceil \frac{n\beta_n}{\log(n\beta_n)} \Big\rceil$ whenever n is sufficiently large. Thus, $n\beta_n \to \infty$ yields

$$\frac{\beta_n}{\log(n\beta_n)} \le \frac{m_n}{n} < \frac{\beta_n}{\log(n\beta_n)} + \frac{1}{n} = \frac{\beta_n}{\log(n\beta_n)}(1+o(1)) \quad \text{as} \quad n \to \infty$$

and

$$\begin{aligned}\frac{m_n}{n}\log(m_n) &= \frac{\beta_n}{\log(n\beta_n)}(1+o(1))\left[\log\left(\frac{n\beta_n}{\log(n\beta_n)}\right) + \log\left(\frac{\log(n\beta_n)}{n\beta_n}\cdot m_n\right)\right] \\ &\sim \beta_n \quad \text{as} \quad n \to \infty,\end{aligned}$$

i. e. $\frac{m_n}{n}\log(m_n)$ approaches zero at the same rate as β_n does. Conversely, $m_n \to \infty$ and $\beta_n \sim \frac{m_n}{n}\log(m_n) \to 0$ as $n \to \infty$ imply

$$m_n \sim \frac{n\beta_n}{\log(n\beta_n)} \cdot \frac{\log(n\beta_n)}{\log(m_n)} = \frac{n\beta_n}{\log(n\beta_n)} \cdot \frac{\log(m_n\log(m_n)) + o(1)}{\log(m_n)} \sim \frac{n\beta_n}{\log(n\beta_n)}$$

as $n \to \infty$, cf. de Bruijn (1981, Section 2.4) and Corless et al. (1996, Section 4). □

According to Lemma 3.4.7, the rate at which m_n tends to infinity is driven by the rate at which $\frac{m_n}{n}\log(m_n)$ approaches zero, and vice versa. In particular, if $(\beta_n)_{n\in\mathbb{N}}$ and

$(m_n)_{n\in\mathbb{N}}$ are given as in Lemma 3.4.7 (i), then

$$c_n := (n\beta_n)^{-\frac{1}{1+2\delta}} \sim (m_n \log(m_n))^{-\frac{1}{1+2\delta}} \to 0 \quad \text{as} \quad n \to \infty,$$

which corresponds to Example 3.4.5. For a given sample size n, the task of finding a suitable tuple (m_n, c_n) of parameters can, thus, be reduced to specifying a single value β_n. Motivated by (3.4.3), we aim at choosing the new parameter β_n such that $m_n c_n$ is large. Since

$$n\beta_n \to \infty \quad \text{and} \quad m_n c_n \sim \frac{(n\beta_n)^{\frac{2\delta}{1+2\delta}}}{\log(n\beta_n)} \quad \text{as} \quad n \to \infty,$$

it is reasonable to require $\beta_n > \frac{1}{n}\exp(1+\frac{1}{2\delta})$; note that the function $(1,\infty) \ni x \mapsto \frac{x^\alpha}{\log(x)}$ is strictly increasing for $x \geq \exp(\frac{1}{\alpha})$ if $\alpha > 0$. In order to observe a sufficiently large number of data in the extremal regions, even for relatively small sample sizes, we choose a sequence $(\beta_n)_{n\in\mathbb{N}}$ with a very low rate of convergence towards zero, cf. Example 3.4.9. Due to the representation $m_n = f(c_n)$, where the function f is defined by

$$f(c) := \min\left\{\left\lceil \frac{-1}{(1+2\delta)\, c^{1+2\delta}\log(c)} \right\rceil, n\right\} \quad \text{for} \quad c \in (0,1], \tag{3.4.8}$$

the graphical tools of the exploratory approach will carry over to our current setting.

Example 3.4.9 For $\beta_n := \left[\log(\log(n))\right]^{-1}\exp\left(1+\frac{1}{2\delta}\right)$, $n \geq 3$, we obtain

$$c_n = \left(\frac{\log(\log(n))}{n}\right)^{\frac{1}{1+2\delta}} \exp\left(-\frac{1}{2\delta}\right) \quad \text{and} \quad m_n = f(c_n).$$

Since $nc_n^{1+2\delta} = \beta_n^{-1}$, we have the asymptotic properties $\frac{m_n}{n}\log(m_n) \sim \left(nc_n^{1+2\delta}\right)^{-1} \to 0$, $m_n c_n \sim \left[-(1+2\delta)\, c_n^{2\delta}\log(c_n)\right]^{-1} \to \infty$, and $m_n c_n^{1+2\delta} \sim \left[-(1+2\delta)\log(c_n)\right]^{-1} \to 0$ as $n \to \infty$.

The goodness of the linkage of m_n and c_n via f may be checked visually by plotting the function $\tilde{\gamma}_*(c) := \frac{1}{c\,f(c)}\,\hat{N}_{\boldsymbol{X}}^{(f(c),n)}(c)$, $c \in (0,1]$. If Hypothesis 2.1.8 is true, then $\tilde{\gamma}_*$ should be — apart from random fluctuations — constant on some interval $\tilde{I} \subset (0,1]$.

We obtain an approach for deriving reasonable parameters which is similar to the case where copula data are observed:

(i) For a finite sample size n, compute c_n from some sequence $(\beta_n)_{n\in\mathbb{N}}$ as above. Put $m_n := f(c_n)$ and plot the function $\hat{\gamma}_*^{(m_n)}$, cf. (3.4.6) and (3.4.8). Provided that Hypothesis 2.1.8 is valid, we should observe an almost linear graph on some interval

$I \subset (0,1]$. Moreover, we expect $c_n \in I \cap \tilde{I}$ and we may compute asymptotic confidence intervals for the underlying generator constant from

$$\left(\frac{m_n c_n}{\hat{\gamma}_*^{(m_n)}(c_n)}\right)^{\frac{1}{2}} \left(\hat{\gamma}_*^{(m_n)}(c_n) - \|\mathbf{1}\|_D\right) \xrightarrow{D} \boldsymbol{\mathcal{N}}(0,1) \quad \text{as} \quad n \to \infty, \tag{3.4.10}$$

where $\mathbf{1}$ has to be interpreted as $\mathbb{1}_{[0,1]}$ for functional data, cf. (3.4.6), Lemma 2.2.19, Lemma 2.3.10, and Lemma 2.4.8. The remaining parameter k can then be chosen such that both of the following conditions are met: On the one hand, we require $\frac{c_n}{k} \in I$. On the other hand, there must be sufficiently many observations exceeding the threshold corresponding to the level $1 - \frac{c_n}{k}$. Typically, we put $k := 2$. In the special case that we observe d-dimensional data which emerged from a whole process, cf. Lemma 2.4.8, we also require $\frac{d^2}{m_n c_n}$ to be sufficiently small. Recall that this condition is obsolete if the underlying sample continuous processes emerge from a copula, cf. Lemma 2.4.5.

(ii) As before, the differences

$$\frac{(j+1)\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\left(\frac{c_n}{j+1}\right)}{j\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\left(\frac{c_n}{j}\right)} - 1, \quad j = 1, \dots, k-1.$$

should be close to zero if Hypothesis 2.1.8 is true and the parameters c_n, m_n, and k are chosen well. Moreover, the function $\hat{\gamma}^{(m_n)}(c) := \frac{1}{k}\sum_{j=1}^{k} \frac{j}{m_n c}\,\hat{N}_{\boldsymbol{X}}^{(m_n,n)}\left(\frac{c}{j}\right) = \frac{1}{k}\sum_{j=1}^{k} \hat{\gamma}_*^{(m_n)}\left(\frac{c}{j}\right)$, $c \in (0,1]$, should be approximately constant on some interval $J_{m_n} \subset (0,1]$. If $c_n \in J_{m_n}$, the parameters appear to be chosen well. If however $c_n > \sup J_{m_n}$, then return to step (i) with $c_n := \sup J_{m_n}$. Due to Lemma 2.2.19, Lemma 2.3.10, and Lemma 2.4.8, we obtain an asymptotic confidence interval for the generator constant from

$$\left(\frac{6k m_n c_n}{(k+1)(2k+1)\,\hat{\gamma}^{(m_n)}(c_n)}\right)^{\frac{1}{2}} \left(\hat{\gamma}^{(m_n)}(c_n) - \|\mathbf{1}\|_D\right) \xrightarrow{D} \boldsymbol{\mathcal{N}}(0,1) \tag{3.4.11}$$

as $n \to \infty$, where $\mathbf{1}$ has to be interpreted as $\mathbb{1}_{[0,1]}$ for functional observations.

With the above reasoning in mind, we move on to Chapter 4. By using the results of Section 3.1 and Section 3.2, we will generate random samples from copulas, for which we *know* whether Hypothesis 2.1.8 is true. Then we will apply our strategies from above in order to compare the tests derived in Chapter 2 with one another.

4 Simulations

This final chapter of the thesis combines the results of Chapter 2 and Chapter 3 and applies them to simulated data, i. e. we *know* whether the underlying copula is in a δ-neighborhood of a GPC, cf. Hypothesis 2.1.8. Thus, we are able to check the performance of the tests derived in Chapter 2 together with the strategies for their application, cf. Section 3.4. The simulation of the data will be based on Section 3.2, where the methods described in Section 3.3 allow to generate even *functional* observations. However, the interpolation techniques of Corollary 3.3.6 and Lemma 3.3.9 have the property that the interpolated process attains its maximum at one of its interpolation points, cf. Falk et al. (2014, Lemmas 3.3 and 4.4). Due to the definition of the sojourn time, the test results for the finite dimensional observations and those for the interpolating processes are identical, apart from errors resulting from numerical integration. Since functional simulations are by far more time intensive, we restrict ourselves to multivariate data in what follows. Recall that Section 2.4 has shown that functional test results can be reasonably approximated by the test results based on finite dimensional projections, provided that the number of observation points tends to infinity at a proper rate and the observation grid gets arbitrarily fine as the sample size grows to infinity. As before, Section 4.1 concentrates on copula data, whereas Section 4.2 will assume that the margins of the data are unknown.

All simulations were performed using the software `R 3.2.1`, cf. R Core Team (2015), with its automatically loaded base packages `base`, `datasets`, `graphics`, `grDevices`, `methods`, `stats`, and `utils`. Moreover, the packages `copula` by Hofert et al. (2015), `CompQuadForm` by Duchesne and Lafaye de Micheaux (2010) and Lafaye de Micheaux (2013), and `tcltk` by R Core Team (2015) were loaded via `R`'s `requireNamespace` function. They were used to generate random deviates from standard copulas, to compute p-values from a weighted χ^2-distribution as in Theorem 2.2.10, and to display the status of the current simulation, respectively. Table 4.0.1 summarizes the loaded packages — including those that were invoked indirectly due to package dependencies, excluding the automatically loaded base packages named above — together with their version numbers.

Due to hardware restrictions and since some simulations were very time consuming, the following results were performed with a sample size of $n = 200$, where each simulation was done 50 times.[1] This allows, on the one hand, to compare the results of different

`ADGofTest 0.3`	`lattice 0.20-33`	`stats4 3.2.1`
`CompQuadForm 1.4.1`	`Matrix 1.2-2`	`tcltk 3.2.1`
`copula 0.999-13`	`mvtnorm 1.0-2`	`tools 3.2.1`
`grid 3.2.1`	`pspline 1.0-17`	
`gsl 1.9-10`	`stabledist 0.7-0`	

Table 4.0.1 List of loaded `R` packages.

simulations with one another and, on the other hand, to compute the relative frequency of repetitions that reject the null hypothesis. Moreover, a sample size of $n = 200$ appears to be more realistic in applications than, e. g., $n = 10\,000$. Nevertheless, whenever time consumption was sufficiently low, the data of all 50 repetitions were combined to a data set of size 10 000 in order to perform the tests for this combined data. The simulations presented below all have this property, such that we obtain an impression of how the sample size influences the test results.

4.1 Copula Data

We start with the case where our data consist of independent copies $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(n)}$ of a random element $\boldsymbol{U}$ which follows a (finite dimensional or functional) copula C. Table 4.1.1 summarizes the three different cases that were considered in Chapter 2 and refers to the corresponding results. As outlined previously, we focus on the finite dimensional case; see

Data	Test Statistics	Theoretical Results
Vectors in $\mathbb{R}^d$	$T_n, \mathcal{T}_n, \tau_n$	Theorem 2.2.10, Theorem 2.2.12
Functions in $\boldsymbol{C}[0,1]$	$T_n, \mathcal{T}_n, \tau_n$	Theorem 2.3.3
Functions observed at a grid	$T_n^*, \mathcal{T}_n^*, \tau_n^*$	Theorem 2.4.7

Table 4.1.1 Overview of the different test statistics for copula data.

(2.4.2), Lemma 2.4.5, and Theorem 2.4.7 for details on how the third case in Table 4.1.1 links the first two ones with one another.

Following the arguments in Section 3.4, we choose $c_n := (n \log(n))^{-\frac{1}{1+2\delta}}$ where $n \geq 2$ denotes the sample size and $\delta > 0$ specifies the kind of neighborhood in the null hypothesis,

[1] Actually, it has been very time demanding task to determine a proper sample size and a number of repetitions that the present hardware was able to deal with in a reasonable amount of time.

cf. Example 3.4.5. Note that, according to the parts (i) and (ii) of Lemma 2.1.6, the proposed test statistics T_n, $\mathcal{T}_n$, and τ_n actually test the null hypothesis

$H_{0,\delta}$: There is some $\delta_0 \in [\delta, \infty]$ such that the copula underlying the observed data is in the δ_0-neighborhood of a GPC,

cf. Corollary 3.4.4. Due to Lemma 3.1.1, the case $\delta = 1$ is of particular interest since the hypothesis $H_{0,1}$ includes the class of all EVCs and, of course, the class of all GPCs. Note that an EVC is, due to (1.1.22), tail equivalent with its corresponding GPC. Thus, it appears to be natural to require $\delta = 1$ for the simulations and to compare the results based on T_n, $\mathcal{T}_n$, and τ_n with those of the test proposed by Kojadinovic et al. (2011), which exploits (1.1.16) in order to test for an EVC. Compared with other tests for an EVC, see e. g. Ghoudi et al. (1998) and Ghorbal et al. (2009), the one by Kojadinovic et al. (2011) is not restricted to bivariate data.

GPCs and Related Copulas

We begin with the family of copulas introduced in Lemma 3.2.1, which is indexed with a parameter $\lambda \in \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$. Recall that for $\lambda = 0$, the corresponding copula is a GPC with D-norm given by

$$\|\boldsymbol{x}\|_D = \|\boldsymbol{x}\|_1 - \frac{|x_1|\,|x_2|}{\|\boldsymbol{x}\|_1} \quad \text{for} \quad \boldsymbol{x} = (x_1, x_2)^\intercal \in \mathbb{R}^2 \setminus \{\mathbf{0}\},$$

whereas $\lambda \neq 0$ implies that the copula is not in the domain of attraction of an EVC. Thus, $H_{0,1}$ is satisfied if and only if $\lambda = 0$, in which case the generator constant is $\|\mathbf{1}\|_D = \frac{3}{2}$. In order to compare the results for different values of λ, 10 000 realizations of the generator $2\,(U, 1-U)^\intercal$ were generated. Moreover, for each $\lambda \in \left\{-\frac{\sqrt{2}}{2}, -0.2, 0, 0.2, \frac{\sqrt{2}}{2}\right\}$, 10 000 realizations of a random variable with distribution function H_λ were simulated, cf. (3.2.2), where the rejection method was used in the case $\lambda \neq 0$, see e. g. Falk et al. (2011, Algorithm 5.7.1). Then, for every λ, the copula data were obtained according to (3.2.3) and (3.2.4), where the same generator realizations were used for all λ.

For a sample size of $n = 10\,000$ and for $\delta = 1$, we choose the parameter $c_n = 0.02214$ according to Example 3.4.5. Note that all numbers are rounded to the five decimal places. Table 4.1.2 shows the number of exceedances N_j above the thresholds $\left(1 - \frac{c_n}{j}\right)\mathbf{1}$, $j = 1, 2, 3$, together with their weighted ratios $w_j := \frac{(j+1)N_{j+1}}{jN_j}$, $j = 1, 2$, cf. Lemma 2.2.14 and Section 3.4. As expected, w_1 and w_2 are close to one under $H_{0,1}$. However, this is also true for $\lambda \neq 0$, which is a first indicator that c_n might not be chosen properly.

λ	N_1	N_2	N_3	w_1	w_2
0.00000	366	174	119	0.95082	1.02586
0.70711	306	153	108	1.00000	1.05882
0.20000	322	174	120	1.08075	1.03448
−0.20000	343	169	107	0.98542	0.94970
−0.70711	376	185	126	0.98404	1.02162

Table 4.1.2 Number of exceedances and their weighted ratios.

Now we also consider the estimators of the generator constant under $H_{0,1}$, i. e. $\lambda = 0$, that are given in (3.4.1) and (3.4.2), cf. Corollary 2.2.9. For convenience, Figure 4.1.3 focuses on $k = 3$ and the estimator in (3.4.2), including the approximate 95 % confidence intervals. The remaining plots are very similar. Recall that the given confidence intervals

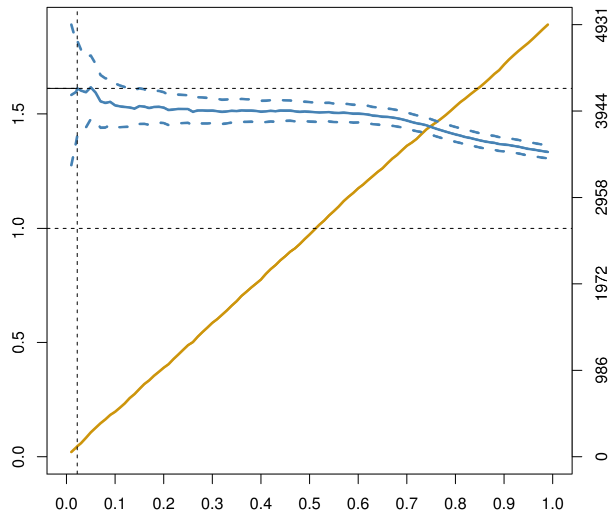

Figure 4.1.3 Estimated generator constant (blue, left scale) for $\lambda = 0$ and $k = 3$ as a function of $c \in (0, 1)$ together with the pointwise approximate 95 % confidence intervals (blue, dashed), cf. (3.4.2). The yellow line (right scale) displays the number of exceedances above the highest threshold, i. e. $(1-\frac{c}{3})\mathbf{1}$. The vertical dashed line emphasizes the value $c_n = 0.02214$, whereas the upper horizontal line marks the corresponding estimate.

have to be interpreted with care. Nevertheless, the relatively large range of the confidence interval at $c = 0.02214$ indicates that c_n was chosen too small. Since we know that the true generator constant is $\frac{3}{2}$, we should select at least a value of 0.5 for c_n.

Finally, Table 4.1.4 summarizes the approximate p-values for the test statistics T_n, $\mathcal{T}_n$, and τ_n. The p-value of the test statistic proposed by Kojadinovic et al. (2011), denoted by ϑ_n, was provided by the R function `evTestC`, which is part of the `copula` package. Note that for $k = 2$, the test statistics T_n, $\mathcal{T}_n$, and τ_n yield almost identical p-values

k	λ	T_n	$\mathcal{T}_n$	τ_n	ϑ_n
2	0.00000	0.34078	0.34076	0.34076	0.00050
	0.70711	1.00000	1.00000	1.00000	0.00050
	0.20000	0.15583	0.15545	0.15545	0.00050
	−0.20000	0.78642	0.78642	0.78642	0.00050
	−0.70711	0.75604	0.75605	0.75605	0.00050
3	0.00000	0.68289	0.56710	0.73626	0.00050
	0.70711	0.56374	0.59498	0.47117	0.00050
	0.20000	0.18783	0.30295	0.14702	0.00050
	−0.20000	0.51940	0.62496	0.39466	0.00050
	−0.70711	0.92345	0.87507	0.94176	0.00050

Table 4.1.4 Approximate p-values for 10 000 copula data from Lemma 3.2.1 and various values of k and λ, where $\delta = 1$ and $c_n = 0.02214$.

since we have in this case $2T_n = \mathcal{T}_n = \tau_n^2$. While the test by Kojadinovic et al. (2011) rejects $H_{0,1}$ in any case, our test statistics never reject the null hypothesis. Since we already supposed that c_n was not chosen properly, a poor performance of the latter tests was to be expected. Altogether, the alternative is not detected by T_n, $\mathcal{T}_n$, and τ_n at a significance level of 5 %.

However, if we consider the p-value as a function of $c \in (0, 1)$, we observe that the shape of its graph depends on the cases $\lambda = 0$, $\lambda > 0$, and $\lambda < 0$. Note that Figure 4.1.5 omits the plots for $\lambda = 0.2$ and $\lambda = -0.2$ since the corresponding shapes are similar to those for $\lambda = 0.70711$ and $\lambda = -0.70711$, respectively. For $\lambda = 0$, the p-values of T_n, $\mathcal{T}_n$, and τ_n are typically above the 5 % line for $c \in (0, 0.5]$. Even if the curve falls below this line on this range, it normally returns to greater values almost instantly. Opposed to that, a p-value curve has for $\lambda > 0$ typically some high peaks for small values of c and then falls below the 5 % line. After another set of peaks for intermediate values of c, the graph normally attains values smaller than 0.05. However, $\lambda < 0$ appears to yield a curve that is above the 5 % line on some interval with left endpoint zero and then falls and stays below this line. Although Figure 4.1.5 suggests that the right endpoint of this interval is relatively close to zero as well, the plots for $\lambda = -0.2$ indicate that the interval may also include intermediate values of c, but with a downward trend. Opposed to that, the curves for $\lambda = 0$ tend to attain large values for c close to 0.5 and then fall below the 5 % line abruptly.

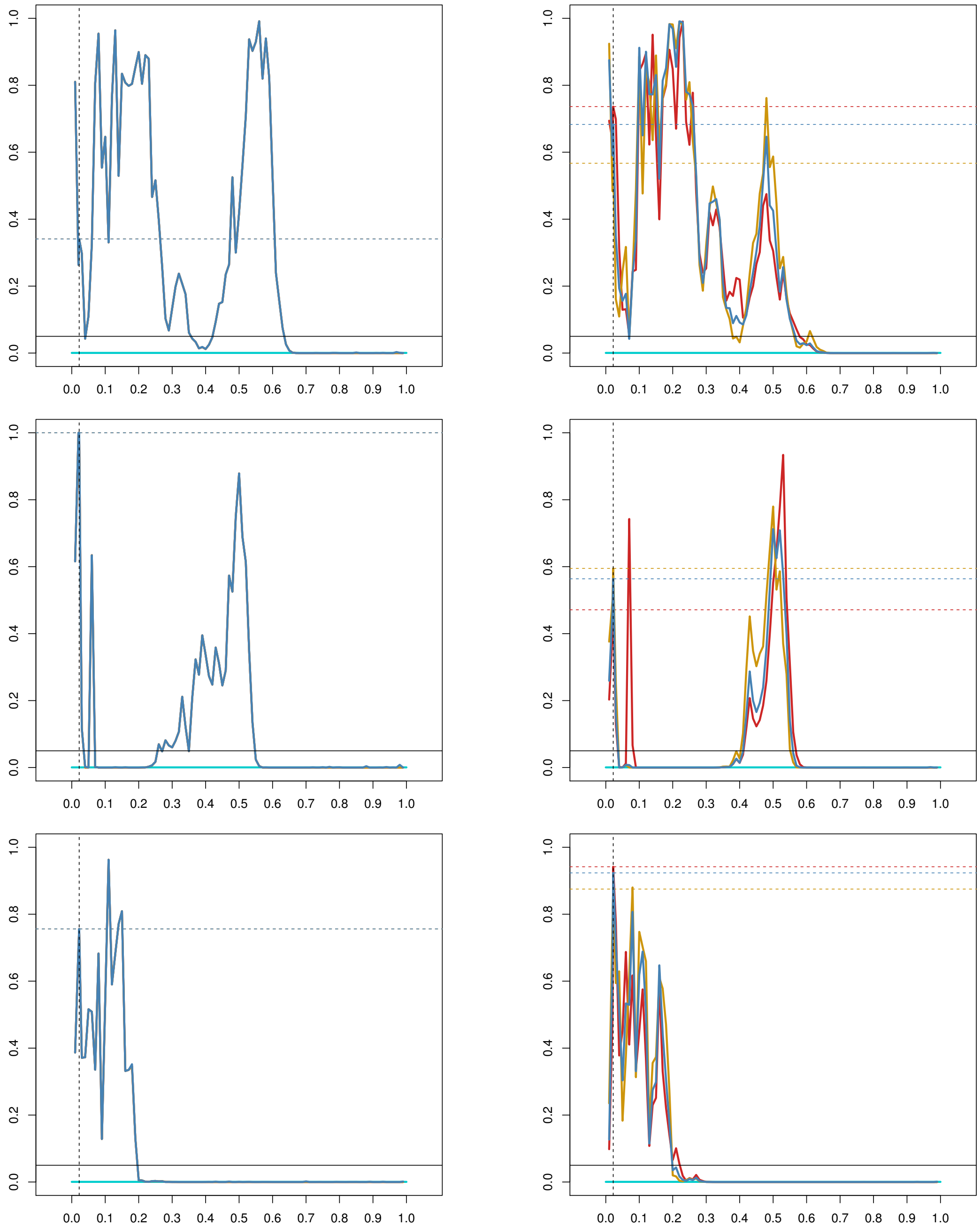

Figure 4.1.5 p-values of T_n (dark blue), $\mathcal{T}_n$ (yellow), τ_n (red), and ϑ_n (light blue) as a function of $c \in (0,1)$. The vertical dashed line emphasizes the value $c = 0.02214$, whereas the horizontal dashed lines mark the corresponding p-values. Top left: $\lambda = 0$, $k = 2$. Middle left: $\lambda = 0.70711$, $k = 2$. Bottom left: $\lambda = -0.70711$, $k = 2$. Top right: $\lambda = 0$, $k = 3$. Middle right: $\lambda = 0.70711$, $k = 3$. Bottom right: $\lambda = -0.70711$, $k = 3$.

In order to complement the above results, we divide for each $\lambda \in \{-\frac{\sqrt{2}}{2}, -0.2, 0, 0.2, \frac{\sqrt{2}}{2}\}$ the corresponding sample of 10 000 copula observations into 50 subsamples. This implies a sample size of $n = 200$ and yields $c_n = 0.09809$. This sample size appears to be too small to obtain stable results since the corresponding p-value curves of the cases $\lambda = 0$, $\lambda > 0$, and $\lambda < 0$ were hardly distinguishable. Among the 50 subsamples, the mean number of exceedances above the thresholds $(1 - \frac{c_n}{2})\mathbf{1}$ and $(1 - \frac{c_n}{3})\mathbf{1}$ were 15.08000 and 10.14000, respectively, for $\lambda = 0$. This indicates that there may be too few observations exceeding the thresholds in order to justify the required approximate normal distribution. However, the quantile plots of the p-values were for $\lambda = 0$ quite close to the main diagonal, both for $k = 2$ and $k = 3$, which indicates a sufficiently well normal approximation. Although the same was true for $\lambda < 0$ and $\lambda = 0.2$, the case $\lambda = 0.70711$ showed a deviation from the main diagonal, cf. Figure 4.1.6.

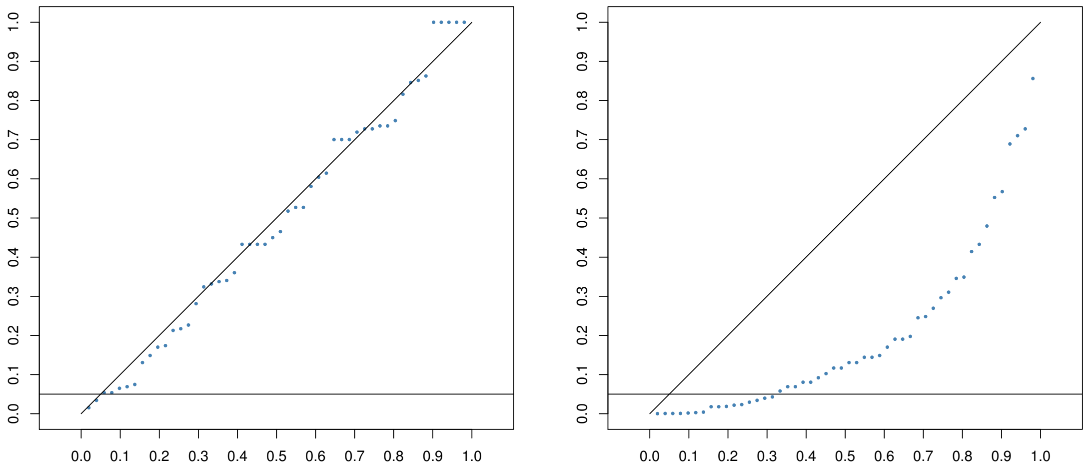

Figure 4.1.6 Quantile plots of the p-values of T_n for $\lambda = 0$ (left) and $\lambda = 0.70711$ (right), where $k = 2$. The x-axis gives the expected order statistics under $H_{0,1}$, whereas the y-axis gives the observed order statistics.

If we consider the rate of rejection — i. e. the number of subsamples where $H_{0,1}$ is rejected divided by the total number of subsamples — and plot it as a function of $c \in (0, 1)$, cf. Figure 4.1.7, we observe that the tests based on T_n, $\mathcal{T}_n$, and τ_n seem to satisfy the type I error of 5 %. Moreover, there is a peak for c slightly larger than 0.1 if $\lambda = 0.70711$, and none of the test statistics T_n, $\mathcal{T}_n$, and τ_n appears to be superior to the others. The test statistic ϑ_n by Kojadinovic et al. (2011), which does not depend on c, rejects $H_{0,1}$ in all subsamples, no matter whether the hypothesis is true or not. This indicates that the test is quite sensitive to condition (1.1.16) and that it ignores the tail equivalence of an EVC and its GPC.

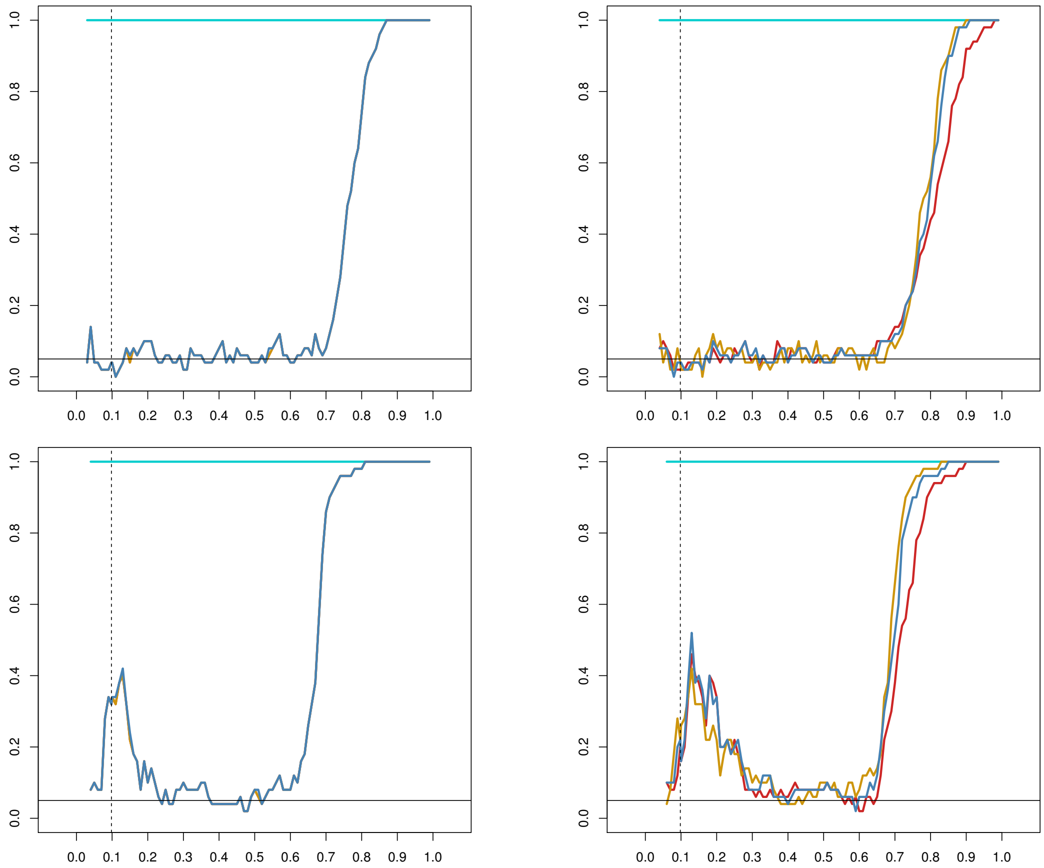

Figure 4.1.7 Rates of rejection for T_n (dark blue), $\mathcal{T}_n$ (yellow), τ_n (red), and ϑ_n (light blue) among the test results of the 50 subsamples. Top left: $\lambda = 0$, $k = 2$. Top right: $\lambda = 0$, $k = 3$. Bottom left: $\lambda = 0.70711$, $k = 2$. Bottom right: $\lambda = 0.70711$, $k = 3$.

Figure 4.1.7 motivates to choose, e. g., $c_n = 0.11$. In order to verify the performance of this value, a new data set was generated for each $\lambda \in \{-\frac{\sqrt{2}}{2}, -0.2, 0, 0.2, \frac{\sqrt{2}}{2}\}$ as described above. But this time, c_n was put to 0.11, regardless of whether the sample size was 10 000 or 200. For $n = 10\,000$, Table 4.1.8 suggests that $c_n = 0.11$ performs much better than the original value of c_n. For both, $k = 2$ and $k = 3$, the null hypothesis is rejected for $\lambda = 0.70711$. Moreover, the case $k = 3$ seems to perform better than the case $k = 2$; note in particular the p-values for $\lambda = -0.70711$. Moreover, the test statistic τ_n seems to perform slightly better for $k = 3$ than T_n and $\mathcal{T}_n$. However, all three tests have difficulties to detect the alternative when λ is relatively close to zero. Splitting the whole data set into 50 subsamples led to similar results as above, i. e. there was no observable gain of performance.

k	λ	T_n	$\mathcal{T}_n$	τ_n	ϑ_n
2	0.00000	0.15428	0.15422	0.15422	0.00050
	0.70711	0.00008	0.00000	0.00000	0.00050
	0.20000	0.54651	0.54649	0.54649	0.00050
	−0.20000	0.94131	0.94131	0.94131	0.00050
	−0.70711	0.34410	0.34506	0.34506	0.00050
3	0.00000	0.28727	0.09467	0.87410	0.00050
	0.70711	0.00000	0.00000	0.00000	0.00050
	0.20000	0.76429	0.62867	0.91527	0.00050
	−0.20000	0.87486	0.87281	0.75431	0.00050
	−0.70711	0.13958	0.25927	0.10585	0.00050

Table 4.1.8 Approximate p-values for 10 000 copula data from Lemma 3.2.1 and various values of k and λ, where $\delta = 1$ and $c_n = 0.11$.

Higher Dimensional Copulas

While the above reasoning mainly dealt with deriving an adequate value for c_n, we now briefly consider the impact of the dimension of the data on the simulation results. As above, a similar analysis for the copula family in Lemma 3.2.7 suggests to choose $c_n = 0.11$. Note that this family can easily be extended to higher dimensions: If $U_1, \dots, U_d$ are independent and on $[0,1]$ uniformly distributed random variables, then

$$-\frac{V}{2}\left(\frac{1}{U_1}, \dots, \frac{1}{U_d}\right)^{\mathsf{T}}$$

follows a standard GPD if $V \sim \mathcal{U}[0,1]$. However, if $V \sim H_\lambda$ for $\lambda \neq 0$, then the above random vector is not in the domain of attraction of an MSD. As it turned out that $c_n = 0.11$ is also valid for $d = 5$, we briefly compare the results for $d = 2$ and $d = 5$ with one another. However, due to technical restrictions in the simulation program, the data sets for $d = 2$ and $d = 5$ are independent.

Table 4.1.9 shows that for the present copula family, the performance of detecting the alternative is much better than for the copulas underlying Table 4.1.8. For $k = 2$, almost all alternatives under consideration are detected, apart from the case $\lambda = -0.20000$. Another difference compared to Table 4.1.8 is that the overall performance seems to decrease slightly if k is increased from 2 to 3, cf. Table 4.1.10. Moreover, a comparison of Table 4.1.9 with Table 4.1.10 suggests that a higher dimension requires a larger sample size in order to detect the alternative: For $d = 5$, we observe that $H_{0,1}$ is rejected only in the case $\lambda = 0.70711$. However, these differences could also be due to random fluctuations since the samples underlying Table 4.1.9 and Table 4.1.10 are independent.

k	λ	$T_n^{(2)}$	$\mathcal{T}_n^{(2)}$	$\tau_n^{(2)}$	$\vartheta_n^{(2)}$
2	0.00000	0.15404	0.15372	0.15372	0.00050
	0.70711	0.00002	0.00000	0.00000	0.00050
	0.20000	0.00161	0.00159	0.00159	0.00050
	−0.20000	0.43441	0.43441	0.43441	0.00050
	−0.70711	0.03528	0.03540	0.03540	0.00050
3	0.00000	0.27252	0.08550	0.85387	0.00050
	0.70711	0.00000	0.00000	0.00000	0.00050
	0.20000	0.01496	0.00631	0.03477	0.00050
	−0.20000	0.44287	0.27067	0.65694	0.00050
	−0.70711	0.06357	0.09665	0.06159	0.00050

Table 4.1.9 Approximate p-values for 10 000 copula data from Lemma 3.2.7 and various values of k and λ, where $\delta = 1$, $c_n = 0.11$, and $d = 2$.

k	λ	$T_n^{(5)}$	$\mathcal{T}_n^{(5)}$	$\tau_n^{(5)}$	$\vartheta_n^{(5)}$
2	0.00000	0.63010	0.63010	0.63010	0.00050
	0.70711	0.00007	0.00000	0.00000	0.00050
	0.20000	0.24500	0.24501	0.24501	0.00050
	−0.20000	0.13132	0.13130	0.13130	0.00050
	−0.70711	0.07823	0.07808	0.07808	0.00050
3	0.00000	0.89392	0.88689	0.78292	0.00050
	0.70711	0.00000	0.00000	0.00000	0.00050
	0.20000	0.55954	0.37505	0.78853	0.00050
	−0.20000	0.17679	0.27559	0.14325	0.00050
	−0.70711	0.16378	0.20738	0.15429	0.00050

Table 4.1.10 Approximate p-values for 10 000 copula data from Lemma 3.2.7 and various values of k and λ, where $\delta = 1$, $c_n = 0.11$, and $d = 5$.

4.2 More General Data

In Section 4.1 we have seen that determining a proper value for c_n can be quite difficult. In the more general case, where the margins of the data are unknown, we also need to choose another parameter m_n. Therefore, this thesis ends with an approach on how to exploit the results of Section 3.4 in that framework. Analogously to Section 4.1, we consider for each $\lambda \in \{-\frac{\sqrt{2}}{2}, -0.2, 0, 0.2, \frac{\sqrt{2}}{2}\}$, the data sets of size 10 000 which were computed at the beginning of that section. However, we assume we would not know that the data follow a copula.

Again, we choose $\delta = 1$. Now initial values for c_n and m_n are computed from Example 3.4.9, which yields $c_n = 0.03673$ and $m_n = 2037$. In particular, m_n is assumed to be a function of c_n. Analogously to Figure 4.1.3, we can thus plot the estimators in (3.4.10) and (3.4.11), cf. Figure 4.2.1. Note that plotting the p-value curves directly by exploiting Example 3.4.9 leads typically to non-distinguishable curves, no matter whether $H_{0,1}$ is true or not.

For convenience, we use the plots for the data corresponding to $\lambda = 0$. However, similar results would be obtained if we observed, e. g., the data set with $\lambda = 0.70711$. According to (2.2.17), the empirical threshold is computed by transforming the vector $(1 - \frac{c_n}{j})\mathbf{1}$ componentwise with the empirical quantile functions of the margins. This is why we call $1 - \frac{c_n}{j}$ a threshold *level*.

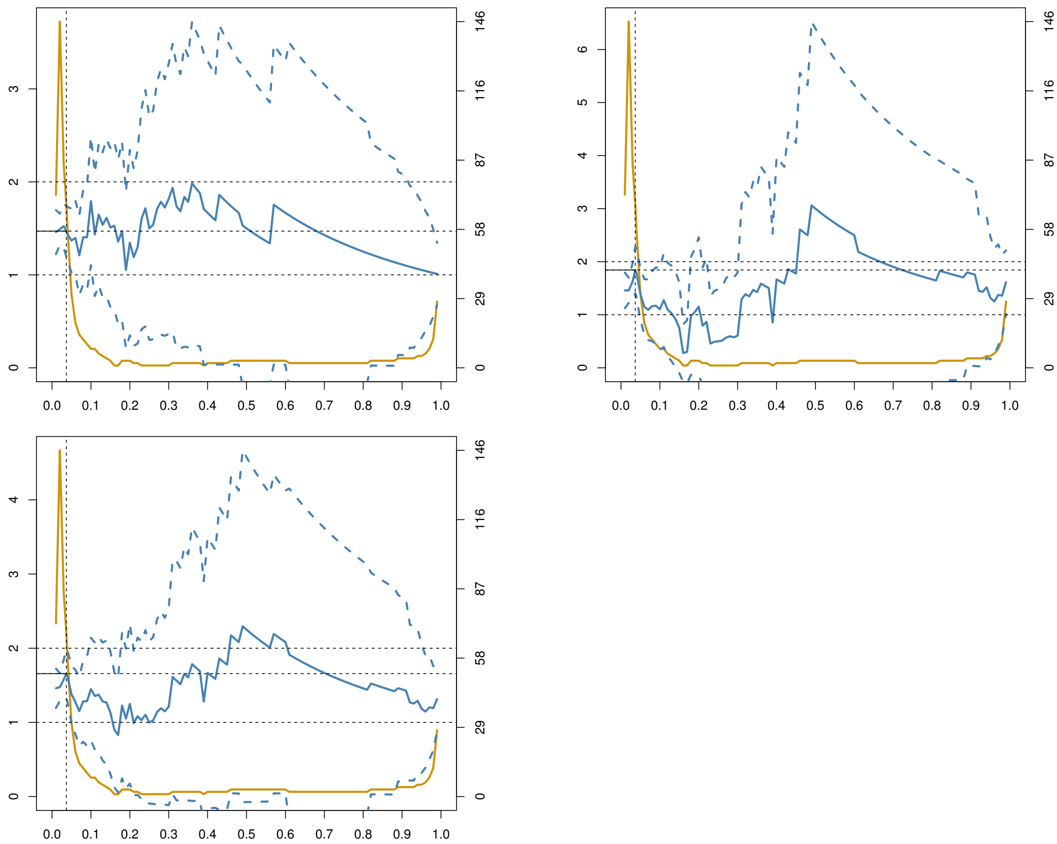

Figure 4.2.1 Estimated generator constant (blue, left scale) for $\lambda = 0$ and $k = 2$ as a function of $c \in (0, 1)$ together with the pointwise approximate 95 % confidence intervals (blue, dashed), cf. (3.4.10) and (3.4.11). The yellow line (right scale) displays the number of exceedances above the highest threshold level, i. e. $1 - \frac{c}{2}$. Top left: (3.4.10) with $j = 1$. Top right: (3.4.10) with $j = 2$. Bottom: (3.4.11).

Due to the functional dependence $m_n = m_n(c)$, m_n is close to zero for intermediate values of c and increases when c approaches zero. This is why the yellow curves in Figure 4.2.1 are not linear. Indeed, the peak close to zero could be misleading: It is *not* desirable to choose c_n such that the number of exceedances above the highest threshold level is maximized! On the one hand, the peaks in the yellow curves represent the case $m_n = n$; recall that we require $\frac{m_n}{n}\log(m_n) \to 0$ as $n \to \infty$. On the other hand, we already noticed in Section 4.1 that $c_n = 0.03673$ should be too small.

However, we have 69 exceedances above the highest threshold level among the first m_n observations, which, roughly speaking, seems to be enough to justify the desired normal approximation. Now we choose c_n such that the estimated generator constant in Figure 4.2.1 is almost constant for $c \leq c_n$ and such that the approximate confidence intervals have a rather small range, e. g. $c_n = 0.11$, as before.

This heuristically motivated approach suggests therefore to put $m_n = 2037$ and $c_n = 0.11$. In order to verify these values, we generate — as in Section 4.1 — a new sample for each $\lambda \in \{-\frac{\sqrt{2}}{2}, -0.2, 0, 0.2, \frac{\sqrt{2}}{2}\}$, where we choose $m_n = 2037$ *independently* of c_n. Indeed, the resulting p-value curves had the characteristic shapes as discussed in Figure 4.1.5 and yielded similar results as in Section 4.1, i. e. $H_{0,1}$ was rejected for $\lambda = 0.70711$, but not for $\lambda \in \{-\frac{\sqrt{2}}{2}, -0.2, 0, 0.2\}$. Note that a further analysis for subsamples of the data set was skipped, due to very few observations above a high threshold even in the case where the margins are known. Since a sample size of $n = 200$ would imply $m_n = 86$, the number of exceedances would become even smaller.

Summary and Outlook

This thesis began with a review of generalized Pareto distributions, which are known to be crucial for modeling extremal events, in finite and infinite dimensions. Since both these frameworks share some crucial properties and due to Sklar's theorem, we were able to define certain δ-neighborhoods of a generalized Pareto copula simultaneously for both cases. Then we considered several tests for these neighborhoods and obtained that the finite dimensional versions of these tests reasonably approximate their functional counterparts under certain regularity conditions.

Although we had to assume in the functional case that a continuous copula process exists, Chapter 3 has shown that this assumption is, roughly speaking, not too restrictive. Furthermore, we considered examples of both, copulas that are in a δ-neighborhood of a GPC and those that are not. These examples covered again the finite dimensional as well as the functional framework.

After a discussion of how the parameters of the proposed tests could be chosen, we applied these strategies to simulated data. Sadly, technical restrictions did not allow to increase the sample sizes any further such that the results of this simulation study are on the one hand promising, but on the other hand rather weak. However, the simulations did provide some hints for the application to real data, which worked in some cases even for small sample sizes. Although the tail equivalence of an extreme value copula and the corresponding generalized Pareto copula motivated to compare our tests with the one by Kojadinovic et al. (2011), it seems like the latter is too restrictive to small deviations from an extreme value copula.

Natural extensions for future research are, of course, to simulate larger sample sizes with more powerful hardware. It could be also useful to implement time demanding tasks not in `R` but in another programming language like `C++`, which was however beyond the scope of this thesis. Moreover, it is desirable to deduce more theoretical results about the asymptotic behavior of the proposed tests. In particular, it should be possible to replace the condition $\frac{m_n}{n}\log(m_n) \to 0$ as $n \to \infty$ with $\frac{m_n}{n} \to 0$.

Particularly the simulation results with large sample sizes indicate a high potential of the proposed tests. A very interesting topic for future research!

Bibliography

Anderson, T. W., and Stephens, M. A. (1997). The continuous and discrete brownian bridges: representations and applications. *Linear Algebra Appl.* **264**, 145–171. doi:10.1016/S0024-3795(97)00015-3.

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

Aulbach, S., and Falk, M. (2012). Testing for a generalized Pareto process. *Electron. J. Stat.* **6**, 1779–1802. doi:10.1214/12-EJS728.

Aulbach, S., Falk, M., and Hofmann, M. (2012*b*). The multivariate piecing-together approach revisited. *J. Multivariate Anal.* **110**, 161–170. doi:10.1016/j.jmva.2012.02.002.

Aulbach, S., Falk, M., and Hofmann, M. (2013). On max-stable processes and the functional D-norm. *Extremes* **16**, 255–283. doi:10.1007/s10687-012-0160-3.

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

Aulbach, S., Falk, M., and Zott, M. (2015*b*). The space of D-norms revisited. *Extremes* **18**, 85–97. doi:10.1007/s10687-014-0204-y.

Balkema, A. A., and de Haan, L. (1974). Residual life time at great age. *Ann. Probab.* **2**, 792–804. doi:10.1214/aop/1176996548.

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

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications.* Wiley Series in Probability and Statistics. Wiley, Chichester, UK. doi:10.1002/0470012382.

Berghaus, B., Bücher, A., and Dette, H. (2013). Minimum distance estimators of the Pickands dependence function and related tests of multivariate extreme-value dependence. *J. SFdS* **154**, 116–137, http://journal-sfds.fr/index.php/J-SFdS/article/view/158.

Billingsley, P. (2012). *Probability and Measure.* Wiley Series in Probability and Statistics, Anniversary ed. Wiley, New York.

de Bruijn, N. G. (1981). *Asymptotic Methods in Analysis.* Dover Publications, New York. Unabridged and corrected republication of the third edition (1970) of the work originally published by North-Holland Publishing, Amsterdam, 1958.

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

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). On the Lambert W function. *Adv. Comput. Math.* **5**, 329–359. doi:10.1007/BF02124750.

Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Pub. Inst. Stat. Univ. Paris* **23**, 1–36.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.), 117–130. D. Reidel, Dordrecht.

Dombry, C., Engelke, S., and Oesting, M. (2015). Exact simulation of max-stable processes. Tech. Rep. arXiv:1506.04430 [math.ME].

Dombry, C., and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface* **8**, 9–17. doi:10.4310/SII.2015.v8.n1.a2.

Duchesne, P., and Lafaye de Micheaux, P. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Comput. Statist. Data Anal.* **54**, 858–862. doi:10.1016/j.csda.2009.11.025.

Einmahl, J. H. J., de Haan, L., and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist.* **34**, 1987–2014. doi:10.1214/009053606000000434.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, *Applications of Mathematics - Stochastic Modelling and Applied Probability*, vol. 33. Springer, Berlin. doi:10.1007/978-3-642-33483-2.

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., and Hofmann, M. (2011). Sojourn times and the fragility index. *Stochastic Process. Appl.* **122**, 1110–1128. doi:10.1016/j.spa.2011.11.009.

Falk, M., Hofmann, M., and Zott, M. (2014). On generalized max-linear models and their statistical interpolation. Tech. Rep., University of Würzburg. To appear in *J. Appl. Probab.*, arXiv:1303.2602v2 [math.ST].

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

Falk, M., and Reiss, R.-D. (2002). A characterization of the rate of convergence in bivariate extreme value models. *Statist. Probab. Lett.* **59**, 341–351. doi:10.1016/S0167-7152(02)00209-2.

Ferreira, A., and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20**, 1717–1737. doi:10.3150/13-BEJ538.

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.

Fortiana, J., and Cuadras, C. M. (1997). A family of matrices, the discretized brownian bridge, and distance-based regression. *Linear Algebra Appl.* **264**, 173–188. doi:10.1016/S0024-3795(97)00051-7.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. Wiley Series in Probability and Mathematical Statistics, 1st ed. Wiley, New York.

Ghorbal, N. B., Genest, C., and Nešlehová, J. (2009). On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. *Canad. J. Statist.* **37**, 534–552. doi:10.1002/cjs.10034.

Ghoudi, K., Khoudraji, A., and Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.* **26**, 187–197. doi:10.2307/3315683.

Giné, E., Hahn, M., and Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probab. Theory Related Fields* **87**, 139–165. doi:10.1007/BF01198427.

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.

Gudendorf, G., and Segers, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009* (P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, eds.), *Lecture Notes in Statistics*, vol. 198, 127–145. Springer, Berlin. doi:10.1007/978-3-642-12465-5_6.

Gut, A. (2013). *Probability: A Graduate Course*, *Springer Texts in Statistics*, vol. 75. 2nd ed. Springer, New York. doi:10.1007/978-1-4614-4708-5.

de Haan, L., and Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* Springer Series in Operations Research and Financial Engineering. Springer, New York. doi:10.1007/0-387-34471-3. See http://people.few.eur.nl/ldehaan/EVTbook.correction.pdf and http://home.isa.utl.pt/~anafh/corrections.pdf for corrections and extensions.

de Haan, L., and Lin, T. (2001). On convergence toward an extreme value distribution in $C[0,1]$. *Ann. Probab.* **29**, 467–483. doi:10.1214/aop/1008956340.

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

Hashorva, E. (2005). Asymptotics and bounds for multivariate Gaussian tails. *J. Theoret. Probab.* **18**, 79–97. doi:10.1007/s10959-004-2577-3.

Hashorva, E., and Hüsler, J. (2003). On multivariate Gaussian tails. *Ann. Inst. Statist. Math.* **55**, 507–522. doi:10.1007/BF02517804.

Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2015). *copula: Multivariate Dependence with Copulas.* R package version 0.999-13, https://cran.r-project.org/package=copula.

Hofmann, D. (2009). *Characterization of the D-Norm Corresponding to a Multivariate Extreme Value Distribution.* Ph.D. thesis, University of Würzburg, http://opus.bibliothek.uni-wuerzburg.de/volltexte/2009/4134/.

Hofmann, M. (2012). *Contributions to Extreme Value Theory in the Space* $C[0,1]$. Ph.D. thesis, University of Würzburg, http://opus.bibliothek.uni-wuerzburg.de/volltexte/2012/7440/.

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

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q.J.R. Meteorol. Soc.* **81**, 158–171. doi:10.1002/qj.49708134804.

Kaufmann, E., and Reiss, R.-D. (1995). Approximation rates for multivariate exceedances. *J. Statist. Plann. Inference* **45**, 235–245. doi:10.1016/0378-3758(94)00074-3.

Kojadinovic, I., Segers, J., and Yan, J. (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *Canad. J. Statist.* **39**, 703–720. doi:10.1002/cjs.10110.

Kortschak, D., and Albrecher, H. (2009). Asymptotic results for the sum of dependent non-identically distributed random variables. *Methodol. Comput. Appl. Probab.* **11**, 279–306. doi:10.1007/s11009-007-9053-3.

Lafaye de Micheaux, P. (2013). *CompQuadForm: distribution function of quadratic forms in normal variables.* R package version 1.4.1, https://cran.r-project.org/package=CompQuadForm.

Larsson, M., and Nešlehová, J. (2011). Extremal behavior of Archimedean copulas. *Adv. in Appl. Probab.* **43**, 195–216. doi:10.1239/aap/1300198519.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer Series in Statistics. Springer, New York. doi:10.1007/978-1-4612-5449-2.

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.

Michel, R. (2006). *Simulation and Estimation in Multivariate Generalized Pareto Models*. Ph.D. thesis, University of Würzburg, http://opus.bibliothek.uni-wuerzburg.de/volltexte/2006/1848/.

Michel, R. (2008). Some notes on multivariate generalized Pareto distributions. *J. Multivariate Anal.* **99**, 1288–1301. doi:10.1016/j.jmva.2007.08.007.

von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique* **1**, 141–160. Reproduced in *Selected Papers of Richard von Mises*, Amer. Math. Soc., Vol. 2, 271–294 (1964).

Moore, D. S., and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3**, 599–616, http://projecteuclid.org/euclid.aos/1176343125.

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 (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131. doi:10.1214/aos/1176343003.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org/.

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., and Thomas, M. (2007). *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields*. 3rd ed. Birkhäuser, Basel. doi:10.1007/978-3-7643-7399-3.

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.

Ribatet, M., and Sedki, M. (2013). Extreme value copulas and max-stable processes. *J. SFdS* **154**, 138–150, http://journal-sfds.fr/index.php/J-SFdS/article/view/160.

Robbins, H., and Pitman, E. J. G. (1949). Application of the method of mixtures to quadratic forms in normal variates. *Ann. Math. Statist.* **20**, 552–560. doi:10.1214/aoms/1177729947.

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

Rüschendorf, L. (2009). On the distributional transform, Sklar's theorem, and the empirical copula process. *J. Statist. Plann. Inference* **139**, 3921–3927. doi:10.1016/j.jspi.2009.05.030.

Savage, I. R. (1962). Mills' ratio for multivariate normal distributions. *J. Res. Nat. Bur. Standards* **66B**, 93–96. doi:10.6028/jres.066B.011.

Schilling, R. L., and Partzsch, L. (2014). *Brownian Motion: An Introduction to Stochastic Processes.* De Gruyter Textbook, 2nd ed. De Gruyter, Berlin.

Schweizer, B., and Sklar, A. (2005). *Probabilistic Metric Spaces.* Dover Publications, Mineola, NY. Unabridged republication of the edition published by Elsevier Science Publishing, New York, 1983. A new Preface to the Dover Edition, list of Errata, Notes, and Supplementary References have been added.

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. L. (1990). Max-stable processes and spatial extremes. Preprint, Univ. North Carolina, http://www.stat.unc.edu/faculty/rs/papers/RLS_Papers.html.

Tajvidi, N. (1996). *Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalised Pareto Distributions.* Ph.D. thesis, Chalmers University of Technology, Gothenburg, http://www.maths.lth.se/matstat/staff/nader/fullpub.html.

Takahashi, R. (1988). Characterizations of a multivariate extreme value distribution. *Adv. in Appl. Probab.* **20**, 235–236. doi:10.2307/1427279.

Tong, Y. L. (1990). *The Multivariate Normal Distribution.* Springer Series in Statistics. Springer, New York. doi:10.1007/978-1-4613-9655-0.

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 16–27, 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.

Wang, Y., and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Adv. in Appl. Probab.* **43**, 461–483, http://projecteuclid.org/euclid.aap/1308662488.