Université catholique de Louvain
Institut de statistique, biostatistique et sciences actuarielles
Thèse de doctorat

Nonparametric Estimation of Multivariate Extreme-Value Copulas

Membres du jury
Prof. Rainer von Sachs
Prof. Johan Segers
Prof. Ingrid van Keilegom
Prof. Christian Hafner
Prof. Roel Braekers

Thèse présentée en vue de l'obtention
du grade de Docteur en Sciences
(orientation statistique) par:
Gordon Gudendorf

Louvain-la-Neuve
Septembre 2012

Contents

List of Symbols

$\vee$	Maximum	10
$\wedge$	Minimum	15
$\rightsquigarrow$	Weak convergence	36
Δ_{p-1}	Unit-simplex in $\mathbb{R}^{p-1}$	11
$\mathscr{C}(\mathscr{W})$	Space of real-valued continuous functions on $\mathscr{W}$	37
$L^2(\mathscr{W})$	Space of real-valued square-integrable functions on $\mathscr{W}$	79
$\ell^\infty(\mathscr{W})$	Space of real-valued bounded functions on $\mathscr{W}$	50
$\mathscr{A}$	Class of Pickands dependence functions	79
A	Pickands dependence function defined on Δ_{p-1}	13
ℓ	Tail dependence function defined on $[0, +\infty)^p$	11
τ_K	Kendall's tau	23
ρ_S	Spearman's rho	7
$F^{\leftarrow}$	Generalized inverse function	6
$F_{n,j}$	Empirical marginal distribution with division by $n+1$	56
$E_{n,j}$	Empirical marginal distribution with division by n	76
C_n	Multivariate empirical distribution function	56
α_n	Multivariate empirical distribution process	56
$\hat{C}_n$	Empirical copula	56
$\mathbb{C}_n$	Empirical copula process	56
$\hat{C}_n^D$	Empirical copula process as defined in Deheuvels [1979]	76
$\hat{C}_n^A$	Nonparametric extreme-value copula estimate	93
$\hat{U}_{i,j}$	Rank-based observations	56
$U_{i,j}$	Uniform random variables	27

Acknowledgements

First and foremost, I am deeply grateful to my supervisor Johan Segers for introducing me to the field of extreme-value copulas and for sharing his insights with me during extensive discussions over the last 4 years. Without his invaluable scientific advice, his enthusiasm and encouragement, it would have been impossible to realize this work.

Furthermore, my gratitude goes also to Ivan Kojadinovic for his patience and friendly help during his last visit in the Institute of Statistics.

I also benefited a lot from the support of Alain Guillet and Jean-Luc Marrion, particularly during the time when I started to program in C/C++.

Last but not least, I need also to thank my mother and my colleagues in the Institute for their moral support.

Introduction

The notion of copula introduced for the first time in Sklar [1959] provides a flexible tool to describe the interrelation of several random variables, as it allows to link the univariate distributions to the multivariate distribution. To be precise, a copula is a function $C : [0,1]^p \to [0,1]$ with

$$F(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p)), \quad (x_1, \dots, x_p) \in \mathbb{R}^p$$

with F being the joint distribution function and F_j, for $j \in \{1, \dots, p\}$, being the univariate marginal distributions.

The financial crisis starting in 2008 has shown that the majority of the statistical models used by the financial industry were not able to capture the joint behavior of extreme events. Practicioners tend to use the multivariate Normal distribution for its tractable properties even though it is widely known that the multivariate Normal distribution is not able to model strong dependence in the tails of the distribution. In this context the paper of Li [2000] has reached world fame. To model the relationship between the default time for loans in the mortgage market, he used the Gaussian copula, which is derived from the multivariate Normal distribution and hence suffers from the same drawbacks.

In various domains, as for example finance, insurance or environmental science, joint extreme events can have a serious impact and therefore need careful modeling. Think for instance of daily water levels at two different locations in a lake during a year. Calculation of the probability that there is a flood exceeding a certain benchmark requires knowledge of the joint distribution of maximal heights during the forecasting period. This is a typical field of application for extreme-value theory. In such situations, extreme-value copulas can be considered to provide appropriate models for the dependence structure between exceptional events.

One of the first applications of bivariate extreme-value analysis must be due to Gumbel and Goldstein [1964], who analyze the maximal annual discharges of the Ocmulgee River in Georgia at two different stations, a dataset that has been taken up again in Jiménez et al. [2001a]. The joint behavior of extreme returns in the foreign exchange rate market is investigated in Stărică [1999], whereas the comovement of equity markets characterized by high volatility levels is studied in Longin and Solnik [2001]. An application in the insurance domain can be found in Cebrián et al. [2003].

Extreme-value copulas do not only arise naturally in the domain of extreme events, but they can also be a convenient choice to model data with positive dependence. An advantage with respect to the much more popular class of Archimedean copulas, for instance, is that they are not necessarily symmetric. Incidentally, a hybrid class containing both the Archimedean and the extreme-value copulas as a special case are the Archimax copulas in Capéraà et al. [2000].

The interest of this work is focused on the nonparametric estimation of extreme-value copulas. A multivariate extreme-value copula admits the following expression

$$C(\boldsymbol{u}) = \exp\left\{ \left(\sum_{j=1}^{p} \log u_j \right) A \left(\frac{\log u_1}{\sum_{j=1}^{p} \log u_j}, \ldots, \frac{\log u_{p-1}}{\sum_{j=1}^{p} \log u_j} \right) \right\}$$

with $\boldsymbol{u} = (u_1, \ldots, u_p) \in (0,1]^p \setminus \{(1, \ldots, 1)\}$ and where A is known as the Pickands dependence function introduced in Pickands [1981].

Let us now review the structure of the thesis explaining the motivations behind the different projects. The different aspects in bivariate extreme-value copulas have already known some treatment, while their higher-dimensional homologues have not experienced the same degree of attention up to that date. Hence, the objective of this work is to fill partially that gap and in particular concentrate on the nonparametric estimation of extreme-value copulas in arbitrary dimensions.

After a brief overview of the concept of copulas, a large part of Chapter 1 will be dedicated to an in-depth introduction of the class of extreme-value copulas starting from the definition of extreme-value copulas in the most general case, i.e. using the integral representation involving the associated spectral measure. After highlighting the dependence properties of extreme-value copulas, this chapter will be completed by a review of some parametric models.

A brief guide of some of the literature about the estimation of extreme-value copulas will be given in Chapter 2 giving the chance to present some recent approaches as for example the minimum-distance estimators in Peng et al. [2011] and Bücher et al. [2011] or some parametric methods on which we will not insist further. The new research in this work will focus more on the classical estimators referred hereafter as moment estimators as for example the estimator presented in Pickands [1981] or the estimator in Capéraà et al. [1997].

The subsequent chapters will deal with the new research results achieved in the last years; some of them have already been published. First, we consider the hypothetical situation where the marginal distributions are known. In this more theoretical framework, we succeed to simplify the expressions for the nonparametric estimator introduced in Zhang et al. [2008], in fact the multivariate version of the CFG estimator introduced in Capéraà et al. [1997]. To impose some shape constraints on the estimates of A, the estimator requires some weight functions whose influence on the quality of the fit had not been investigated before. The main achievement of this project is to determine the weight functions leading to the CFG-estimator with minimum asymptotic variance. In practice, this optimal estimator can be computed as the intercept in a certain linear regression framework. A significant reduction of the variance can be ascertained in the simulation studies.

In Chapter 4, we will drop the assumption of known margins and replace the marginal distributions by their empirical counterparts. This chapter will then be completely dedicated to establish the asymptotic properties of the classical moment estimators in the multivariate setting, something which had not been done before. In fact, this work boils down to adapt the already established results for bivariate copulas in Genest and Segers [2009] and Segers [2012] to the higher-dimensional case.

As will be made clear in Chapter 1, the Pickands dependence function A needs to verify certain boundary constraints, which none of the estimators presented up to that point can verify. Therefore, we will extend in Chapter 5 the projection methodology presented initially in Fils-Villetard et al. [2008] allowing to transform any initial estimate for A into a valid Pickands dependence function. Although the context and the objective are the same as in the bivariate case, the methodology cannot been taken over straightforwardly. In fact, in the general case the projection process consists in modeling the spectral measure associated to the Pickands dependence function which could be avoided in dimension $p = 2$ due to the properties of extreme-value copulas.

Finally, the last chapter of this work presents a test for multivariate extreme-value dependence testing whether a copula belongs to the class of extreme-value copulas or not. The only other test currently available for this purpose in Kojadinovic et al. [2011] is based on the max-stable characterization of extreme-value copulas. In Chapter 6, we will try to justify why it might preferable to build a test using the characterization using Pickands dependence function. The new test statistic measures the distance between the empirical copula and an extreme-value copula estimate using the estimators presented in the previous chapters. The implementation is based on the multiplier resampling method introduced in Rémillard and Scaillet [2009] to resample under the null hypothesis. To establish the asymptotic properties of this test procedure, one can heavily rely on the results in Chapter 4.

1 Extreme-value copulas

This chapter, inspired from the survey paper Gudendorf and Segers [2010], will introduce the class of extreme-value copulas and its properties to the reader. In Section 1.1, the definition of copulas, some associated dependence concepts and a small number of parametric models are given. The definition, the origin, and some basic properties of extreme-value copulas are presented in Section 1.2. Multivariate extreme-value copulas are determined by their so-called spectral measure H; details will be given in 1.3. A number of useful and popular parametric families are reviewed in Section 1.4. Section 1.5 provides a discussion of the most important dependence coefficients associated to extreme-value copulas. Finally, some further topics and pointers to the literature are gathered in Section 1.6.

1.1 Copulas

1.1.1 Introduction

A p-dimensional copula C can be defined as a multivariate distribution function with standard uniform marginal distributions.

The notion of copulas was introduced in Sklar [1959], where it was established that, under the necessary continuity assumptions, the construction of a multivariate model can be separated in two parts: the marginals and the copula. Hence, the copula allows the margin-free modeling of the dependence structure of a multivariate general distribution.

Theorem 1.1. *Let F be a joint cumulative distribution function with continuous margins $F_1, \dots, F_p$. Then there exists a unique copula $C : [0,1]^p \to [0,1]$ such that*

$$F(x_1, \dots, x_p) = C\left(F_1(x_1), \dots, F_p(x_p)\right) \tag{1.1}$$

holds. Conversely, if C is a copula and $F_1, \ldots, F_p$ are distribution functions, then the function F given by (1.1) *is a joint distribution function with margins $F_1, \ldots, F_p$.*

We extract a unique copula from a multivariate distribution function F with continuous margins $F_1, \ldots, F_p$ as follows

$$C(\boldsymbol{u}) = F\left(F_1^{\leftarrow}(u_1), \ldots, F_p^{\leftarrow}(u_p)\right), \quad \boldsymbol{u} = (u_1, \ldots, u_p) \in [0,1]^p \qquad (1.2)$$

where $F_1^{\leftarrow}, \ldots, F_p^{\leftarrow}$ are the generalized inverses of $F_1, \ldots, F_p$ defined as $F_j^{\leftarrow}(y) = \inf\{x : F_j(x) \geqslant y\}$ for $y \in [0,1]$.

Even though there exist alternative ways to introduce the concept of copula, as for example done in Nelsen [2006], the probably more natural and intuitive approach in Sklar's theorem is preferred in this text.

The use of copulas in the description of multivariate dependence structures is partially justified by the following theorem.

Theorem 1.2. *Consider the random variables $X_1, \ldots, X_p$ with copula function C. Let $T_i : \mathbb{R} \to \mathbb{R}$ for $i \in \{1, \ldots, p\}$ be strictly increasing functions. Then the copula of the random vector $\big(T_1(X_1), \ldots, T_p(X_p)\big)$ is given by C.*

This theorem states that the associated copula function of a multivariate distribution is unaffected by monotone increasing transformations of the original data.

Geometric constraints on copulas are given by the so-called Fréchet-Hoeffding bounds

$$W(\boldsymbol{u}) \leqslant C(\boldsymbol{u}) \leqslant M(\boldsymbol{u}).$$

The upper limit, given by $M(\boldsymbol{u}) = \min(u_1, \ldots, u_p)$ is associated to perfect positive dependence of the random variables $X_1, \ldots, X_p$, while the lower bound is equal to

$$W(\boldsymbol{u}) = \max\left(1 - p + \sum_{i=1}^{p} u_i, 0\right)$$

and does not correspond to a valid copula function except in dimension $p = 2$, where W is the copula of countermonotonic random variables. In particular, the independence copula $\Pi(\boldsymbol{u}) = u_1 \cdot \ldots \cdot u_p$ can be derived directly from Sklar's theorem.

1.1.2 Dependence measures

Most often, practitioners rely only on the linear correlation to describe the degree of dependence between two or more variables; an approach that can lead to quite misleading conclusions as this measure is only capable to capture linear relationships. In this section, we will present some copula-based concepts of dependence that are not suffering from this drawback.

Consider $\boldsymbol{X} = (X_1, X_2)$ and $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2)$ independent random vectors from F. Kendall's τ is defined as the probability of concordance minus the probability of discordance

$$\tau_K = \mathrm{P}((\tilde{X}_1 - X_1)\,(\tilde{X}_2 - X_2) > 0) - \mathrm{P}((\tilde{X}_1 - X_1)\,(\tilde{X}_2 - X_2) < 0).$$

In a more compact way, τ_K can be expressed as the expectation of the sign of $(\tilde{X}_1 - X_1)\,(\tilde{X}_2 - X_2)$,

$$\tau_K = \mathrm{E}\left(\mathrm{sign}((\tilde{X}_1 - X_1)\,(\tilde{X}_2 - X_2))\right).$$

To give a geometric interpretation of this measure, consider the observations (x_1, x_2) and $(\tilde{x}_1, \tilde{x}_2)$ from F. In the case of positive dependence, one can expect the line joining both points to be increasing, being equivalent to $(x_1 - \tilde{x}_1)\,(x_2 - \tilde{x}_2)$ being positive. The case of negative dependence will be characterized by lines with negative slopes. A positive Kendall's τ associates a higher probability to the event of two components of a random vector to move in the same direction than in opposite directions.

Spearman's ρ is defined as the correlation of the uniformized variables $F_1(X_1)$ and $F_2(X_2)$

$$\rho_S = \mathrm{Corr}(F_1(X_1), F_2(X_2)).$$

From the previous definitions, one can see that Kendall's τ_K and Spearman's ρ_S are invariant under strictly increasing componentwise transformations. Straightforward calculus permits to express τ_K and ρ_s as functionals of C:

$$\begin{aligned}
\tau_K &= 4 \iint_{[0,1]^2} C(u_1, u_2)\,\mathrm{d}C(u_1, u_2) - 1 \\
\rho_S &= 12 \iint_{[0,1]^2} C(u_1, u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2 - 3
\end{aligned}$$

For a proof see Theorem 5.1.3 and Theorem 5.1.6 in Nelsen [2006]. As for the linear correlation, the values of τ and ρ range from 1 in the case of perfect dependence to -1 for countermonotonic random variables.

For higher-dimensional vectors, Kendall's τ_K and Spearman's ρ_S can be considered pairwise. Alternatively, one possible extension to the higher-dimensional case of Kendall's τ could be

$$\tau = \frac{1}{2^{p-1}-1}\left[2^p \, \mathrm{E}[C(\boldsymbol{u})] - 1\right]. \tag{1.3}$$

We will use this measure in the simulation parts in the later chapters to compare copulas with similar degree of dependence. Further information on this topic can be found in Schmid et al. [2010].

Tail dependence relates to the amount of dependence in the extremes, either in the upper tail or in the lower tail. Still in the framework of bivariate data, the coefficient of upper tail dependence is defined as

$$\begin{aligned}\lambda_U &= \lim_{u \to 1^-} \mathrm{P}(X_2 > F_2^{\leftarrow}(u) | X_1 > F_1^{\leftarrow}(u)) \\ &= \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1-u}.\end{aligned}$$

In words, the coefficient of upper tail dependence consists in the probability that one variable will be larger than a certain treshold, given that the other variable has already exceeded a value with the same exceedance probability.

In the same way, the coefficient of lower tail dependence is defined as

$$\begin{aligned}\lambda_L &= \lim_{u \to 0^+} \mathrm{P}(X_2 < F_2^{\leftarrow}(u) | X_1 < F_1^{\leftarrow}(u)) \\ &= \lim_{u \to 0^+} \frac{C(u,u)}{u}.\end{aligned}$$

1.1.3 Parametric copula models

For the purpose of illustration, a restricted number of frequently used copulas will be listed below. The class of EV-Copulas being introduced with more details later, we will focus in a first time on other copulas.

According to expression (1.2) one can define the Gaussian copula as follows

$$C_{\Sigma}^{Ga}(\boldsymbol{u}) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)),$$

where Φ_Σ denotes the cumulative distribution function of the $p-$variate Gaussian distribution with zero mean and covariance matrix Σ and Φ the univariate cumulative distribution function of a standard Gaussian distribution. In the same spirit, the $p-$variate $t-$copula with ν degrees of freedom and covariance matrix Σ can be written

$$C^t_{\nu,\Sigma}(\boldsymbol{u}) = t_{\nu,\Sigma}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_p)),$$

with the definitions of $t_{\nu,\Sigma}$ and t_ν being obvious. Due to its tail dependence properties, the $t-$distribution enjoys some popularity in financial modelling; see for example Demarta and McNeil [2005]. The two previous copulas belong to the class of elliptical copulas presented in Fang et al. [1990].

In contrast to the previously mentioned copulas, Archimedean copulas can admit explicit and relatively simple expressions, which explains their success to some extent. A copula is Archimedean if it can be written as

$$C_\phi(u_1, \ldots, u_p) = \phi^{\leftarrow}\big(\phi(u_1) + \cdots + \phi(u_p)\big), \qquad (u_1, \ldots, u_p) \in [0,1]^p \tag{1.4}$$

with generator $\phi : [0,1] \to [0,\infty]$ and inverse $\phi^{\leftarrow}(t) = \inf\{u \in [0,1] : \phi(u) \leqslant t\}$; the function ϕ should be strictly decreasing and convex and satisfy $\phi(1) = 0$, and $\phi^{\leftarrow}$ should be p-monotone on $(0,\infty)$, see McNeil and Nešlehová [2009] and Genest et al. [2011].

To mention just two members of the class of Archimedean copulas, the Frank copula introduced in Frank [1979]

$$C^F(\boldsymbol{u}) = -\alpha^{-1} \log\left(1 + \frac{(\exp(-\alpha u_1) - 1) \cdots (\exp(-\alpha u_p) - 1)}{\exp(-\alpha) - 1}\right) \tag{1.5}$$

can be constructed from the generator function

$$\phi^F(t) = -\log\left(\frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1}\right), \quad \alpha \geqslant 0,$$

and the Clayton copula presented in Clayton [1978]

$$C^{Cl}(\boldsymbol{u}) = \left(1 - p + \sum_{i=1}^p u_i^{-\alpha}\right)^{-1/\alpha}, \quad \alpha > 0.$$

is associated to the generator function $\phi^{Cl}(t) = t^{-\alpha} - 1$.

Belonging to both the Archimedean and extreme-value copula class, the Gumbel-Hougaard copula will be more highlighted in Subsection 1.4.1.

1.2 Extreme-value copulas: Foundations

Let $\boldsymbol{X_i} = (X_{i,1}, \ldots, X_{i,p})$, $i \in \{1, \ldots, n\}$ be a sample of independent and identically distributed random vectors with common distribution function F, margins $F_1, \ldots, F_p$ and copula C. Consider the vector of componentwise maxima:

$$\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,p}), \qquad \text{where } M_{n,j} = \bigvee_{i=1}^{n} X_{i,j}, \tag{1.6}$$

with '$\vee$' denoting the maximum. Since the joint and marginal distribution functions of $\boldsymbol{M}_n$ are given by F^n and $F_1^n, \ldots, F_p^n$ respectively, it follows that the copula, C_n, of $\boldsymbol{M}_n$ is given by

$$C_n(u_1, \ldots, u_p) = C_F(u_1^{1/n}, \ldots, u_p^{1/n})^n, \qquad (u_1, \ldots, u_p) \in [0,1]^p.$$

The family of extreme-value copulas arises as the limits of these copulas C_n as the sample size n tends to infinity.

Definition 1.3. *A copula C is called an* extreme-value copula *if there exists a copula C_F such that*

$$C_F(u_1^{1/n}, \ldots, u_p^{1/n})^n \to C(u_1, \ldots, u_p) \qquad (n \to \infty) \tag{1.7}$$

for all $(u_1, \ldots, u_p) \in [0,1]^p$. The copula C_F is said to be in the domain of attraction *of C.*

Historically, this construction dates back at least to Deheuvels [1984] and Galambos [1978]. The representation of extreme-value copulas can be simplified using the concept of max-stability.

Definition 1.4. *A p-variate copula C is* max-stable *if it satisfies the relationship*

$$C(u_1, \ldots, u_p) = C(u_1^{1/m}, \ldots, u_p^{1/m})^m \tag{1.8}$$

for every integer $m \geqslant 1$ and all $(u_1, \ldots, u_p) \in [0,1]^p$.

From the previous definitions, it is trivial to see that a max-stable copula is in its own domain of attraction and thus must be itself an extreme-value copula. The converse is true as well.

Theorem 1.5. *A copula is an extreme-value copula if and only if it is max-stable.*

The proof of Theorem 1.5 is standard: for fixed integer $m \geqslant 1$ and for $n = mk$, write

$$C_n(u_1, \ldots, u_p) = C_k(u_1^{1/m}, \ldots, u_p^{1/m})^m.$$

Let k tend to infinity on both sides of the previous display to get (1.8).

By definition, the family of extreme-value copulas coincides with the set of copulas of *extreme-value distributions*, that is, the class of limit distributions with non-degenerate margins of

$$\left(\frac{M_{n,1} - b_{n,1}}{a_{n,1}}, \ldots, \frac{M_{n,p} - b_{n,p}}{a_{n,p}} \right)$$

with $M_{n,j}$ as in (1.6), centering constants $b_{n,j} \in \mathbb{R}$ and scaling constants $a_{n,j} > 0$. Representations of extreme-value distributions then yield representations of extreme-value copulas. Let $\Delta_{p-1} = \{(w_1, \ldots, w_p) \in [0,1]^p : \sum_j^p w_j = 1\}$ be the unit simplex in $\mathbb{R}^p$; see Figure 1.1. Alternatively, one can consider the unit-simplex as a subset of $\mathbb{R}^{p-1}$ as follows $\Delta_{p-1} = \{(w_1, \ldots, w_{p-1}) \in [0,1]^{p-1} : \sum_{j=0}^{p-1} w_j \leqslant 1\}$. Depending on the situation, we will consider Δ_{p-1} either as a subset of $\mathbb{R}^p$ or $\mathbb{R}^{p-1}$.

The following theorem is adapted from Pickands [1981], which is based in turn on de Haan and Resnick [1977].

Theorem 1.6. *A p-variate copula C is an extreme-value copula if and only if there exists a finite Borel measure H on Δ_{p-1}, called* spectral measure, *such that*

$$C(u_1, \ldots, u_p) = \exp\bigl(-\ell(-\log u_1, \ldots, -\log u_p)\bigr), \qquad (u_1, \ldots, u_p) \in (0,1]^p,$$

where the tail dependence function $\ell : [0,\infty)^p \to [0,\infty)$ *is given by*

$$\ell(x_1, \ldots, x_p) = \int_{\Delta_{p-1}} \bigvee_{j=1}^{p} (v_j x_j) \, dH(v_1, \ldots, v_p), \qquad (x_1, \ldots, x_p) \in [0,\infty)^p. \tag{1.9}$$

The spectral measure H is arbitrary except for the p moment constraints

$$\int_{\Delta_{p-1}} v_j \, dH(v_1, \ldots, v_p) = 1, \qquad j \in \{1, \ldots, p\}. \tag{1.10}$$

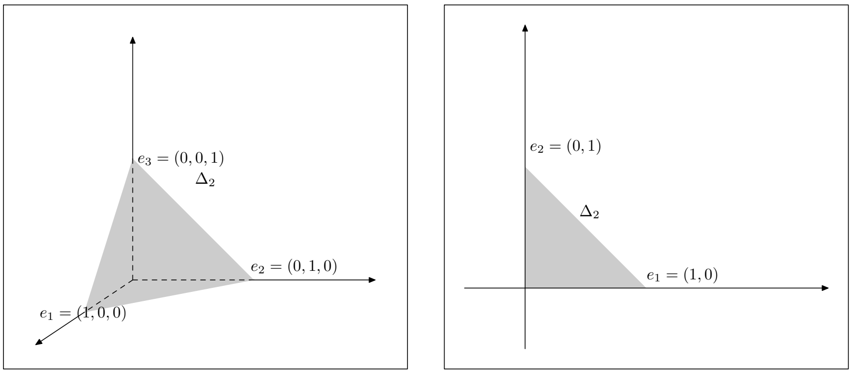


Figure 1.1: On the left-hand side Δ_2 is represented in $\mathbb{R}^3$, which is equivalent to the representation in $\mathbb{R}^2$ on the right side.

The p moment constraints on H in (1.10) stem from the requirement that the margins of C be standard uniform. They imply that $H(\Delta_{p-1}) = p$.

By a linear expansion of the logarithm and the exponential function, the domain-of-attraction equation (1.7) is equivalent to

$$\begin{aligned} \lim_{t \downarrow 0} t^{-1}\big(1 - C_F(1 - tx_1, \ldots, 1 - tx_p)\big) &= -\log C(e^{-x_1}, \ldots, e^{-x_p}) \\ &= \ell(x_1, \ldots, x_p) \qquad (1.11) \end{aligned}$$

for all $(x_1, \ldots, x_p) \in [0, \infty)^p$; see for instance Drees and Huang [1998]. The tail dependence function ℓ in (1.9) is convex, homogeneous of order one, that is $\ell(cx_1, \ldots, cx_p) = c\,\ell(x_1, \ldots, x_p)$ for $c > 0$, and satisfies $\max(x_1, \ldots, x_p) \leqslant \ell(x_1, \ldots, x_p) \leqslant x_1 + \cdots + x_p$ for all $(x_1, \ldots, x_p) \in [0, \infty)^p$. By homogeneity, it is characterized by the *Pickands dependence function* $A : \Delta_{p-1} \to [1/p, 1]$, which is simply the restriction of ℓ to the unit simplex:

$$\ell(x_1, \ldots, x_p) = (x_1 + \cdots + x_p)\, A(w_1, \ldots, w_{p-1}),$$

for $(x_1, \ldots, x_p) \in [0, \infty)^p \setminus \{0\}$ and $w_j = \frac{x_j}{x_1 + \cdots + x_p}$ so that

$$A(w_1, \ldots, w_{p-1}) = \int_{\Delta_{p-1}} \bigvee_{j=1}^{p} (v_j w_j)\, dH(v_1, \ldots, v_p), \qquad (w_1, \ldots, w_p) \in \Delta_{p-1}. \qquad (1.12)$$

The extreme-value copula C can be expressed in terms of A via

$$C(u_1,\dots,u_p)=\exp\left\{\left(\sum_{j=1}^{p}\log u_j\right)A\left(\frac{\log u_1}{\sum_{j=1}^{p}\log u_j},\dots,\frac{\log u_{p-1}}{\sum_{j=1}^{p}\log u_j}\right)\right\}. \tag{1.13}$$

The function A is convex as well and satisfies

$$\max(w_1,\dots,w_p)\leqslant A(w_1,\dots,w_{p-1})\leqslant 1, \forall(w_1,\dots,w_{p-1})\in\Delta_{p-1}. \tag{1.14}$$

However, these properties do not characterize the class of Pickands dependence functions unless $p=2$. To illustrate this, let us consider the counterexample on p. 257 in Beirlant et al. [2004] with

$$A(\boldsymbol{w})=(w_1+w_2)\vee(1-w_1)\vee(1-w_2), \quad \boldsymbol{w}\in\Delta_2,$$

corresponding to a convex function satisfying the inequalities in . Observe that $A(1/2,0)=A(0,1/2)=A(1/2,1/2)=1$ implying pairwise independence and hence full independence corresponding to $A(\boldsymbol{w})=1$ for all $\boldsymbol{w}\in\Delta_2$, which is however violated since $A(1/3,1/3)=2/3$. To justify the previous reasoning, remark that pairwise independence for extreme-value data results into a system of equations

$$\int_{\Delta_{p-1}}\Big\{(w_iv_i+w_jv_j)-(w_iv_i\vee w_jv_j)\Big\}H(\mathrm{d}\boldsymbol{v})=0,$$

with $\boldsymbol{v},\boldsymbol{v}\in\Delta_{p-1}$ for all $i\neq j\in\{1,\dots,p\}$. This system of equations can only be verified if the spectral measure concentrates all the mass on the vertices. For more details, check the developments in Beirlant et al. [2004].

In the bivariate case, we identify the unit simplex $\Delta_1=\{(1-t,t):t\in[0,1]\}$ in $\mathbb{R}^2$ with the interval $[0,1]$.

Theorem 1.7. *A bivariate copula C is an extreme-value copula if and only if*

$$C(u,v)=(uv)^{A(\log(v)/\log(uv))}, \qquad (u,v)\in(0,1]^2\setminus\{(1,1)\}, \tag{1.15}$$

where $A:[0,1]\to[1/2,1]$ is convex and satisfies $t\vee(1-t)\leqslant A(t)\leqslant 1$ for all $t\in[0,1]$.

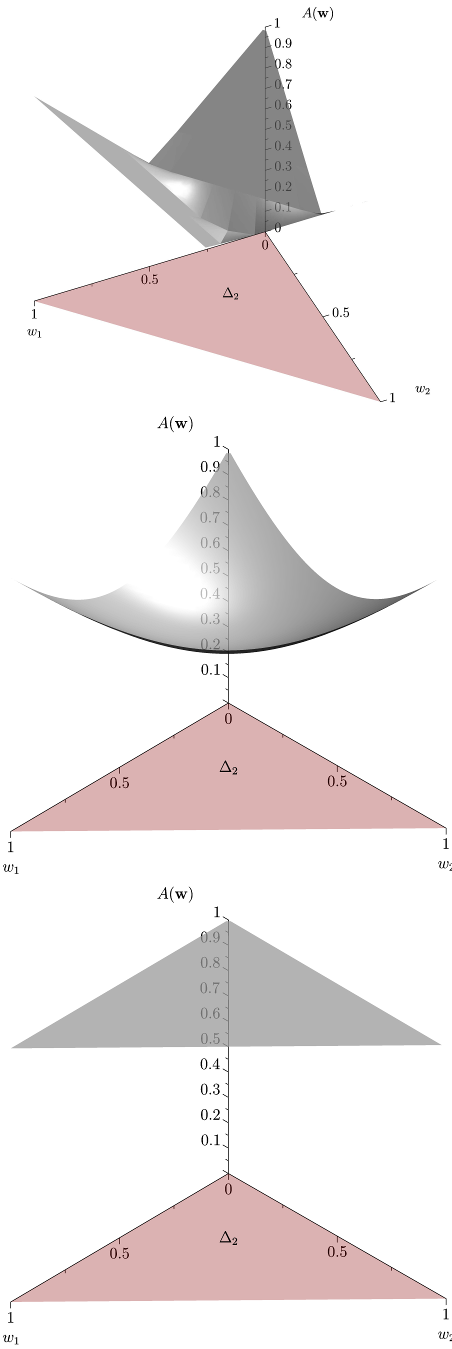


Table 1.1: Sample of Pickands dependence functions associated to three-dimensional extreme-value copulas. Top: Perfect dependence $A(\boldsymbol{w}) = \max(w_1, w_2, w_3)$. Middle: Logistic copula with parameter $\alpha = 0.5$ (see section 1.4.1). Bottom: Independence $A(\boldsymbol{w}) = 1$.

It is worth stressing that in the bivariate case, any function A satisfying the two constraints from Theorem 1.7 corresponds to an extreme-value copula. These functions lie in the shaded area of Figure 1.2; in particular, $A(0) = A(1) = 1$.

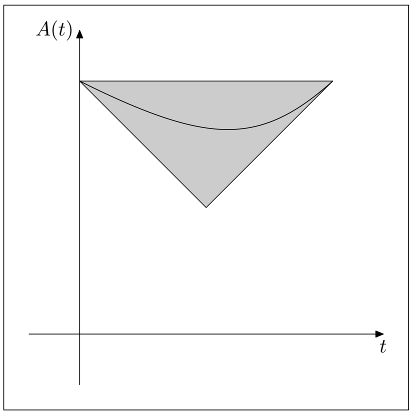


Figure 1.2: A typical Pickands dependence function A together with the region $t \vee (1-t) \leqslant A(t) \leqslant 1$ in Theorem 1.7.

The upper and lower bounds for A have special meanings: the upper bound $A(t) = 1$ corresponds to independence, $C(u,v) = uv$, whereas the lower bound $A(t) = t \vee (1-t)$ corresponds to perfect dependence (comonotonicity) $C(u,v) = u \wedge v$. In general, the inequality $A(t) \leqslant 1$ implies $C(u,v) \geqslant uv$, that is, extreme-value copulas are necessarily *positive quadrant dependent.*

1.3 Spectral representation of extreme-value copulas

For a non-empty vector of indices $I \subseteq \{1, 2, \ldots, p\}$, we will define the associated subset of the Δ_{p-1} as follows

$$\Delta_I^{(|I|-1)} = \{\boldsymbol{v} \in \Delta_{p-1} : v_i > 0 \text{ if } i \in I \quad \text{and} \quad v_i = 0 \text{ if } i \notin I\}. \tag{1.16}$$

In particular, for dimension $p = 3$ the $2^3 - 1$ lower-dimensional subsets of Δ_2 according to the definition (1.16) are listed below:

- the relative interior of the unit simplex

$$\Delta^{(2)}_{\{1,2,3\}} = \{\boldsymbol{v} \in \Delta_2 : \, v_i > 0 \, , \forall i \in \{1,2,3\}\},$$

- the one-dimensional borders

$$\begin{aligned}
\Delta^{(1)}_{\{1,2\}} &= \{\boldsymbol{v} \in \Delta_2 : v_3 = 0, v_1 > 0, v_2 > 0\} \\
\Delta^{(1)}_{\{3,1\}} &= \{\boldsymbol{v} \in \Delta_2 : v_2 = 0, v_1 > 0, v_3 > 0\} \\
\Delta^{(1)}_{\{2,3\}} &= \{\boldsymbol{v} \in \Delta_2 : v_1 = 0, v_2 > 0, v_3 > 0\},
\end{aligned}$$

- the zero-dimensional faces corresponding to the vertices

$$\begin{aligned}
\Delta^{(0)}_{\{1\}} &= \{\boldsymbol{v} \in \Delta_2 : v_2 = v_3 = 0\} = \{(1,0,0)\} \\
\Delta^{(0)}_{\{2\}} &= \{(0,1,0)\} \\
\Delta^{(0)}_{\{3\}} &= \{(0,0,1)\}.
\end{aligned}$$

Remark that the faces of dimension 1 can be parametrized by either of the two components that are nonzero. For example,

$$\Delta^{(1)}_{\{1,2\}} = \{(v_1, v_2, 0) : \, v_1 + v_2 = 1; v_1, v_2 > 0\}$$

can be identified with $\big\{(v_1, 1 - v_1) : \, v_1 \in (0,1)\big\}$.

Henceforth we will consider the case where A is absolutely continuous. In Coles and Tawn [1991] it was shown that the spectral measure H may have *spectral densities* on the interior of Δ_{p-1} and all of its lower-dimensional faces.

Taking into account this decomposition, a Pickands dependence function A in $p = 3$ is given by

$$\begin{aligned}
A(\boldsymbol{w}) &= \int_{\Delta^{(2)}_{\{1,2,3\}}} \max\big(w_1 v_1, w_2 v_2, w_3 v_3\big)\, h^{(2)}_{\{1,2,3\}}(v_1, v_2)\, \mathrm{d}v_1\, \mathrm{d}v_2 \\
&+ \int_0^1 \max\big(w_1 v_1, w_2\,(1 - v_1)\big)\, h^{(1)}_{\{1,2\}}(v_1)\, \mathrm{d}v_1 \\
&+ \int_0^1 \max\big(w_2 v_2, w_3\,(1 - v_2)\big)\, h^{(1)}_{\{2,3\}}(v_2)\, \mathrm{d}v_2 \\
&+ \int_0^1 \max\big(w_1\,(1 - v_3), w_3\, v_3\big)\, h^{(1)}_{\{3,1\}}(v_3)\, \mathrm{d}v_3 \\
&+ h^{(0)}_{\{1\}} w_1 + h^{(0)}_{\{2\}} w_2 + h^{(0)}_{\{3\}} w_3 \qquad (1.17)
\end{aligned}$$

with $\boldsymbol{w} \in \Delta_2$. For C to be a copula function, the spectral measure H has to be a positive measure satisfying the constraints given in (1.10), that can be made explicit for dimension $p = 3$

$$\int_{\Delta_2} v_1 \, H(\mathrm{d}\boldsymbol{v}) = \int_{\Delta^{(2)}_{\{1,2,3\}}} v_1 \, h^{(2)}_{\{1,2,3\}}(v_1, v_2) \, \mathrm{d}v_1 \, \mathrm{d}v_2$$
$$+ \int_{\Delta^{(1)}_{\{1,2\}}} v_1 h^{(1)}_{\{1,2\}}(v_1) \, \mathrm{d}v_1 + \int_{\Delta^{(2)}_{\{3,1\}}} v_1 \, h^{(1)}_{\{3,1\}}(1 - v_1) \, \mathrm{d}v_1 + h^{(0)}_{\{1\}} = 1$$

$$\int_{\Delta_2} v_2 \, H(\mathrm{d}\boldsymbol{v}) = \int_{\Delta^{(2)}_{\{1,2,3\}}} v_2 \, h^{(2)}_{\{1,2,3\}}(v_1, v_2) \, \mathrm{d}v_1 \, \mathrm{d}v_2$$
$$+ \int_{\Delta^{(2)}_{\{1,2\}}} (1 - v_1) \, h^{(1)}_{\{1,2\}}(v_1) \, \mathrm{d}v_1 + \int_{\Delta^{(2)}_{\{2,3\}}} v_2 \, h^{(1)}_{\{2,3\}}(v_2) \, \mathrm{d}v_2 + h^{(0)}_{\{2\}} = 1$$

$$\int_{\Delta_2} v_3 \, H(\mathrm{d}\boldsymbol{v}) = \int_{\Delta^{(2)}_{\{1,2,3\}}} (1 - v_1 - v_2) \, h^{(2)}_{\{1,2,3\}}(v_1, v_2) \, \mathrm{d}v_1 \, \mathrm{d}v_2$$
$$+ \int_{\Delta^{(1)}_{\{2,3\}}} (1 - v_2) \, h^{(1)}_{\{2,3\}}(v_2) \, \mathrm{d}v_2 + \int_{\Delta^{(2)}_{\{3,1\}}} v_3 \, h^{(2)}_{\{3,1\}}(v_3) \, \mathrm{d}v_3 + h^{(0)}_{\{3\}} = 1$$

To deduce the spectral densities from the tail dependence function ℓ, let us define

$$V(\boldsymbol{z}) = \ell\,(1/\boldsymbol{z})\,, \quad \boldsymbol{z} \in (0, \infty]^p. \tag{1.18}$$

Consider a subset

$$I = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, p\}, \quad 0 < m \leqslant p.$$

In Coles and Tawn [1991] it was proved that the spectral density defined on $\Delta_I^{(|I|-1)}$ can be computed by

$$h_I^{(m-1)}(v_{i_1}, \ldots, v_{i_{m-1}}) = \left(\sum_{i \in I} z_i \right)^{m+1} \lim_{\substack{z_j \to 0 \\ \forall j \in \{1, \ldots, p\} \setminus I}} \frac{\partial^m}{\partial z_{i_1} \ldots \partial z_{i_m}} V(\boldsymbol{z}), \tag{1.19}$$

where $v_{i_j} = z_{i_j} / \sum_{i \in I} z_i, \forall j \in \{1, \ldots, m\}$.

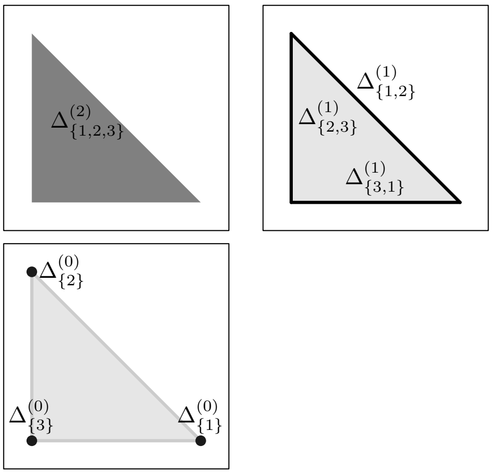


Table 1.2: The interior $\Delta^{(2)}_{\{1,2,3\}}$ (top left). The one-dimensional boundaries $\Delta^{(1)}_{\{1,2\}}$, $\Delta^{(1)}_{\{3,1\}}$ and $\Delta^{(1)}_{\{2,3\}}$ (top right). The vertices $\Delta^{(0)}_{\{1\}}$, $\Delta^{(0)}_{\{1\}}$ and $\Delta^{(0)}_{\{3\}}$ (bottom left).

Example 1.8. In particular for dimension $p = 3$, the spectral densities for the asymmetric logistic dependence function, used in our simulation studies, can be calculated by the means of formula (1.19). For $(v_1, v_2) \in \Delta^{(2)}_{\{1,2,3\}}$, the spectral density defined on the interior of the unit-simplex is given by

$$h^{(2)}_{\{1,2,3\}}(v_1, v_2) = \left(\psi \prod_{j=1}^{3}(j/\alpha - 1)\right)\left(\prod_{i=1}^{3} v_i^{-1-1/\alpha}\right)\left(\sum_{i=1}^{3} v_i^{-1/\alpha}\right)^{\alpha-3},$$

where $v_3 = 1-v_1-v_2$. For example, the one-dimensional spectral density associated to the subspace $\Delta^{(1)}_{\{1,2\}}$ can be calculated as follows

$$h^{(1)}_{\{1,2\}}(v_1) = \left(\frac{1-\alpha}{\alpha}\right)\theta^{1/\alpha}\phi^{1/\alpha}\left(\theta^{1/\alpha}v_1^{-1/\alpha} + \phi^{1/\alpha}(1-v_1)^{-1/\alpha}\right)^{\alpha-2} v_1^{-1-1/\alpha}(1-v_1)^{-1-1/\alpha},\ v_1 \in (0,1),$$

the two remaining densities admitting similar expressions by symmetry.

Finally, the same mass is distributed on the three vertices

$$h^{(0)}_{\{1\}} = h^{(0)}_{\{2\}} = h^{(0)}_{\{3\}} = 1 - \theta - \phi - \psi.$$

The logistic dependence function arises as the special case with parametrization $(\theta, \phi, \psi) = (0, 0, 1)$ and admits only a non-zero spectral density on $\Delta^{(2)}_{\{1,2\}}$.

Special cases appear for the case of independence with all the mass concentrated on the vertices and the case of perfect dependence degenerated to a Dirac delta function on the interior of the unit-simplex. □

1.4 Parametric models

By Theorems 1.6 and 1.7, the class of extreme-value copulas is infinite-dimensional. Parametric submodels can be constructed in a number of ways: by calculating the limit ℓ in (1.11) for a given initial copula C_F; by specifying a spectral measure H; in dimension $p = 2$, by constructing a Pickands dependence function A. In this section, we employ the first of these methods to introduce some of the more popular families. For more extensive overviews, see e.g. Beirlant et al. [2004], Kotz and Nadarajah [2000].

1.4.1 Logistic model or Gumbel–Hougaard copula

Consider again the definition of Archimedean copula with generator ϕ in (1.4). If the following limit exists,

$$\frac{1}{\alpha} = -\lim_{s \downarrow 0} \frac{s\,\phi'(1-s)}{\phi(1-s)} \in [1, \infty] \tag{1.20}$$

then the domain-of-attraction condition (1.11) is verified for C_F equal to C_ϕ, the tail dependence function being

$$\ell(x_1, \ldots, x_p) = \begin{cases} (x_1^{1/\alpha} + \cdots + x_p^{1/\alpha})^\alpha & \text{if } 0 < \alpha \leqslant 1 \;, \\ x_1 \vee \cdots \vee x_p & \text{if } \alpha = 0, \end{cases} \tag{1.21}$$

for $(x_1, \ldots, x_p) \in [0, \infty)^p$; see Capéraà et al. [2000], Charpentier and Segers [2009]. The associated Pickands dependence function A evaluated in $\boldsymbol{w} = (w_1, \ldots, w_{p-1}) \in \Delta_{p-1}$ can be deduced immediately

$$A(w_1, \ldots, w_{p-1}) = \begin{cases} (w_1^{1/\alpha} + \cdots + w_p^{1/\alpha})^\alpha & \text{if } 0 < \alpha \leqslant 1, \\ w_1 \vee \cdots \vee w_p & \text{if } \alpha = 0, \end{cases} \tag{1.22}$$

with $w_p = 1 - \sum_{j=1}^{p-1} w_j$. The range $[0, 1]$ for the parameter α in (1.20) is not an assumption but rather a consequence of the properties of ϕ. The parameter α measures the degree of dependence, ranging from independence ($\alpha = 1$) to complete dependence ($\alpha = 0$).

The extreme-value copula associated to ℓ in (1.21) is

$$C(u_1, \ldots, u_p) = \exp\bigl\{-\bigl((-\log u_1)^{1/\alpha} + \cdots + (-\log u_p)^{1/\alpha}\bigr)^\alpha\bigr\},$$

known as the *Gumbel–Hougaard* or *logistic* copula. Dating back to Gumbel [1960, 1961], it is (one of) the oldest multivariate extreme-value models. It was discovered independently in survival analysis in Crowder [1989] and Hougaard [1986]. It happens to be the only copula that is at the same time Archimedean and extreme-value as proved in Genest and Rivest [1989].

The bivariate asymmetric logistic model introduced in Tawn [1988] adds further flexibility to the basic logistic model. Multivariate extensions of the asymmetric logistic model were studied already in McFadden [1978] and later in Coles and Tawn [1991] and Joe [1994]. These distributions can be generated via mixtures of certain extreme-value distributions over stable distributions, a representation that yields large possibilities for modelling that have yet begun to be explored; see Fougères et al. [2009] and Toulemonde et al. [2009].

The multivariate asymmetric logistic copula function to which we will refer in this work was introduced in Tawn [1990]. Let B be the set of all nonempty subsets of $\{1, \ldots, p\}$, then the asymmetric logistic copula can be written as

$$C(\boldsymbol{u}) = \exp\left\{-\sum_{b \in B} \bigl(\sum_{i \in b} \theta_{i,b}^{1/\alpha_b} (-\log u_i)^{1/\alpha_b}\bigr)^{\alpha_b}\right\}, \tag{1.23}$$

with $\theta_{i,b} \geqslant 0$, $0 \leqslant \alpha_b \leqslant 1$ and $\sum_{\{b \in B : i \in b\}} \theta_{i,b} = 1$ for all $i \in \{1, \ldots, p\}$ and all $b \in B$.

In dimension $p = 3$, B contains all the subsets of $\{1, 2, 3\}$, i.e.

$$B = \Bigl\{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\Bigr\}.$$

Frequently, we will use the following version that is more parsimonious in the parameters

$$A(w_1, w_2) = (\theta^{1/\alpha} w_1^{1/\alpha} + \phi^{1/\alpha} w_2^{1/\alpha})^\alpha + (\theta^{1/\alpha} w_2^{1/\alpha} + \phi^{1/\alpha} w_3^{1/\alpha})^\alpha$$
$$+ (\theta^{1/\alpha} w_3^{1/\alpha} + \phi^{1/\alpha} w_1^{1/\alpha})^\alpha + \psi(w_1^{1/\alpha} + w_2^{1/\alpha} + w_3^{1/\alpha})^\alpha + 1 - \theta - \phi - \psi \tag{1.24}$$

with $\boldsymbol{w} \in \Delta_{p-1}$.

This model includes the logistic dependence function as particular case corresponding to parameter values $(\phi, \psi, \theta) = (0, 1, 0)$.

1.4.2 Negative logistic model or Galambos copula

Let $\hat{C}_\phi$ be the survival copula of the Archimedean copula C_ϕ in (1.4). Specifically, if C_ϕ is the distribution function of the random vector $(U_1, \ldots, U_d)$, then $\hat{C}_\phi$ is the distribution function of the random vector $(1 - U_1, \ldots, 1 - U_d)$. If the following limit exists,

$$\theta = -\lim_{s \downarrow 0} \frac{\phi(s)}{s\,\phi'(s)} \in [0, \infty] \tag{1.25}$$

then the domain-of-attraction condition (1.11) is verified for C_F equal to $\hat{C}_\phi$, the tail dependence function being

$$\ell(x_1, \ldots, x_p) = \begin{cases} x_1 + \cdots + x_p & \text{if } \theta = 0, \\ x_1 + \cdots + x_p - \displaystyle\sum_{\substack{I \subset \{1,\ldots,p\} \\ |I| \geqslant 2}} (-1)^{|I|} \left(\textstyle\sum_{i \in I} x_i^{-\theta}\right)^{-1/\theta} & \text{if } 0 < \theta < \infty, \\ x_1 \vee \cdots \vee x_p & \text{if } \theta = \infty, \end{cases}$$

for $(x_1, \ldots, x_p) \in [0, \infty)^p$; see Capéraà et al. [2000] and Charpentier and Segers [2009]. In case $0 < \theta < \infty$, the sum is over all subsets I of $\{1, \ldots, p\}$ of cardinality $|I|$ at least 2. The amount of dependence ranges from independence ($\theta = 0$) to complete dependence ($\theta = \infty$).

The resulting extreme-value copula is known as the *Galambos* or *negative logistic* copula, dating back to Galambos [1975]. Asymmetric extensions have been proposed in Joe [1990, 1994].

1.4.3 Hüsler–Reiss model

For the bivariate normal distribution with correlation coefficient ρ smaller than one, it is known since Sibuya [1960] that the marginal maxima $M_{n,1}$ and $M_{n,2}$ are asymptotically independent, that is, the domain-of-attraction condition (1.7) holds with limit copula $C(u,v) = uv$. However, for ρ close to one, better approximations to the copula of $M_{n,1}$ and $M_{n,2}$ arise within a somewhat different asymptotic framework. More precisely, as in Hüsler and Reiss [1989], consider the situation where the correlation coefficient ρ associated to the bivariate Gaussian copula C_ρ is allowed to change with the sample size, $\rho = \rho_n$, in such a way that $\rho_n \to 1$ as $n \to \infty$. If

$$(1 - \rho_n) \log n \to \lambda^2 \in [0, +\infty] \qquad (n \to \infty),$$

then one can show that

$$C_{\rho_n}(u^{1/n}, v^{1/n})^n \to C_A(u,v) \qquad (n \to \infty), \qquad (u,v) \in [0,1]^2,$$

where the *Hüsler–Reiss* copula C_A is the bivariate extreme-value copula with Pickands dependence function

$$A(w) = (1-w)\,\Phi\left(\lambda + \frac{1}{2\lambda} \log \frac{1-w}{w}\right) + w\,\Phi\left(\lambda + \frac{1}{2\lambda} \log \frac{w}{1-w}\right)$$

for $w \in [0,1]$, with Φ representing the standard normal cumulative distribution function. The parameter λ measures the degree of dependence, going from independence ($\lambda = \infty$) to complete dependence ($\lambda = 0$).

1.4.4 The t-EV copula

In financial applications, the t-copula is sometimes preferred over the Gaussian copula because of the larger weight it assigns to the tails. The bivariate t-copula with $\nu > 0$ degrees of freedom and correlation parameter $\rho \in (-1,1)$ is the copula of the bivariate t-distribution with the same parameters and is given by

$$C_{\nu,\rho}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{\pi\nu|P|^{1/2}} \frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\boldsymbol{x}' P^{-1} \boldsymbol{x}}{\nu}\right)^{-\nu/2+1} d\boldsymbol{x},$$

where t_ν represents the distribution function of the univariate t-distribution with ν degrees of freedom and P represents the 2×2 correlation matrix with off-diagonal element ρ. In Demarta and McNeil [2005], it is shown that $C_{\nu,\rho}$ is in the domain of attraction of the bivariate extreme-value copula C_A with Pickands dependence function

$$\begin{aligned} &A(w) = w\, t_{\nu+1}(z_w) + (1-w)\, t_{\nu+1}(z_{1-w}), \\ &\text{where } z_w = (1+\nu)^{1/2}[\{w/(1-w)\}^{1/\nu} - \rho](1-\rho^2)^{-1/2}, \qquad w \in [0,1]. \end{aligned} \tag{1.26}$$

This extreme-value copula was coined the *t-EV* copula. Building upon results in Ghoudi et al. [2005], Hashorva [2005], exactly the same extreme-value attractor is found in Asimit and Jones [2007] for the more general class of (meta-)elliptical distributions whose generator has a regularly varying tail.

1.5 Dependence properties of extreme-value copulas

Let (U, V) be a bivariate random vector with distribution function C, a bivariate extreme-value copula with Pickands dependence function A as in (1.15). As mentioned already, the inequality $A \leqslant 1$ implies that $C(u,v) \geqslant uv$ for all $(u,v) \in [0,1]^2$, that is, C is positive quadrant dependent. In fact, in Guillem [2000] it was shown that extreme-value copulas are *monotone regression dependent*, that is, the conditional distribution of U given $V = v$ is stochastically increasing in v and *vice versa*; see also Theorem 5.2.10 in Resnick [1987].

All measures of dependence of C such as Kendall's τ_K or Spearman's ρ_S can be expressed in terms of A via

$$\begin{aligned} \tau_K &= 4 \iint_{[0,1]^2} C(u,v)\, \mathrm{d}C(u,v) - 1 = \int_0^1 \frac{t(1-t)}{A(t)}\, \mathrm{d}A'(t), \\ \rho_S &= 12 \iint_{[0,1]^2} uv\, \mathrm{d}C(u,v) - 3 = 12 \int_0^1 \frac{1}{(1+A(t))^2}\, \mathrm{d}t - 3. \end{aligned}$$

The Stieltjes integrator $\mathrm{d}A'(t)$ is well-defined since A is a convex function on $[0,1]$; if the dependence function A is twice differentiable, it can be replaced by $A''(t)\,\mathrm{d}t$. For a proof of the identities above, see for instance Hürlimann [2003], where it is shown that τ and ρ_S satisfy

$-1 + \sqrt{1 + 3\tau_K} \leqslant \rho_S \leqslant \min\left(\frac{3}{2}\tau_K, 2\tau_K - \tau_K^2\right)$, a pair of inequalities first conjectured in Hutchinson and Lai [1990].

The *Kendall distribution function* associated to a general bivariate copula C is defined as the distribution function of the random variable $C(U, V)$, that is,

$$K(w) = \mathrm{P}[C(U, V) \leqslant w], \qquad w \in [0, 1].$$

The reference to Kendall stems from the link with Kendall's τ, which is given by $\tau = 4\,\mathrm{E}[C(U, V)] - 1$. For bivariate Archimedean copulas, for instance, the function K not only identifies the copula Genest and Rivest [1993], convergence of Archimedean copulas is actually equivalent to weak convergence of their Kendall distribution functions Charpentier and Segers [2008]. For bivariate extreme-value copulas, the function K takes the remarkably simple form

$$K(w) = w - (1 - \tau)\, w \log w, \qquad w \in [0, 1], \tag{1.27}$$

as shown in Ghoudi et al. [1998]. In fact, in that paper the conjecture was formulated that if the Kendall distribution function of a bivariate copula is given by (1.27), then C is a bivariate extreme-value copula, a conjecture which to the best of our knowledge still stands. In the same paper, equation (1.27) was used to formulate a test that a copula belongs to the family of extreme-value copulas; see also Ghorbal et al. [2009].

In the context of extremes, it is natural to study the *coefficient of upper tail dependence*. For a bivariate copula C_F in the domain of attraction of an extreme-value copula with tail dependence function ℓ and Pickands dependence function A, we find

$$\begin{aligned}\lambda_U &= \lim_{u \uparrow 1} \mathrm{P}(U > u \mid V > u) = \lim_{t \downarrow 0} t^{-1}\bigl(2t - 1 + C(1 - t, 1 - t)\bigr) \\ &= 2 - \ell(1, 1) = 2\,\bigl(1 - A(1/2)\bigr) \in [0, 1].\end{aligned}$$

Graphically this quantity can be represented as the length between the upper boundary and the curve of the Pickands dependence function evaluated in the mid-point 1/2, see Figure 1.3. The coefficient λ_U ranges from 0 ($A = 1$, independence) to 1 (complete dependence). Multivariate extensions are proposed in Li [2009].

The related quantity $\ell(1, 1) = 2\,A(1/2)$ is called the *extremal coefficient* in Smith et al. [1990]. For a bivariate extreme-value copula, we

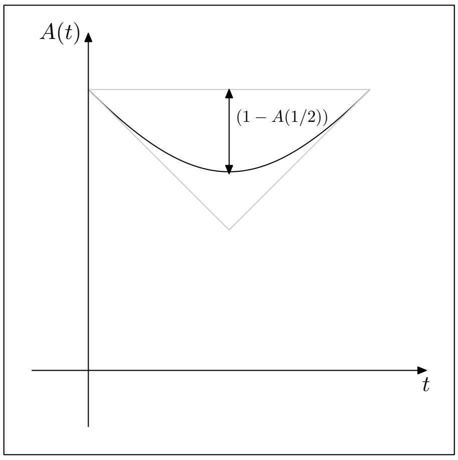


Figure 1.3: The coefficient of upper tail dependence λ_U is equal to twice the length of the double arrow in the upper part of the graph.

find

$$\mathrm{P}(U \leqslant u, V \leqslant u) = u^{2\,A(1/2)}, \qquad u \in [0,1],$$

so that $2\,A(1/2) \in [1,2]$ can be thought of as the (fractional) number of independent components in the copula. Multivariate extensions have been studied in Schlather and Tawn [2002].

For the *lower tail dependence coefficient*, the situation is trivial:

$$\lambda_L = \lim_{u \downarrow 0} \mathrm{P}(U \leqslant u \mid V \leqslant u) = \lim_{u \downarrow 0} u^{(2A(1/2)-1)} = \begin{cases} 0 & \text{if } A(1/2) > 1/2, \\ 1 & \text{if } A(1/2) = 1/2. \end{cases}$$

In words, except for the case of perfect dependence, $A(1/2) = 1/2$, extreme-value copulas have asymptotically independent lower tails.

1.6 Further reading

About the first monograph to treat multivariate extreme-value dependence is the one by Galambos [1978], with a major update in the second edition Galambos [1987]. Extreme-value copulas are treated extensively in the monographs Beirlant et al. [2004], Kotz and Nadarajah [2000] and briefly in the 2006 edition of Nelsen [2006]. The regular-variation approach to multivariate extremes is emphasized in the books Resnick [1987, 2007] and de Haan and Ferreira [2006]. A highly readable introduction to extreme-value analysis is the book by Coles [2001].

The first representations of bivariate extreme-value distributions are due to Finkelstein [1953], Tiago de Oliveira [1958], Geffroy [1958, 1959] and Sibuya [1960]. Incidentally, the 1959 paper by Geffroy appeared in the same issue as the famous paper by Sklar [1959]. The equivalence of all these representations was shown in Gumbel [1962]; see also the more recent paper by Obretenov [1991]. However, their representations of multivariate extreme value distributions have not enjoyed the same success as the one proposed by Pickands [1981]. The domain-of-attraction condition seems to have been formulated for the first time by Berman [1962], his standardization being to the standard exponential distribution rather than the uniform one.

A particular class of extreme-value copulas arises if the spectral measure H in Theorem 1.6 is discrete. In that case, the stable tail dependence function ℓ and the Pickands dependence function A are piecewise linear. In general, such distributions arise from max-linear combinations of independent random variables; see Schlather and Tawn [2002]. An early example of such a distribution is the bivariate model studied in Tiago de Oliveira [1974, 1980, 1989], which has a spectral measure with exactly two atoms; see also Einmahl et al. [2008]. The bivariate distribution of Marshall and Olkin [1967] corresponds to a spectral measure with exactly three atoms, $\{0, 1/2, 1\}$ (after identification of the unit simplex in $\mathbb{R}^2$ with the unit interval); see Mai and Scherer [2009] for a multivariate extension.

2 Inference for extreme-value copulas

The aim of this chapter is to give an overview of the state-of-the-art of parametric and nonparametric inference methods for extreme-value copulas. Inference on extreme-value copulas usually proceeds via its Pickands dependence function A defined in (1.12), a convex function on the unit-simplex satisfying certain inequality constraints, such that a vast literature has been dedicated to the development of nonparametric estimators. A selection of them is listed in Section 2.1 and 2.2 ranging from the classical moment estimators dating back to Pickands [1981] for the earliest publication in this domain to the more recent approaches based on weighted distance criteria. In Section 2.3, we will review different methods to transform initial estimates of A into valid Pickands dependence functions. Some approaches in the domain of parametric estimation are treated in Section of 2.4.

This chapter refers to the second part of the survey paper Gudendorf and Segers [2010] with updates where necessary.

2.1 Moment estimators

Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,p})$, $i \in \{1, \ldots, n\}$, be n independent random vectors from a $p-$variate distribution F and associated extreme-value copula C with Pickands dependence function A. Assume for the moment that the marginal distribution functions $F_1 \ldots F_p$ are known. Put $U_{i,j} = F_j(X_i)$ and $Y_{i,j} = -\log U_{i,j}$ for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$. Note that $Y_{i,j}$ are standard exponential random variables. The class of Pickands dependence functions is infinite-dimensional, a fact that justifies the use of nonparametric methods.

2.1.1 Pickands estimator

The first estimator presented in the context of extreme-value copulas was already given in Pickands [1981].

For $\boldsymbol{w} \in \Delta_{p-1}$, define

$$\xi_i(\boldsymbol{w}) = \bigwedge_{j=1}^{p} \frac{Y_{i,j}}{w_j}. \tag{2.1}$$

with the obvious conventions for division by zero. A characterizing property of extreme-value copulas is that the distribution of $\xi_i(\boldsymbol{w})$ is exponential as well, now with mean $1/A(\boldsymbol{w})$. For $x > 0$,

$$\mathrm{P}[\xi_i(\boldsymbol{w}) > x] = C(e^{-w_1 x}, \ldots, e^{-w_p x}) = e^{-x\, A(\boldsymbol{w})}. \tag{2.2}$$

giving rise to the original estimator presented in Pickands [1981]:

$$\frac{1}{\hat{A}^P(\boldsymbol{w})} = \frac{1}{n} \sum_{i=1}^{n} \xi_i(\boldsymbol{w}). \tag{2.3}$$

A major drawback of this estimator is that it does not verify any of the constraints imposed on the family of the Pickands dependence functions. Besides establishing the asymptotic properties of the original Pickands estimator, the main achievement in Deheuvels [1991] consists in the linear adjustments of the initial Pickands estimator imposing the endpoint constraints $A(\boldsymbol{e_j}) = 1$ for $j \in \{1, \ldots, p\}$. In detail, given continuous functions $\lambda_1, \ldots, \lambda_p : \Delta_{p-1} \to \mathbb{R}$ verifying $\lambda_j(\boldsymbol{e}_k) = \delta_{jk}$ (Kronecker delta) for all $j, k \in \{1, \ldots, p\}$, the modified estimator is defined as

$$\frac{1}{\hat{A}^{\mathrm{P}}_{\boldsymbol{\lambda},n}(\boldsymbol{w})} = \frac{1}{\hat{A}^{\mathrm{P}}_n(\boldsymbol{w})} - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \left(\frac{1}{\hat{A}^{\mathrm{P}}_n(\boldsymbol{e}_j)} - 1 \right). \tag{2.4}$$

The weight functions, as suggested in Zhang et al. [2008], are $\lambda_j(\boldsymbol{w}) = w_j$ for $j \in \{1, \ldots, p\}$ and can be understood as pragmatic choices. Optimal weight functions leading to minimum variance estimators in the case of known margins were derived in Gudendorf and Segers [2011a]. We will come back to this topic in Chapter 3.

2.1.2 Estimator of Capéraà-Fougéres-Genest

A different starting point was chosen in Capéraà et al. [1997] as they showed that in the two-dimensional case the distribution function of the random variable $Z_i = \log(U_{i,2})/\log(U_{i,1}\,U_{i,2})$ is given by

$$\mathrm{P}(Z_i \leqslant z) = z + z(1-z)\,\frac{A'(z)}{A(z)}, \qquad 0 \leqslant z < 1,$$

where A' denotes the right-hand derivative of A. Solving the resulting differential equation for A and replacing unknown quantities by their sample versions yields the CFG-estimator. This approach was generalized to higher dimensions in Zhang et al. [2008]. For a fixed value $j \in \{1,\dots,p\}$, define the n i.i.d. random variables

$$Z_{i,j} = \frac{\bigwedge_{l\neq j} \frac{Y_{i,l}}{w_l}}{\frac{Y_{i,j}}{1-w_j} + \bigwedge_{l\neq j} \frac{Y_{i,l}}{w_l}}, \quad i \in \{1,\dots,n\} \tag{2.5}$$

with distribution function given by

$$H_j(z) = P(Z_{i,j} \leqslant z) = z + z(1-z) \\ \frac{\partial}{\partial z} \log A\left(\frac{zw_1}{1-w_j},\dots,\frac{zw_{j-1}}{1-w_j}, 1-z, \frac{zw_{j+1}}{1-w_j},\dots,\frac{zw_{p-1}}{1-w_j}\right).$$

The previous differential equation can be solved

$$\log A_j(\boldsymbol{w}) = \int_0^{1-w_j} \frac{H_j(z)-z}{z(1-z)} \mathrm{d}z.$$

This motivates the construction of the estimator in Zhang et al. [2008]

$$\log \hat{A}_n^{\mathrm{ZWP}}(\boldsymbol{w}) = \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \int_0^{1-w_j} \frac{n^{-1}\sum_{i=1}^n \mathbf{1}\{Z_{ij}(\boldsymbol{w}) \leqslant z\} - z}{z(1-z)} \,\mathrm{d}z \tag{2.6}$$

with weight functions λ_j satisfying

$$\lambda_j(\boldsymbol{w}) \geqslant 0,\ \forall j \in \{1,\dots,p\} \quad \text{and} \quad \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) = 1, \quad \boldsymbol{w} \in \Delta_{p-1}. \tag{2.7}$$

Starting from a different approach, one can find back the same estimator as given in (2.6). Therefore, define

$$\log \hat{A}_n(\boldsymbol{w}) = -\frac{1}{n}\sum_{i=1}^{n} \log \xi_i(\boldsymbol{w}) - \gamma. \tag{2.8}$$

justified by the relationship between exponential and Gumbel distributions as

$$\mathrm{E}[-\log \xi_i(\boldsymbol{w})] = \log A(\boldsymbol{w}) + \gamma, \qquad \boldsymbol{w} \in \Delta_{p-1},$$

the Euler–Mascheroni constant $\gamma = 0.5772\ldots$ being equal to the mean of the standard Gumbel distribution. The endpoint-corrected version admits the following expression

$$\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \log \hat{A}_n(\boldsymbol{e}_j). \tag{2.9}$$

Both estimators $\hat{A}_n^{\mathrm{CFG}}$ and $\hat{A}_n^{\mathrm{ZWP}}$ are equivalent as long as the assumptions given in (2.7) on the weight functions λ_j are valid. This will be shown in the next chapter.

2.1.3 Hall-Tajvidi estimator

Hall and Tajvidi [2000] followed another approach to improve the small-sample properties of the Pickands estimator at the boundary points. For all $\boldsymbol{w} \in \Delta_{p-1}$ and $i \in \{1,\ldots,n\}$, define

$$\bar{\xi}_i(\boldsymbol{w}) = \bigwedge_{j=1}^{p} \frac{Y_{i,j}/\bar{Y}_{\cdot,j}}{w_j}$$

with $\bar{Y}_{\cdot,j} = \frac{1}{n}\sum_{i=1}^{n} Y_{i,j}$. The estimator presented in Hall and Tajvidi [2000] is given by

$$\frac{1}{\hat{A}^{HT}(\boldsymbol{w})} = \frac{1}{n}\sum_{i=1}^{n} \bar{\xi}_i(\boldsymbol{w}).$$

This construction ensures that the endpoint conditions are verified, i.e. $\hat{A}^{HT}(\boldsymbol{e}_j) = 1, \forall j \in \{1,\ldots,p\}$. In addition the new estimator verifies the lower boundary constraint as $\hat{A}^{HT}(\boldsymbol{w}) \geqslant \vee w_j$, $\forall \boldsymbol{w} \in \Delta_{p-1}$.

2.2 Minimum distance estimators

2.2.1 Estimator of Bücher, Dette and Volgushev

For dimension $p = 2$, a new estimator is presented in Bücher et al. [2011] minimizing the following criterion

$$M_\lambda(\tilde{C}_n, A) = \int_{(0,1)^2} \Big(\log \tilde{C}_n(y^{1-t}, y^t) - \log(y)\, A(t)\Big)^2\, \lambda(y)\, \mathrm{d}y\, \mathrm{d}t$$

over the set of all functions $A : [0,1] \to [1/2, 1]$. The minimum is attained for

$$\hat{A}_{n,\lambda}(t) = \frac{1}{\int_0^1 \lambda^*(y)\, \mathrm{d}y} \int_0^1 \frac{\log \tilde{C}_n(y^{1-t}, y^t)}{\log y}\, \lambda^*(y)\, \mathrm{d}y$$

with $\lambda^*(y) = \log(y)\, \lambda(y)$. For technical reasons, the authors preferred to work with a slightly modified version $\tilde{C}_n$ of $\hat{C}_n$.

Although the estimator is constrained to lie in the shadowed area in Figure 1.2, i.e. $t \vee (1-t) \leqslant \hat{A}_{n,\lambda}(t) \leqslant 1$, the convexity constraints are not necessarily fulfilled, such that the methodology presented in Fils-Villetard et al. [2008] can still be applied.

Simulation studies in Bücher et al. [2011] suggest that their new estimator performs slightly better than the estimators presented earlier in this chapter. Nevertheless, optimal weight functions can only be derived in theory such that in real-world applications one has to rely on suggestions and experiences shared by the authors. Using the same methodology, the authors were able to derive another test for extreme-value dependence.

2.2.2 Estimator of Peng, Qian and Yang.

A similar approach has been chosen in Peng et al. [2011], who prefer however to minimize for every fixed value of $t \in [0,1]$ the following criterion with respect to α

$$\int_0^1 \{\hat{C}_n(u^{1-t}, u^t) - u^\alpha\}^2 \bar{\lambda}(u,t)\, \mathrm{d}u, \quad t \in [0,1]. \tag{2.10}$$

The optimal value can be found by differentiating with respect to α in (2.10) to yield the following non-linear equation in α

$$\int_0^1 \{\hat{C}_n(u^{1-t}, u^t) - u^\alpha\}\lambda(u,t)\, \mathrm{d}u = 0$$

with a newly defined weight function λ. The Pickands and the CFG-estimator arise as special cases for the weight functions $\lambda(u,t) = u^{-1}$ and $\lambda(u,t) = \{u \log(u)\}^{-1}$. The above presented method in combination with the empirical likelihood methodology allows to create pointwise confidence intervals for their estimates of A with good empirical coverage rates.

None of the estimators figuring in this section has yet been extended to the higher-dimensional case.

2.3 Enforcing the shape constraints

The previous estimators do typically not fulfill the *shape constraints* on A as given in Theorem 1.7. In the bivariate case, a natural way to enforce these constraints is by modifying a pilot estimate $\hat{A}$ into the convex minorant of $\big(\hat{A}(t) \vee (1-t) \vee t\big) \wedge 1$, see Deheuvels [1991], Jiménez et al. [2001a], Pickands [1981]. It can be shown that this transformation cannot cause the L^∞ error of the estimator to increase.

A different way to impose the shape constraints is by constrained spline smoothing Abdous and Ghoudi [2005], Hall and Tajvidi [2000] or by constrained kernel estimation of the derivative of A Smith et al. [1990].

The L^2-viewpoint was chosen in Fils-Villetard et al. [2008]. The set $\mathscr{A}$ of Pickands dependence functions being a closed and convex subset of the space $L^2([0,1], \mathrm{d}x)$, it is possible to find for a pilot estimate $\hat{A}$ a Pickands dependence function $A \in \mathscr{A}$ that minimizes the L^2-distance $\int_0^1 (\hat{A} - A)^2$. By general properties of orthogonal projections, the L^2-error of the projected estimator cannot increase. This last approach will be extended to the higher-dimensional case in Chapter 5.

2.4 Parametric estimation

Assume that the extreme-value copula C belongs to a parametric family $(C_\theta : \theta \in \Theta)$ with $\Theta \subset \mathbb{R}^d$ with d the dimension of the parameter; for instance, one of the families described in Section 1.4. Inference on C then reduces to inference on the parameter vector θ. The usual way to proceed is by maximum likelihood. The likelihood is to be constructed from the copula density

$$c_\theta(u_1, \ldots, u_p) = \frac{\partial^p}{\partial u_1 \cdots \partial u_p} C_\theta(u_1, \ldots, u_p), \qquad (u_1, \ldots, u_p) \in (0,1)^p.$$

In order for this density to exist and to be continuous, the spectral measure H should be absolute continuous with continuous Radon–Nikodym derivative on all $2^p - 1$ faces of the unit simplex with respect to the Hausdorff measure of the appropriate dimension; see Coles and Tawn [1991]. In dimension $p = 2$, the Pickands dependence function $A : [0, 1] \to [1/2, 1]$ should be twice continuously differentiable on $(0, 1)$, or equivalently, the spectral measure H should have a continuous density on $(0, 1)$ (after identification of the unit simplex in $\mathbb{R}^2$ with the unit interval).

In case the margins are unknown, they may be estimated by the (properly rescaled) empirical distribution functions

$$\hat{F}_{n,j}(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{1}(X_{i,j} \leqslant x), \qquad x \in \mathbb{R},\ j \in \{1, \ldots, p\}. \tag{2.11}$$

(The denominator is $n + 1$ rather than n in order to avoid boundary effects in the pseudo-loglikelihood below.) Estimation of θ then proceeds by maximizing the pseudo-loglikelihood

$$\sum_{i=1}^n \log c_\theta\big(\hat{F}_{n,1}(X_{i,1}), \ldots, \hat{F}_{n,p}(X_{i,p})\big),$$

see Genest et al. [1995]. The resulting estimator is consistent and asymptotically normal, and its asymptotic variance can be estimated consistently.

If the margins are modelled parametrically as well, a fully parametric model for the joint distribution F arises, and the parameter vector of F may be estimated by ordinary maximum likelihood. An explicit expression for the 5×5 Fisher information matrix for the bivariate distribution with Weibull margins and Gumbel copula is calculated in Oakes and Manatunga [1992]. A multivariate extension with arbitrary generalized extreme value margins is presented in Shi [1995].

Special attention to the boundary case of independence is given in Tawn [1988]. In this case, the dependence parameter lies on the boundary of the parameter set and the Fisher information matrix is singular, implying that the normal assumptions for validity of the likelihood method are no longer valid.

A robustified version of the maximum likelihood estimator is introduced in Dupuis and Morgenthaler [2002]. The effects of misspecification of the dependence structure are studied in Dupuis and Tawn [2001].

2.5 Further reading

Even more challenging than the estimation problem considered in the context of this chapter is when the random sample comes from a distribution which is merely in the domain of attraction of a multivariate extreme-value distribution. See for instance Boldi and Davison [2007], Coles and Tawn [1991], de Haan et al. [2008], Einmahl et al. [2008], Joe et al. [1992], Klüppelberg et al. [2007, 2008], Ledford and Tawn [1996] for some (semi-)parametric approaches and Abdous and Ghoudi [2005], Capéraà and Fougères [2000], Einmahl et al. [2001, 2006], Einmahl and Segers [2009], Schmidt and Stadtmüller [2006] for some nonparametric ones.

For an overview of software related to extreme value analysis, see Stephenson and Gilleland [2005]. Particularly useful are the `R` packages `evd` presented in Stephenson [2002], which provides algorithms for the computation, simulation developed in Stephenson [2003] and estimation of certain univariate and multivariate extreme-value distributions, as well as the more general `copula` package Yan [2007].

3 Inference: Known margins

The multivariate extension of the CFG-estimator was presented in Zhang et al. [2008] and was reproduced in equation (2.6) in Chapter 2. This estimator was shown to be uniformly consistent and pointwise asymptotically normal. However, the joint asymptotic distribution of the estimator as a random function on Δ_{p-1} was not provided. Moreover, implementation of the estimator requires the choice of p weight functions λ_j on Δ_{p-1}, the issue of their optimal selection being left unresolved. Using a simplified representation of the above-mentioned estimator, we are able to uncover its asymptotic distribution in the space $\mathscr{C}(\Delta_{p-1})$ of continuous, real-valued functions on Δ_{p-1}.

Moreover, we give explicit expressions for the weight functions λ_j that minimize the pointwise asymptotic variance of the estimator. These optimal weight functions depend on the unknown distribution. We show that the CFG-estimator with estimated variance-minimizing weight functions can be implemented as the intercept estimator in a certain linear regression model via ordinary least squares. The OLS-estimator is data-adaptive in the sense that the asymptotic distribution is the same as if the optimal weight functions were used. In a simulation study, the gain in efficiency is shown to be quite sizeable.

As in Zhang et al. [2008], the setting here is that of a random sample from a distribution whose margins are known and whose copula is an extreme-value copula. The results from this chapter have already been published in Gudendorf and Segers [2011a].

3.1 CFG-estimator and variants

Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,p})$, $i \in \{1, \ldots, n\}$, be i.i.d. random vectors from a p-variate, continuous distribution function F with multivariate extreme-value copula C and Pickands dependence function A.

To begin, we show that the ZWP-estimator introduced in his original form (2.6) can be reexpressed in a simplified way as already stated in

Chapter 2.

Lemma 3.1. *If the constraints on the weight functions given in* (2.7) *are valid, then:*

$$\log \hat{A}_n^{ZWP}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \log \hat{A}_n(\boldsymbol{e}_j) \tag{3.1}$$

Hence, we will drop the superscript ZWP from now on and refer to the estimator in 3.1 as the CFG-estimator. The proof of Lemma 3.1 is essentially the same as the one in Segers [2007] for the bivariate case.

Proposition 3.2 (Naive estimator)**.** *The naive estimator $\hat{A}_n$ in* (2.8) *satisfies*

$$\sup_{\boldsymbol{w}\in\Delta_{p-1}} |\hat{A}_n(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely,} \tag{3.2}$$

and in $\mathscr{C}(\Delta_{p-1})$,

$$\sqrt{n}(\hat{A}_n - A) \rightsquigarrow A\,\zeta, \qquad n \to \infty, \tag{3.3}$$

where ζ is a centered Gaussian process with covariance function

$$\operatorname{cov}\bigl(\zeta(\boldsymbol{v}), \zeta(\boldsymbol{w})\bigr) = \operatorname{cov}\bigl(-\log \xi_i(\boldsymbol{v}),\, -\log \xi_i(\boldsymbol{w})\bigr), \qquad \boldsymbol{v}, \boldsymbol{w} \in \Delta_{p-1}, \tag{3.4}$$

with $\xi_i(\,\cdot\,)$ as in (2.1).

There is no reason for the weight functions $\lambda_j(\cdot)$ to still be positive and sum up to 1 in expression (3.1). The asymptotics of the naive estimator and the CFG-estimator follow from standard empirical process theory as presented for instance in van der Vaart and Wellner [1996] and van der Vaart [1998].

Theorem 3.3 (CFG-estimator)**.** *If, in addition to the assumptions in Proposition 3.2, the functions $\lambda_1, \ldots, \lambda_p : \Delta_{p-1} \to \mathbb{R}$ are continuous, then*

$$\sup_{\boldsymbol{w}\in\Delta_{p-1}} |\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely,} \tag{3.5}$$

and in $\mathscr{C}(\Delta_{p-1})$,

$$\sqrt{n}(\hat{A}_n^{\mathrm{CFG}} - A) \rightsquigarrow A\,\eta, \qquad n \to \infty, \tag{3.6}$$

where η is a centered Gaussian process defined by

$$\eta(\boldsymbol{w}) = \zeta(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w})\,\zeta(\boldsymbol{e}_j), \qquad \boldsymbol{w} \in \Delta_{p-1}, \tag{3.7}$$

with ζ as in Proposition 3.2.

Remark 3.4 (Covariance function)**.** *The covariance function* (3.4) *can be expressed in terms of A as follows. An application of the identity* $\log(x) = \int_0^\infty \{\mathbf{1}(s \leqslant x) - \mathbf{1}(s \leqslant 1)\}\, s^{-1}\, ds$ *for $x \in (0, \infty)$ yields, by Fubini's theorem,*

$$\begin{aligned}
&\operatorname{cov}\bigl(-\log \xi_i(\boldsymbol{v}),\, -\log \xi_i(\boldsymbol{w})\bigr) \\
&= \int_0^\infty \int_0^\infty \Bigl(P\bigl(\xi_i(\boldsymbol{v}) \geqslant s,\, \xi_i(\boldsymbol{w}) \geqslant t\bigr) - P\bigl(\xi_i(\boldsymbol{v}) \geqslant s\bigr)\, P\bigl(\xi_i(\boldsymbol{w}) \geqslant t\bigr)\Bigr) \frac{ds}{s}\,\frac{dt}{t} \\
&= \int_0^\infty \int_0^\infty [\exp\{-\ell((w_1 s) \vee (v_1 t), \ldots, (w_p s) \vee (v_p t))\} \\
&\qquad\qquad - \exp\{-s\,A(\boldsymbol{v})\}\exp\{-t\,A(\boldsymbol{w})\}] \frac{ds}{s}\,\frac{dt}{t}.
\end{aligned}$$

where $\ell(\boldsymbol{y}) = |\boldsymbol{y}|\,A(\boldsymbol{y}/|\boldsymbol{y}|)$ and $|\boldsymbol{y}| = |y_1| + \cdots + |y_p|$. Replacing A by any estimator of it results in an estimator of the covariance function. However, a more practical way to estimate this function is by the sample covariance of the pairs $(-\log \xi_i(\boldsymbol{v}),\, -\log \xi_i(\boldsymbol{w}))$; see also (the proof of) Theorem 3.7.

Remark 3.5 (Shape constraints)**.** *A further enhancement to the CFG-estimator is to replace it by the convex minorant of the function*

$$\min[\max\{\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}), w_1, \ldots, w_p\}, 1], \qquad \boldsymbol{w} \in \Delta_{p-1},$$

as in Deheuvels [1991] and Jiménez et al. [2001b] for the bivariate case. Although the resulting estimator would be a convex function respecting the bounds $\max(w_1, \ldots, w_p) \leqslant A(\boldsymbol{w}) \leqslant 1$, *in case $p \geqslant 3$ this would still not guarantee it to be a genuine Pickands dependence function. Hitherto, the only valid procedure to transform initial estimates for A into valid projection estimates has been presented in Gudendorf and Segers [2011b] and can be found in Chapter 5 of this work.*

3.2 The OLS-estimator

The question remains which weight functions λ_j to choose in the CFG-estimator (2.9). In Zhang et al. [2008], the choice $\lambda_j(\boldsymbol{w}) = w_j$ was recommended as a pragmatic one. The option of using variance-minimizing functions λ_j was mentioned but not carried out. By casting the estimation problem in a linear regression framework, we will obtain an estimator with the same asymptotic performance as the CFG-estimator with those optimal weights. In this section, we define the estimator and prove its consistency and asymptotic normality, both in the functional sense. In the next section, the gain in efficiency is assessed by means of simulations.

In view of Theorem 3.3, for each $\boldsymbol{w} \in \Delta_{p-1}$ we have

$$\sqrt{n}\big(\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - A(\boldsymbol{w})\big) \rightsquigarrow A(\boldsymbol{w})\,\eta(\boldsymbol{w}), \qquad n \to \infty,$$

where $\eta(\boldsymbol{w})$ is a zero-mean normal random variable. We will look for those $\lambda_j(\boldsymbol{w})$ that minimize the variance of $\eta(\boldsymbol{w})$. Let ζ be the Gaussian process on $\mathscr{C}(\Delta_{p-1})$ in Proposition 3.2. For ease of notation, put

$$\boldsymbol{\lambda}(\boldsymbol{w}) = \big(\lambda_1(\boldsymbol{w}), \ldots, \lambda_p(\boldsymbol{w})\big)^\top, \qquad \boldsymbol{\zeta}(\boldsymbol{e}) = \big(\zeta(\boldsymbol{e}_1), \ldots, \zeta(\boldsymbol{e}_p)\big)^\top,$$

the symbol "$\top$" denoting matrix transposition. Then

$$\begin{aligned}\operatorname{var}\eta(\boldsymbol{w}) &= \operatorname{var}\big(\zeta(\boldsymbol{w}) - \boldsymbol{\lambda}(\boldsymbol{w})^\top \boldsymbol{\zeta}(\boldsymbol{e})\big) \\ &= \operatorname{var}\zeta(\boldsymbol{w}) - 2\,\boldsymbol{\lambda}(\boldsymbol{w})^\top E[\boldsymbol{\zeta}(\boldsymbol{e})\,\zeta(\boldsymbol{w})] + \boldsymbol{\lambda}(\boldsymbol{w})^\top E[\boldsymbol{\zeta}(\boldsymbol{e})\,\boldsymbol{\zeta}(\boldsymbol{e})^\top]\,\boldsymbol{\lambda}(\boldsymbol{w}).\end{aligned}$$

Note that

$$\Sigma = E[\boldsymbol{\zeta}(\boldsymbol{e})\,\boldsymbol{\zeta}(\boldsymbol{e})^\top] \tag{3.8}$$

is the covariance matrix of $(-\log U_{i,1}, \ldots, -\log U_{i,p})^\top$. Provided this matrix is non-singular, $\operatorname{var}\eta(\boldsymbol{w})$ attains a unique global minimum for $\boldsymbol{\lambda}(\boldsymbol{w})$ equal to

$$\boldsymbol{\lambda}^{\mathrm{opt}}(\boldsymbol{w}) = \Sigma^{-1} E[\boldsymbol{\zeta}(\boldsymbol{e})\,\zeta(\boldsymbol{w})]. \tag{3.9}$$

With this choice of the weight functions, the variance of

$$\eta_{\mathrm{opt}}(\boldsymbol{w}) = \zeta(\boldsymbol{w}) - \boldsymbol{\lambda}^{\mathrm{opt}}(\boldsymbol{w})^\top \boldsymbol{\zeta}(\boldsymbol{e}) \tag{3.10}$$

is equal to

$$\operatorname{var}\eta_{\mathrm{opt}}(\boldsymbol{w}) = \operatorname{var}\zeta(\boldsymbol{w}) - E[\zeta(\boldsymbol{w})\,\boldsymbol{\zeta}(\boldsymbol{e})^\top]\,\Sigma^{-1} E[\boldsymbol{\zeta}(\boldsymbol{e})\,\zeta(\boldsymbol{w})]. \tag{3.11}$$

This variance is minimal over all possible choices of weight functions λ_j.

The optimal weight functions λ_j^{opt} in (3.9) depend on the unknown Pickands dependence function A. Fortunately, replacing these weight functions by uniformly consistent estimators $\hat{\lambda}_{n,j}$ is just as good asymptotically. For such estimated weight functions, define the adaptive CFG-estimator by

$$\log \hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \hat{\lambda}_{n,j}(\boldsymbol{w}) \log \hat{A}_n(\boldsymbol{e}_j), \tag{3.12}$$

Proposition 3.6 (Adaptive CFG-estimator)**.** *Assume that, in addition to the assumptions in Proposition 3.2, the matrix* Σ *in* (3.8) *is non-singular and* $\hat{\lambda}_{n,j}$ *are random elements in* $\mathscr{C}(\Delta_{p-1})$ *such that, for every* $j \in \{1, \ldots, p\}$*,*

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}} |\hat{\lambda}_{n,j}(\boldsymbol{w}) - \lambda_j^{\mathrm{opt}}(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely,}$$

with λ_j^{opt} *as in* (3.9)*. Then the adaptive CFG-estimator in* (3.12) *satisfies*

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}} |\hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely,} \tag{3.13}$$

and in $\mathscr{C}(\Delta_{p-1})$*,*

$$\sqrt{n}(\hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}} - A) \rightsquigarrow A\, \eta_{\mathrm{opt}}, \qquad n \to \infty, \tag{3.14}$$

where η_{opt} *is the zero-mean Gaussian process defined in* (3.10)*.*

Finally we propose a particularly convenient way to implement the adaptive CFG-estimator in (3.12). For $\boldsymbol{w} \in \Delta_{p-1}$, let

$$\hat{\boldsymbol{\beta}}_n(\boldsymbol{w}) = (\hat{\beta}_{n,0}(\boldsymbol{w}), \ldots, \hat{\beta}_{n,p}(\boldsymbol{w}))^\top$$

be the minimizer in $(b_0, \ldots, b_p)^\top$ of

$$\sum_{i=1}^{n} \Big(\big(-\log \xi_i(\boldsymbol{w}) - \gamma\big) - b_0 - \sum_{j=1}^{p} b_j \big(-\log \xi_i(\boldsymbol{e}_j) - \gamma\big) \Big)^2. \tag{3.15}$$

In words, $\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})$ is the ordinary least-squares (OLS) estimator of the vector of regression coefficients in a linear regression of the dependent variable $-\log \xi_i(\boldsymbol{w}) - \gamma$ upon the explanatory variables $-\log \xi_i(\boldsymbol{e}_j) - \gamma$, $j \in \{1, \ldots, p\}$. Define the OLS-estimator of A via the estimated intercept by

$$\log \hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w}) = \hat{\beta}_{n,0}(\boldsymbol{w}), \qquad \boldsymbol{w} \in \Delta_{p-1}.$$

Since the residuals

$$\hat{\epsilon}_{n,i}(\boldsymbol{w}) = \bigl(-\log \xi_i(\boldsymbol{w}) - \gamma\bigr) - \hat{\beta}_{n,0}(\boldsymbol{w}) - \sum_{j=1}^{p} \hat{\beta}_{n,j}(\boldsymbol{w}) \bigl(-\log \xi_i(\boldsymbol{e}_j) - \gamma\bigr)$$

verify $\sum_{i=1}^{n} \hat{\epsilon}_{n,i}(\boldsymbol{w}) = 0$, we have

$$\log \hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w}) = \hat{\beta}_{n,0}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \hat{\beta}_{n,j}(\boldsymbol{w}) \log \hat{A}_n(\boldsymbol{e}_j), \quad (3.16)$$

that is, the OLS-estimator is equal to the adaptive CFG-estimator with estimated weights $\hat{\lambda}_{n,j}(\boldsymbol{w}) = \hat{\beta}_{n,j}(\boldsymbol{w})$. The variance of the (logarithm of the) OLS-estimator can be estimated by the sample variance of the residuals, properly corrected for the loss in number of degrees of freedom,

$$\hat{\sigma}^2_{n,\mathrm{OLS}}(\boldsymbol{w}) = \frac{1}{n-p-1} \sum_{i=1}^{n} \hat{\epsilon}^2_{n,i}(\boldsymbol{w}), \qquad \boldsymbol{w} \in \Delta_{p-1}. \quad (3.17)$$

Theorem 3.7 (OLS-estimator)**.** *Assume that, in addition to the assumptions in Proposition 3.2, the matrix Σ in* (3.8) *is non-singular. Then, with probability tending to one, the minimizer $\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})$ of* (3.15) *is uniquely defined and for $j \in \{1, \ldots, p\}$,*

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}} |\hat{\beta}_{n,j}(\boldsymbol{w}) - \lambda_j^{\mathrm{opt}}(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely.} \quad (3.18)$$

As a consequence, the OLS-estimator in (3.16) *is uniformly consistent,*

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}} |\hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w}) - A(\boldsymbol{w})| \to 0, \qquad n \to \infty, \qquad \textit{almost surely}, \quad (3.19)$$

and in $\mathscr{C}(\Delta_{p-1})$,

$$\sqrt{n}(\hat{A}_n^{\mathrm{OLS}} - A) \rightsquigarrow A\, \eta_{\mathrm{opt}}, \qquad n \to \infty, \quad (3.20)$$

where η_{opt} is the zero-mean Gaussian process defined in (3.10)*. In addition, the variance estimator in* (3.17) *satisfies*

$$\sup_{\boldsymbol{w}\in\Delta_{p-1}} |\hat\sigma^2_{n,\mathrm{OLS}}(\boldsymbol{w}) - \operatorname{var}\eta_{\mathrm{opt}}(\boldsymbol{w})| \to 0, \qquad n\to\infty, \qquad \textit{almost surely.}$$

Remark 3.8 (Non-singularity assumption)**.** *In the bivariate case, the assumption that the covariance matrix* Σ *in* (3.8) *is non-singular is equivalent to the assumption that the copula* C *is not the comonotone copula Segers [2007]. We conjecture that in the general multivariate case, a necessary and sufficient condition for* Σ *to be non-singular is that none of the bivariate margins of* C *is equal to the comonotone copula.*

3.3 Simulations

In order to investigate the finite-sample properties of the estimators discussed in the previous sections, we generated pseudo-random samples from trivariate extreme-value copulas of logistic type as presented in Tawn [1990]:

$$\begin{aligned} A(\boldsymbol{w}) &= (\theta^r w_1^r + \phi^r w_2^r)^{1/r} + (\theta^r w_2^r + \phi^r w_3^r)^{1/r} + (\theta^r w_3^r + \phi^r w_1^r)^{1/r} \\ &\quad + \psi(w_1^r + w_2^r + w_3^r)^{1/r} + 1 - \theta - \phi - \psi, \qquad \boldsymbol{w}\in\Delta_p, \quad (3.21) \end{aligned}$$

for $(r,\theta,\phi,\psi)\in[1,\infty)\times[0,1]^3$. To facilitate comparisons, we opted for the same parameter values as chosen in Zhang et al. [2008]: a symmetric case, $(r,\theta,\phi,\psi) = (3,0,0,1)$, and an asymmetric one, $(r,\theta,\phi,\psi) = (6, 0.6, 0.3, 0)$. For each case 10 000 samples were generated of size $n\in\{50,100,200\}$ using the simulation algorithms in Stephenson [2003] and implemented in the R-package evd Stephenson [2002].

Four estimators were compared: the CFG-estimator $\hat A_n^{\mathrm{CFG}}$ with weight functions $\lambda_j(\boldsymbol{w}) = w_j$ as recommended in Zhang et al. [2008], the OLS-estimator $\hat A_n^{\mathrm{OLS}}$ in (3.16), and the enhanced versions of the original Pickands estimator due to Deheuvels [1991] and Hall and Tajvidi [2000] as presented in Zhang et al. [2008]. To visualize the performances of the estimators, we plotted their biases and mean squared errors along the line $\{\boldsymbol{w}\in\Delta_{p-1} : w_1 = w_2\}$; see Figures 3.1 and 3.2 for the symmetric and asymmetric logistic dependence functions respectively. The

	$n = 25$		$n = 50$		$n = 100$		$n = 200$	
	SLDF	ALDF	SLDF	ALDF	SLDF	ALDF	SLDF	ALDF
MISE $\times 10^{-5}$								
Deheuvels	128	425	64	204	32	97	15	47
HT	23	377	12	184	6	88	3	43
ZWP	21	270	12	128	6	63	3	31
OLS	15	250	7	107	4	50	2	24
MMSE $\times 10^{-5}$								
Deheuvels	347	1290	172	584	87	274	47	135
HT	113	1170	57	580	29	278	14	136
ZWP	63	800	32	380	16	184	8	93
OLS	74	730	33	297	16	137	7	67

Table 3.1: Mean integrated squared errors (MISE) and maximum mean squared errors (MMSE) for 10 000 samples of size $n \in \{25, 50, 100, 200\}$ from the trivariate extreme-value copula C with logistic dependence function A in (6.9) for two different choices of the parameter vector: a symmetric case (SLDF), $(r, \theta, \phi, \psi) = (3, 1, 1, 1)$, and an asymmetric case (ALDF), $(r, \theta, \phi, \psi) = (6, 0.6, 0.3, 0)$.

mean integrated squared errors (MISE) and the maximum mean squared errors (MMSE) are shown in Table 3.1.

In accordance to the theory, the OLS-estimator is in virtually all cases considered more efficient than the CFG-estimator. Moreover, our simulations confirm the findings in Zhang et al. [2008] that the CFG-estimator is typically more efficient than the ones of Deheuvels and Hall–Tajvidi. Note that the finite-sample bias of the OLS-estimator is somewhat larger than for the other estimators. However, thanks to its minimum-variance property it ends up as an overall winner in terms of mean squared error.

3.4 Proofs

3.4.1 Proof of Lemma 3.1

Keeping in mind the definition of $Z_{i,j}$ in (2.5), we are going to manipulate

$$\int_0^{1-w_j} \frac{\mathbf{1}(Z_{i,j} \le z) - z}{z(1-z)} dz$$

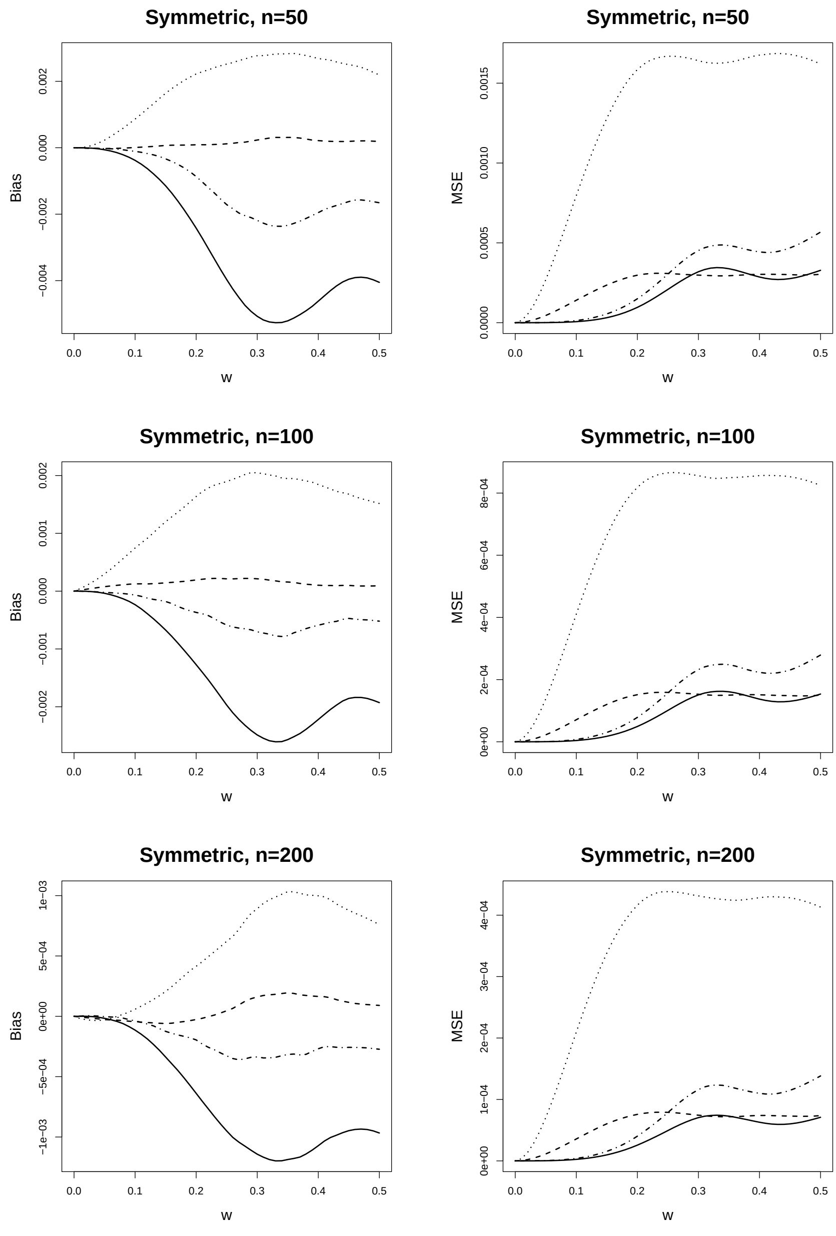


Figure 3.1: Biases (left) and mean squared errors (right) of $\hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w})$ (solid), $\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w})$ (dashed), $\hat{A}_n^{\mathrm{HT}}(\boldsymbol{w})$ (dash-dotted) and $\hat{A}_n^{\mathrm{D}}(\boldsymbol{w})$ (dotted) along the line $w_1 = w_2$ for 10 000 samples of size $n \in \{50, 100, 200\}$ from the trivariate extreme-value copula C with symmetric logistic dependence function $A(\boldsymbol{w}) = (w_1^r + w_2^r + w_3^r)^{1/r}$ at $r = 3$.

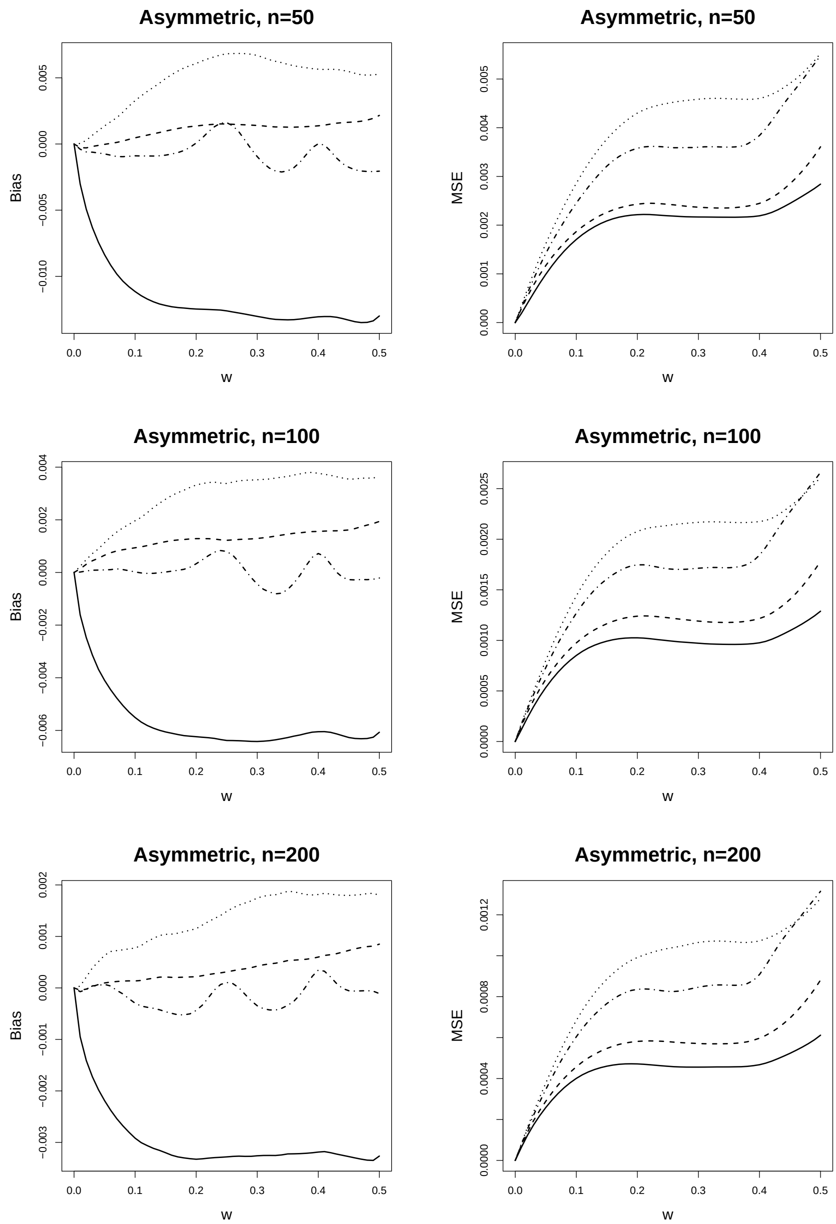


Figure 3.2: Biases (left) and mean squared errors (right) of $\hat{A}_n^{\mathrm{OLS}}(\boldsymbol{w})$ (solid), $\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w})$ (dashed), $\hat{A}_n^{\mathrm{HT}}(\boldsymbol{w})$ (dash-dotted) and $\hat{A}_n^{\mathrm{D}}(\boldsymbol{w})$ (dotted) along the line $w_1 = w_2$ for 10 000 samples of size $n \in \{50, 100, 200\}$ from the trivariate extreme-value copula C with asymmetric logistic dependence function A in (6.9) for $(r, \theta, \phi, \psi) = (6, 0.6, 0.3, 0)$.

using similar computations as in section 4 of Segers [2007] for the bivariate case.

The indicator function in the integral disappears if we use the following decomposition:

$$\begin{aligned}\int_0^{1-w_j} \frac{\mathbf{1}(Z_{i,j} \le z) - z}{z(1-z)} dz &= -\int_0^{(1-w_j)\wedge Z_{i,j}} \frac{1}{1-z} dz + \int_{(1-w_j)\wedge Z_{i,j}}^{(1-w_j)} \frac{1}{z} dz \\ &= \log\{1 - \big((1-w_j)\wedge Z_{i,j}\big)\} + \log(1-w_j) \\ &\quad - \log\{(1-w_j)\wedge Z_{i,j}\}\end{aligned}$$

Rearranging the previous expression and exploiting the relationships between maximum and the minimum in the case of inverse functions or functions of opposite sign, we find:

$$\int_0^{1-w_j} \frac{\mathbf{1}(Z_{i,j} \le z) - z}{z(1-z)} dz = \log\left(-\frac{\log U_{i,j}}{1-w_j}\right) - \log\left(-\frac{\log(U_{i,j})}{w_j(1-w_j)} \bigwedge \left(\bigwedge_{l\neq j}^{p} -\frac{\log(U_{i,l})}{w_l(1-w_j)}\right)\right)$$

We can further simplify to obtain:

$$\begin{aligned}\int_0^{1-w_j} \frac{\mathbf{1}(Z_{i,j} \le z) - z}{z(1-z)} dz &= \log\left(-\frac{\log U_{i,j}}{1-w_j}\right) - \log\left(\bigwedge_{l=1}^{p} -\frac{\log U_{i,l}}{w_l(1-w_j)}\right) \\ &= \log\left(-\log U_{i,j}\right) - \log\left(\bigwedge_{l=1}^{p} -\frac{\log U_{i,l}}{w_l}\right) \\ &= \log\left(-\log U_{i,j}\right) - \log \xi_i(\boldsymbol{w})\end{aligned}$$

using the definition of $\xi_i(\boldsymbol{w})$ given in (2.1). Replacing the previously derived expressions into the ZWP estimator (2.6):

$$\begin{aligned}
\log\left(\hat{A}_n^{ZWP}(\boldsymbol{w})\right) &= \sum_{j=1}^{p} \frac{\lambda_j(\boldsymbol{w})}{n} \sum_{i=1}^{n} \int_0^{1-w_j} \frac{\mathbf{1}(Z_{i,j} \le z) - z}{z(1-z)} dz \\
&= \sum_{j=1}^{p} \frac{\lambda_j(\boldsymbol{w})}{n} \sum_{i=1}^{n} \Big\{ \log(-\log(U_{i,j})) - \log \xi_i(\boldsymbol{w}) \Big\} \\
&= \sum_{j=1}^{p} \frac{\lambda_j(\boldsymbol{w})}{n} \sum_{i=1}^{n} \Big\{ \log(-\log(U_{i,j})) + \gamma - \gamma - \log \xi_i(\boldsymbol{w}) \Big\} \\
&= -\frac{1}{n} \sum_{i=1}^{n} \log\left(\xi_i(\boldsymbol{w})\right) - \gamma \\
&\quad + \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \frac{1}{n} \sum_{i=1}^{n} \Big\{ \log(-\log U_{i,j}) + \gamma \Big\} \\
&= \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \log \hat{A}_n(\boldsymbol{e}_j)
\end{aligned}$$

The last two equalities were obtained by adding and subtracting γ and using the fact that $\sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) = 1$ for every $\boldsymbol{w} \in \Delta_p$.

3.4.2 Proof of Remark 3.4

We will establish an analytical expression for the covariance matrix Σ, i.e. evaluate the expression for

$$\operatorname{Cov}\left(-\log(\xi(\boldsymbol{w})), -\log(\xi(\boldsymbol{v}))\right), \quad \boldsymbol{w}, \boldsymbol{v} \in \Delta_{p-1}.$$

We know that $\xi\big(\boldsymbol{w}\big)$ has a scalar exponential distribution with mean $1/A(\boldsymbol{w})$ implying that $-\log \xi(\boldsymbol{w})$ belongs to a Gumbel location family so that

$$\operatorname{Var}\left(-\log(\xi(\boldsymbol{w}))\right) = \frac{\pi^2}{6}.$$

For the general case with $\boldsymbol{w}$ and $\boldsymbol{v}$ being two different vectors of the unit-simplex Δ_{p-1}, we are going to use the following integral representation for the logarithm

$$\log(x) = \int_0^{+\infty} \{\mathbf{1}(s \le x) - \mathbf{1}(s \le 1)\} \frac{ds}{s}.$$

Then,

$$\begin{aligned}
&\mathrm{E}[\log\xi(\boldsymbol{v})\log\xi(\boldsymbol{w})]\\
=\ &\mathrm{E}\left[\int_0^\infty\{\mathbf{1}(s\le\xi(\boldsymbol{v}))-\mathbf{1}(s\le 1)\}\frac{ds}{s}\int_0^\infty\{\mathbf{1}(t\le\xi(\boldsymbol{w}))-\mathbf{1}(t\le 1)\}\frac{dt}{t}\right]\\
=\ &\mathrm{E}\left[\int_0^\infty\int_0^\infty\{\mathbf{1}(s\le\xi(\boldsymbol{v}))-\mathbf{1}(s\le 1)\}\{\mathbf{1}(t\le\xi(\boldsymbol{w}))-\mathbf{1}(t\le 1)\}\frac{dt}{t}\frac{ds}{s}\right]\\
=\ &\int_0^\infty\int_0^\infty\mathrm{E}\Big[\{\mathbf{1}(s\le\xi(\boldsymbol{v}))-\mathbf{1}(s\le 1)\}\{\mathbf{1}(t\le\xi(\boldsymbol{w}))-\mathbf{1}(t\le 1)\}\Big]\frac{dt}{t}\frac{ds}{s}
\end{aligned}$$

and

$$\mathrm{E}[\log(\xi(\boldsymbol{v}))]\,\mathrm{E}[\log(\xi(\boldsymbol{w}))]=$$
$$\int_0^\infty\int_0^\infty\mathrm{E}\Big[\{\mathbf{1}(s\le\xi(\boldsymbol{v}))-\mathbf{1}(s\le 1)\}\Big]\,\mathrm{E}[\{\mathbf{1}(t\le\xi(\boldsymbol{w}))-\mathbf{1}(t\le 1)\}\Big]\frac{dt}{t}\frac{ds}{s}.$$

The covariance function can be decomposed as follows

$$\begin{aligned}
&\mathrm{Cov}\big(\log(\xi(\boldsymbol{v})),\log(\xi(\boldsymbol{w}))\big)\\
=\ &\mathrm{E}[\log\xi(\boldsymbol{v})\log\xi(\boldsymbol{w})]-\mathrm{E}[\log\xi(\boldsymbol{v})]\,\mathrm{E}[\log\xi(\boldsymbol{w})]\\
=\ &\int_0^\infty\int_0^\infty\Big(\mathrm{E}[1(s\le\xi(\boldsymbol{v}))1(t\le\xi(\boldsymbol{w}))]\\
&\qquad-\mathrm{E}[1(s\le\xi(\boldsymbol{v}))]\,\mathrm{E}[1(t\le\xi(\boldsymbol{w}))]\Big)\frac{ds}{s}\frac{dt}{t}\\
=\ &\int_0^\infty\int_0^\infty\Big(\mathrm{P}(\xi(\boldsymbol{v})\geqslant s,\xi(\boldsymbol{w})\geqslant t)-\mathrm{P}(\xi(\boldsymbol{v})\geqslant s)\,\mathrm{P}(\xi(\boldsymbol{w})\geqslant t)\Big)\frac{ds}{s}\frac{dt}{t}.
\end{aligned}$$

The fact that $\xi(\boldsymbol{v})$ and $\xi(\boldsymbol{w})$ are exponentially distributed leads to the final result as

$$\mathrm{P}(\xi(\boldsymbol{v})\geqslant s)=\exp\{-sA(v_1,\dots,v_p)$$

and

$$\mathrm{P}(\xi(\boldsymbol{v})\geqslant s,\xi(\boldsymbol{w})\geqslant t)=\exp\Big\{-l\Big((v_1s)\vee(w_1t),\dots,(v_ps)\vee(w_pt)\Big)\Big\}.$$

3.4.3 Proof of Proposition 3.2

For $\boldsymbol{w}\in\Delta_{p-1}$, define $f_{\boldsymbol{w}}:(0,\infty)^p\to\mathbb{R}$ by

$$f_{\boldsymbol{w}}(\boldsymbol{y})=-\log\Big(\bigwedge_{j=1}^p\frac{y_j}{w_j}\Big)-\gamma,\qquad \boldsymbol{y}\in(0,\infty)^p. \tag{3.22}$$

We can write

$$\log \hat{A}_n(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n} f_{\boldsymbol{w}}(\boldsymbol{Y}_i).$$

Consider the function class $\mathscr{F} = \{f_{\boldsymbol{w}} : \boldsymbol{w} \in \Delta_{p-1}\}$. We will show that $\mathscr{F}$ is P-Donsker and therefore also P-Glivenko–Cantelli, where P denotes the common probability distribution on $(0,\infty)^p$ of the random vectors $\boldsymbol{Y}_i$. According to Theorem 2.6.8 in van der Vaart and Wellner [1996] and the proof thereof, we need to verify that $\mathscr{F}$ is a pointwise separable Vapnik–Červonenkis-class (VC-class) that admits an envelope function with a finite second moment under P. From this result, it will be trivial to show that the function class

$$\mathscr{G} = \left\{g_{\boldsymbol{w}} : \mathbb{R}_+^p \to \mathbb{R} \text{ where } g_{\boldsymbol{w}}(\boldsymbol{x}) = -f_{\boldsymbol{w}}(\boldsymbol{x}) - \gamma\right\} \tag{3.23}$$

will also be $P-$Donsker and hence also $P-$Glivenko-Cantelli.

1. First, the function class $\mathscr{F}$ is pointwise separable since $\boldsymbol{w} \mapsto \log f_{\boldsymbol{w}}(\boldsymbol{y})$ is continuous for every $\boldsymbol{y} \in (0,+\infty)^p$. So, it suffices to take the intersection between Δ_{p-1} and $\mathbb{Q}^p$ if we identify $\mathscr{F}$ with Δ_{p-1}.

2. $\mathscr{F}$ and $\mathscr{G}$ are Vapnik–Červonenkis subgraph (VC) classes by repeated applications of Lemma 2.6.15 and points (i) and (viii)($\phi(\cdot) = \log(\cdot)$) of Lemma 2.6.18 of van der Vaart and Wellner [1996].

   In detail, consider the p vector spaces of linear functions

   $$\mathscr{F}_j = \left\{ f_{a_j}; f_{a_j} : [0,\infty)^p \to \mathbb{R}\,\text{s.t.}\,(x_1,\dots,x_p) \mapsto \frac{x_j}{a_j} \quad \text{for}\, a_j \in \mathbb{R}\setminus\{0\} \right\},$$

   $j \in \{1,\dots,p\}$, which are VC-subgraph as guaranteed by Lemma 2.6.15. Then, from the aforementioned points in Lemma 2.6.18 one can deduce that

   $$\begin{aligned}\mathscr{F} = \mathscr{F}_1 \wedge \ldots \wedge \mathscr{F}_p = \{f_1 \wedge \ldots \wedge f_p; f_1 \in \mathscr{F}_1, \ldots,\\ f_p \in \mathscr{F}_p \,\text{and}\, (a_1,\dots,a_p) \in \Delta_{p-1}\}\end{aligned}$$

   and

   $$\phi \circ \mathscr{F} = \{\phi \circ f; f \in \mathscr{F}\} \quad \text{with}\ \phi = \log$$

   are also VC-subgraph.

3. We have the following inequalities for any values $\boldsymbol{y} = (y_1, \ldots, y_p) \in (0, +\infty)^p$ and $\boldsymbol{w} = (w_1, \ldots, w_p = 1 - \sum_{j=1}^{p-1} w_j)$

$$\bigwedge_{j=1}^{p} y_j \leq \bigwedge_{j=1}^{p} \frac{y_j}{w_j} \leq p \bigvee_{j=1}^{p} y_j$$

as $w_j \geqslant 1/p$ for all $j \in \{1, \ldots, p\}$. From the strict monotonicity of the logarithm function:

$$\log\left(\bigwedge_{j=1}^{p} y_j\right) \leq \log\left(\bigwedge_{j=1}^{p} \frac{y_j}{w_j}\right) \leq \log(p) + \log\left(\bigvee_{j=1}^{p} y_j\right)$$

We can choose the following envelope function:

$$F(\boldsymbol{y}) = \max\left\{|\log(p \vee_{j=1}^{p} y_j)|, |\log(\wedge_{j=1}^{p} y_j)|\right\} \tag{3.24}$$

from where we can deduce that

$$F^2(\boldsymbol{y}) \leq \underbrace{\log(p \vee_{j=1}^{p} y_j)^2}_{\text{(I)}} + \underbrace{\log(\wedge_{j=1}^{p} y_j)^2}_{\text{(II)}}$$

- For the term (I) , we show that

$$\begin{aligned}
\log(p) + \log(\vee_{j=1}^{p} y_j) &\leq \log(p) + \log\left(\sum_{j=1}^{p} y_j\right) \\
&\leq \log(p) + \log\left(\sum_{j=1}^{p} (y_j \wedge 1)\right) \\
&\leq \log(p) + \sum_{j=1}^{p} \log(y_j \wedge 1) \\
&\leq \log(p) + \sum_{j=1}^{p} |\log(y_j)|
\end{aligned}$$

- Observe that the distribution of $\bigwedge_{j=1}^{p} Y_{i,j}$ is Exponential with mean equal to $\{p\, A(1/p, \ldots, 1/p)\}^{-1} \in [1/p, 1]$.

F yields an envelope function of $\mathscr{F}$ all of whose moments are finite under P.

From the fact that $\mathscr{F}$ is P-Glivenko–Cantelli it follows that

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in\Delta_{p-1}} |\log \hat{A}_n(\boldsymbol{w}) - \log A(\boldsymbol{w})| \\
&= \sup_{\boldsymbol{w}\in\Delta_{p-1}} \left| \frac{1}{n} \sum_{i=1}^{n} f_{\boldsymbol{w}}(\boldsymbol{Y}_i) - E[f_{\boldsymbol{w}}(\boldsymbol{Y})] \right| \to 0, \qquad n \to \infty, \qquad \text{almost surely.}
\end{aligned}$$

(Here, we dropped a subscript i for convenience.) Continuity of the map $\exp : \mathscr{C}(\Delta_{p-1}) \to \mathscr{C}(\Delta_{p-1}) : f \mapsto \exp(f)$ yields uniform consistency as in (3.2).

Moreover, the P-Donsker property entails

$$\sqrt{n}(\log \hat{A}_n^{\mathrm{CFG}} - \log A) \rightsquigarrow \zeta, \qquad n \to \infty, \tag{3.25}$$

in the space $\ell^\infty(\Delta_{p-1})$ of bounded functions from Δ_{p-1} into $\mathbb{R}$ equipped with the topology of uniform convergence, where we identified $\mathscr{F}$ with Δ_{p-1}. The process ζ is zero-mean Gaussian with covariance function given in (3.4). The sample paths of the limit process ζ are continuous with respect to the standard deviation (semi-)metric ρ on Δ_{p-1} defined by

$$\rho(\boldsymbol{v}, \boldsymbol{w}) = [\operatorname{var}\{f_{\boldsymbol{v}}(\boldsymbol{Y}) - f_{\boldsymbol{w}}(\boldsymbol{Y})\}]^{1/2}, \qquad \boldsymbol{v}, \boldsymbol{w} \in \Delta_{p-1}.$$

If $\lim_{n\to\infty} \boldsymbol{v}_n = \boldsymbol{v}$ in Δ_{p-1} according to the Euclidean metric, then by continuity of $f_{\boldsymbol{w}}(\boldsymbol{y})$ in $\boldsymbol{w}$ and by uniform integrability, also $\lim_{n\to\infty} \rho(\boldsymbol{v}_n, \boldsymbol{v}) = 0$. (Uniform integrability is checked by using the bound in (3.24).) It follows that the trajectories of ζ are also continuous with respect to the Euclidean metric on Δ_{p-1}, that is, ζ actually takes its values in $\mathscr{C}(\Delta_{p-1})$. As the trajectories of the left-hand side in (3.25) are continuous too, the convergence in (3.25) takes place not only in $\ell^\infty(\Delta_{p-1})$ but also in $\mathscr{C}(\Delta_{p-1})$.

The convergence in (3.3) follows from the Hadamard-differentiability of the map $\exp : \mathscr{C}(\Delta_{p-1}) \to \mathscr{C}(\Delta_{p-1}) : f \mapsto \exp f$ and the functional delta-method van der Vaart and Wellner [1996, Section 3.9].

3.4.4 Proof of Theorem 3.3

Uniform consistency of $\hat{A}_n^{\mathrm{CFG}}$ in (3.5) follows from uniform consistency of $\hat{A}_n$ in (3.2) and the fact that the functions λ_j are continuous, hence bounded.

To show (3.6), define $L : \mathscr{C}(\Delta_{p-1}) \to \mathscr{C}(\Delta_{p-1})$ by

$$L f(\boldsymbol{w}) = f(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w})\, f(\boldsymbol{e}_j)$$

for $f \in \mathscr{C}(\Delta_{p-1})$ and $\boldsymbol{w} \in \Delta_{p-1}$. The operator L is linear and bounded. We have $\log \hat{A}_n^{\mathrm{CFG}} = L(\log \hat{A}_n)$. Moreover, as $A(\boldsymbol{e}_j) = 1$ for all $j \in \{1, \ldots, p\}$, also $L(\log A) = \log A$. We find

$$\sqrt{n}(\log \hat{A}_n^{\mathrm{CFG}} - \log A) = L\big(\sqrt{n}(\log \hat{A}_n - \log A)\big) \rightsquigarrow L\zeta = \eta, \qquad n \to \infty.$$

The weak convergence in (3.6) follows from the functional delta-method van der Vaart and Wellner [1996, Section 3.9]. The representation $\eta = L\zeta$ coincides with (3.7).

3.4.5 Proof of Proposition 3.6

If the optimal weight functions λ_j^{opt} were known, we could consider the optimal CFG-estimator

$$\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) = \log \hat{A}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j^{\mathrm{opt}}(\boldsymbol{w}) \log \hat{A}_n(\boldsymbol{e}_j), \qquad \boldsymbol{w} \in \Delta_{p-1}.$$

By Theorem 3.3, the optimal CFG-estimator is uniformly consistent (3.5) and is asymptotically normal in the sense of (3.6) with $\eta = \eta_{\mathrm{opt}}$. Now

$$|\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) - \log \hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w})| \leqslant \sum_{j=1}^{p} |\hat{\lambda}_{n,j}(\boldsymbol{w}) - \lambda_j^{\mathrm{opt}}(\boldsymbol{w})|\, |\log \hat{A}_n(\boldsymbol{e}_j)|.$$

By uniform consistency of $\hat{\lambda}_{n,j}$ and asymptotic normality of $\sqrt{n}\log \hat{A}_n(\boldsymbol{e}_j)$, we obtain, as $n \to \infty$,

$$\sup_{\boldsymbol{w}\in\Delta_{p-1}} |\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) - \log \hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w})| \to 0, \qquad \text{almost surely,}$$

$$\sup_{\boldsymbol{w}\in\Delta_{p-1}} \sqrt{n}\,|\log \hat{A}_{n,\mathrm{opt}}^{\mathrm{CFG}}(\boldsymbol{w}) - \log \hat{A}_{n,\mathrm{ad}}^{\mathrm{CFG}}(\boldsymbol{w})| \rightsquigarrow 0.$$

As a consequence, the adaptive CFG-estimator is uniformly consistent (3.13) and asymptotically normal (3.14).

3.4.6 Proof of Theorem 3.7

In analogy to the linear regression framework, define the $n \times (p+1)$ matrix

$$\boldsymbol{X} = \begin{pmatrix} 1 & -\log \xi_1(\boldsymbol{e}_1) - \gamma & \dots & \log \xi_1(\boldsymbol{e}_p) - \gamma \\ \dots & \dots & \dots & \dots \\ 1 & -\log \xi_n(\boldsymbol{e}_1) - \gamma & \dots & \log \xi_n(\boldsymbol{e}_p) - \gamma \end{pmatrix}$$

and the $n \times 1$ vector

$$\boldsymbol{Y}(\boldsymbol{w}) = \bigl(-\log \xi_1(\boldsymbol{w}) - \gamma, \dots, -\log \xi_n(\boldsymbol{w}) - \gamma\bigr)^\top, \qquad \boldsymbol{w} \in \Delta_{p-1}.$$

Provided the matrix $\boldsymbol{X}^\top \boldsymbol{X}$ is non-singular, the OLS-estimator $\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})$ is given by

$$\hat{\boldsymbol{\beta}}_n(\boldsymbol{w}) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}(\boldsymbol{w}).$$

Recall the functions $f_{\boldsymbol{w}}$ in (3.22). For $\boldsymbol{v}, \boldsymbol{w} \in \Delta_{p-1}$, define $g_{\boldsymbol{v},\boldsymbol{w}} : (0,\infty)^p \to \mathbb{R}$ by

$$g_{\boldsymbol{v},\boldsymbol{w}}(\boldsymbol{y}) = f_{\boldsymbol{v}}(\boldsymbol{y})\, f_{\boldsymbol{w}}(\boldsymbol{y}), \qquad \boldsymbol{y} \in (0,\infty)^p.$$

By (3.24) and by Example 2.10.23 in van der Vaart and Wellner [1996], the function class $\{g_{\boldsymbol{v},\boldsymbol{w}} : \boldsymbol{v}, \boldsymbol{w} \in \Delta_{p-1}\}$ is P-Donsker and thus P-Glivenko–Cantelli, where P is the common distribution on $(0,\infty)^p$ of the random vectors $\boldsymbol{Y}_i$. It follows that, almost surely as $n \to \infty$,

$$\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} \to \begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}, \tag{3.26}$$

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}} \left| \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{Y}(\boldsymbol{w}) - \begin{pmatrix} \log A(\boldsymbol{w}) \\ E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})] \end{pmatrix} \right| \to 0, \tag{3.27}$$

As Σ is non-singular, we have

$$\begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma^{-1} \end{pmatrix},$$

while $\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}$ is with probability tending to one a non-singular matrix too. We find, almost surely and uniformly in $\boldsymbol{w} \in \Delta_{p-1}$,

$$\begin{aligned}\hat{\boldsymbol{\beta}}_n(\boldsymbol{w}) &= \left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{Y}(\boldsymbol{w}) \\ &\to \begin{pmatrix}1 & 0\\ 0 & \Sigma^{-1}\end{pmatrix}\begin{pmatrix}\log A(\boldsymbol{w})\\ E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})]\end{pmatrix} = \begin{pmatrix}\log A(\boldsymbol{w})\\ \boldsymbol{\lambda}^{\mathrm{opt}}(\boldsymbol{w})\end{pmatrix}, \qquad n \to \infty.\end{aligned}$$

Equation (3.18) follows. Proposition 3.6 and equation (3.16) then yield equations (3.19) and (3.20).

Finally, for the estimation of the variance, note that it does not matter asymptotically if we divide by n or by $n-p-1$. Elementary calculations yield

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_{n,i}^2(\boldsymbol{w}) &= \frac{1}{n}\big(\boldsymbol{Y}(\boldsymbol{w}) - \boldsymbol{X}\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})\big)^\top\big(\boldsymbol{Y}(\boldsymbol{w}) - \boldsymbol{X}\hat{\boldsymbol{\beta}}_n(\boldsymbol{w})\big) \\ &= \frac{1}{n}\boldsymbol{Y}(\boldsymbol{w})^\top\boldsymbol{Y}(\boldsymbol{w}) - \left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{Y}(\boldsymbol{w})\right)^\top\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{Y}(\boldsymbol{w}).\end{aligned}$$

The Glivenko–Cantelli property yields, almost surely and uniformly in $\boldsymbol{w} \in \Delta_{p-1}$,

$$\begin{aligned}\frac{1}{n}\boldsymbol{Y}(\boldsymbol{w})^\top\boldsymbol{Y}(\boldsymbol{w}) &= \frac{1}{n}\sum_{i=1}^n\big(-\log\xi_i(\boldsymbol{w}) - \gamma\big)^2 \\ &\to E\big[\big(-\log\xi_i(\boldsymbol{w}) - \gamma\big)^2\big] = \operatorname{var}\zeta(\boldsymbol{w}) + \big(\log A(\boldsymbol{w})\big)^2, \qquad n \to \infty.\end{aligned}$$

In combination with (3.26) and (3.27), we obtain that $n^{-1}\sum_{i=1}^n \hat{\epsilon}_{n,i}^2(\boldsymbol{w})$ converges almost surely and uniformly in $\boldsymbol{w} \in \Delta_{p-1}$ to

$$\begin{aligned}\operatorname{var}\zeta(\boldsymbol{w}) + \big(\log A(\boldsymbol{w})\big)^2 - \begin{pmatrix}\log A(\boldsymbol{w})\\ E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})]\end{pmatrix}^\top\begin{pmatrix}1 & 0\\ 0 & \Sigma^{-1}\end{pmatrix}\begin{pmatrix}\log A(\boldsymbol{w})\\ E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})]\end{pmatrix} \\ = \operatorname{var}\zeta(\boldsymbol{w}) - E[\boldsymbol{\zeta}(\boldsymbol{e})^\top\zeta(\boldsymbol{w})]\,\Sigma^{-1}\,E[\boldsymbol{\zeta}(\boldsymbol{e})\zeta(\boldsymbol{w})],\end{aligned}$$

which by (3.11) is equal to $\operatorname{var}\eta_{\mathrm{opt}}(\boldsymbol{w})$.

4 Inference: Unknown margins

In the following chapter we will drop the assumption of known marginal distributions and treat the more realistic case of unknown margins as done in Gudendorf and Segers [2011b]. Initially this topic has been treated in the bivariate case in Jiménez et al. [2001a] for a submodel and in Genest and Segers [2009] for the general model. In Section 4.2 and 4.3 we will concentrate on the convergence results of these estimators in case of unknown margins being estimated by the empirical distribution functions, thus generalizing the main results in Genest and Segers [2009] to arbitrary dimensions.

In the proofs of the asymptotic normality of the Pickands and CFG estimators, a certain expansion of the empirical copula process due to Stute [1984] and Tsukahara [2005] plays a crucial role. The second-order derivatives of extreme-value copulas typically exhibit explosive behaviour near the corners of the hypercube, violating the assumptions in the two papers just cited. In Segers [2012], it was shown that the same expansion continues to hold under much weaker conditions on the partial derivatives. In Section 4.1, these issues are considered for multivariate extreme-value copulas.

4.1 Empirical copula processes

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be an iid sequence of random vectors from a p-variate multivariate distribution F with continuous margins $F_1, \ldots, F_p$. If the margins $F_1, \ldots, F_p$ are known, we can define the empirical cumulative distribution function C_n of the (unobservable) random sample $\boldsymbol{U}_i = (U_{i,1}, \ldots, U_{i,p}) = (F_1(X_{i,1}), \ldots, F_p(X_{i,p}))$ for $i \in \{1, \ldots, n\}$ by

$$C_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left(U_{i,1} \leqslant u_1, \ldots, U_{i,p} \leqslant u_p\right), \qquad \boldsymbol{u} \in [0,1]^p, \tag{4.1}$$

with associated empirical process

$$\alpha_n = n^{1/2}\,(C_n - C)\,. \tag{4.2}$$

For ease of notation, we will write

$$\alpha_{n,j}(u_j) = \alpha_n(1,\dots,1,u_j,1,\dots,1) \quad \text{for } j \in \{1,\dots,p\}. \tag{4.3}$$

In practice, the marginal distributions will need to be estimated. If we are not willing to make any assumptions (except for continuity) we can estimate them by the empirical distribution functions

$$F_{n,j}(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}(X_{i,j} \leqslant x), \qquad j \in \{1,\dots,p\}, \tag{4.4}$$

where we divided by $n+1$ in order to avoid later problems at the borders. By doing so, we can construct n vectors $\hat{\boldsymbol{U}}_i = (\hat{U}_{i,1},\dots,\hat{U}_{i,p})$ via

$$\hat{U}_{i,j} = F_{n,j}(X_{i,j}) = \frac{1}{n+1}\sum_{l=1}^{n} \mathbf{1}(X_{l,j} \leqslant X_{i,j}) \tag{4.5}$$

for $i \in \{1,\dots,n\}$ and $j \in \{1,\dots,p\}$. The empirical copula will be denoted by

$$\hat{C}_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\hat{U}_{i,1} \leqslant u_1,\dots,\hat{U}_{i,p} \leqslant u_p\right), \qquad \boldsymbol{u} \in [0,1]^p \tag{4.6}$$

with associated empirical copula process

$$\mathbb{C}_n = n^{1/2}\big(\hat{C}_n - C\big). \tag{4.7}$$

In Stute [1984] and Tsukahara [2005] it was established that if all second-order derivatives of C exist and are continuous on $[0,1]^p$, the processes α_n in (4.2) and $\mathbb{C}_n$ in (4.7) are related via

$$\mathbb{C}_n(\boldsymbol{u}) = \alpha_n(\boldsymbol{u}) - \sum_{j=1}^{p} \dot{C}_j(\boldsymbol{u})\,\alpha_{n,j}(u_j) + R_n(\boldsymbol{u}), \tag{4.8}$$

the remainder term R_n satisfying

$$\sup_{\boldsymbol{u}\in[0,1]^p} |R_n(\boldsymbol{u})| = O\big(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\big) \quad \text{almost surely.} \tag{4.9}$$

By classical empirical process theory, we have $\alpha_n \rightsquigarrow \alpha$ as $n \to \infty$, the limiting process being a mean-zero tight Gaussian process with continuous trajectories and covariance function given by

$$\operatorname{cov}\big(\alpha(\boldsymbol{u}), \alpha(\boldsymbol{v})\big) = C(\boldsymbol{u} \wedge \boldsymbol{v}) - C(\boldsymbol{u})\, C(\boldsymbol{v}), \qquad \boldsymbol{u}, \boldsymbol{v} \in [0,1]^p, \tag{4.10}$$

where $(\boldsymbol{u} \wedge \boldsymbol{v})_j = \min(u_j, v_j)$. In view of the expansion (4.8), it then follows that in $\ell^\infty([0,1]^p)$, we have $\mathbb{C}_n \rightsquigarrow \mathbb{C}$ as $n \to \infty$, where

$$\mathbb{C}(u) = \alpha(\boldsymbol{u}) - \sum_{j=1}^{p} \dot{C}_j(\boldsymbol{u})\, \alpha_j(u_j) \tag{4.11}$$

and $\alpha_j(u_j) = \alpha(1, \ldots, 1, u_j, 1, \ldots, 1)$.

Like many other copulas, extreme-value copulas do in general not have uniformly bounded second-order partial derivatives. For instance, in the bivariate case, every copula having a positive coefficient of upper tail dependence will have first-order partial derivatives that fail to have a continuous extension to the upper corner $(1,1)$; see Segers [2012, Example 1.1]. As a consequence, the only bivariate extreme-value copula whose density is uniformly bounded is the independence copula. However, as shown in the same paper, for copulas satisfying Assumption 4.2 below, the expansion (4.8)–(4.9) of the empirical copula process remains valid.

Example 4.1. Let us consider the Gumbel-Hougaard copula presented in Section 1.4.1

$$C(\boldsymbol{u}) = \exp\left\{- \left((-\log u_1)^{1/\alpha} + \ldots + (-\log u_p)^{1/\alpha}\right)^{\alpha}\right\}, \forall \boldsymbol{u} \in (0,1]^p,$$

with $0 < \alpha \leqslant 1$. The partial derivatives of C w.r.t. to the first component evaluated in $\boldsymbol{u} = (0, \boldsymbol{u_{-1}})$, where $\boldsymbol{u_{-1}} = (u_2, \ldots, u_p)$, is defined as

$$\dot{C}_1(0, \boldsymbol{u_{-1}}) = \lim_{h \to 0^+} \frac{C(h, \boldsymbol{u_{-1}})}{h}.$$

Using a first-order Taylor series expansion around $x = 0$ for $f(x) = (1+x)^\alpha$, one can show

$$\dot{C}_1(0, \boldsymbol{u_{-1}}) = \lim_{h\to 0^+} \exp\left\{ -(-\log h)^{1-1/\alpha}\, \alpha \sum_{i=2}^{p} (-\log u_i)^{1/\alpha} \right\} = 1.$$

If at least one component in $\boldsymbol{u_{-1}}$ equals 0, the partial derivative $\dot{C}_1(0, \boldsymbol{u_{-1}})$ equals 0 as $C(h, \boldsymbol{u_1}) = 0$.

Let us consider the derivative in the points $(1, \boldsymbol{u_{-1}})$, where one component belongs to the open interval $(0,1)$. In this case,

$$\begin{aligned} \dot{C}_1(1, \boldsymbol{u_{-1}}) &= \lim_{h\to 0^+} \frac{C(1, \boldsymbol{u_{-1}}) - C(1-h, \boldsymbol{u_{-1}})}{h} \\ &= \lim_{h\to 0^+} \frac{h^{1/\alpha}}{h} = 0. \end{aligned}$$

In the special case, where $\boldsymbol{u_{-1}} = (1, \ldots, 1)$,

$$\dot{C}_1(\mathbf{1}) = \lim_{h\to 0^+} \frac{1 - (1-h)}{h} = 1.$$

□

Assumption 4.2. *(C1) For every $j \in \{1, \ldots, p\}$, the first-order partial derivative $\dot{C}_j$ exists and is continuous on the set $V_{p,j} = \{u \in [0,1]^p : 0 < u_j < 1\}$.*

(C2) For every $i, j \in \{1, \ldots, p\}$ (i and j not necessarily distinct), $\ddot{C}_{i,j}$ exists and is continuous on $V_{p,i} \cap V_{p,j}$ and

$$\sup_{u \in V_{p,i} \cap V_{p,j}} \max\{u_i(1-u_i), u_j(1-u_j)\}\, |\ddot{C}_{i,j}(u)| < \infty.$$

In fact, for weak convergence $\mathbb{C}_n \rightsquigarrow \mathbb{C}$ in $\ell^\infty([0,1]^p)$ to hold, condition (C1) is already sufficient. In the context of multivariate extreme-value copulas, it will be of interest to have a readily verifiable condition on the stable tail dependence function ℓ for Assumption 4.2 to hold.

Assumption 4.3. *(L1) For every $j \in \{1, \ldots, p\}$, the first-order partial derivative $\dot{\ell}_j$ exists and is continuous on the set $W_{p,j} = \{x \in [0,\infty)^p : x_j > 0\}$.*

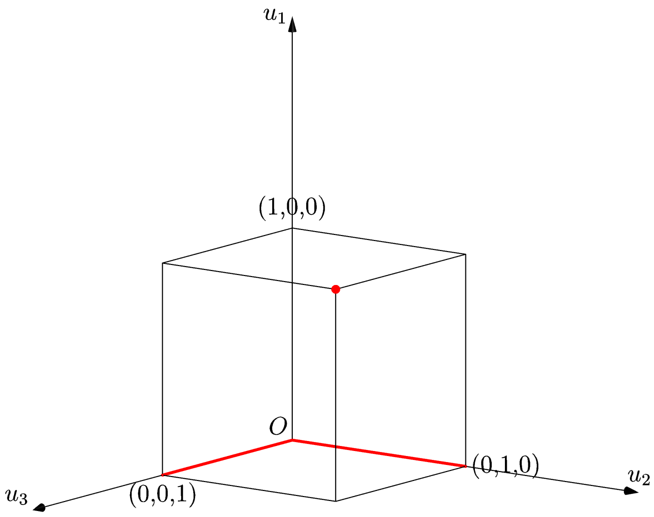


Figure 4.1: Points of discontinuity on the unit-cube for $\dot{C}_1$ with C corresponding to the trivariate logistic copula.

(L2) For every $i, j \in \{1, \dots, p\}$ (i and j not necessarily distinct), $\ddot{\ell}_{i,j}$ exists and is continuous on $W_{p,i} \cap W_{p,j}$ and

$$\sup_{\substack{x \in W_{p,i} \cap W_{p,j} \\ x_1 + \dots + x_p = 1}} \max(x_i, x_j)\, |\ddot{\ell}_{i,j}(x)| < \infty.$$

Proposition 4.4. *For p-variate extreme-value copulas, (L1) implies (C1). If in addition (L2) holds, then (C2) holds as well.*

Example 4.5. Let us show first that condition (L2) is verified by the tail dependence function of the asymmetric logistic copula with the parametrization given in (1.24). We have that

$$\ddot{\ell}_{1,2}(\boldsymbol{x}) = \left(1 - \frac{1}{\alpha}\right) \psi \left(x_1^{1/\alpha} + x_2^{1/\alpha} + x_3^{1/\alpha}\right)^{\alpha-2} x_1^{1/\alpha-1} x_2^{1/\alpha-1}$$
$$+ \left(1 - \frac{1}{\alpha}\right) \left(\theta^{1/\alpha} x_1^{1/\alpha} + \phi^{1/\alpha} x_2^{1/\alpha}\right)^{\alpha-2} \theta^{1/\alpha} \phi^{1/\alpha} x_1^{1/\alpha-1} x_2^{1/\alpha-1}. \quad (4.12)$$

and we will show that

$$\sup_{\substack{x \in W_{3,1} \cap W_{3,2} \\ x_1+x_2+x_3=1}} \max(x_1, x_2)\, |\ddot{\ell}_{1,2}(x)| < \infty.$$

We can split this task into two steps. First, the term involving the three random variables is bounded on the set on which the supremum is taken as

$$\underbrace{\max(x_1, x_2)}_{0<\cdots<1} \; \underbrace{\left(x_1^{1/\alpha} + x_2^{1/\alpha} + x_3^{1/\alpha}\right)^{\alpha-2}}_{\leqslant 3^{\frac{(1-\alpha)\,(2-\alpha)}{\alpha}} \text{ since } x_1+x_2+x_3=1} \; \underbrace{x_1^{1/\alpha-1}\, x_2^{1/\alpha-1}}_{0<\cdots<1}.$$

Without loss of generality, we can suppose that $0 < x_2 < x_1 < 1$, i.e. $\max(x_1, x_2) = x_1$. Then,

$$\max(x_1, x_2) \left(\theta^{1/\alpha} x_1^{1/\alpha} + \phi^{1/\alpha} x_2^{1/\alpha}\right)^{\alpha-2} x_1^{1/\alpha-1} x_2^{1/\alpha-1}$$

is proportional to

$$\frac{\left(\frac{x_2}{x_1}\right)^{1/\alpha-1}}{\left(1 + \left(\frac{\theta}{\phi}\frac{x_2}{x_1}\right)^{1/\alpha}\right)^{2-\alpha}},$$

which is finite as $0 < \frac{x_2}{x_1} < 1$. The same conclusions hold true also for the unconstrained asymmetric logistic model given in (1.23). □

In the bivariate case, sufficient conditions for (C1) and (C2) can be given in terms of the Pickands dependence function $A(w) = \ell(1-w, w)$, where $w \in [0,1]$: (C1) holds as soon as A is continuously differentiable on $(0,1)$, and (C1)–(C2) hold as soon as A is twice continuously differentiable on $(0,1)$ and $\sup_{0<w<1} w(1-w)\, A''(w) < \infty$ [Segers, 2012, Example 5.3].

4.2 Nonparametric estimation of the dependence function

The nonparametric estimators presented in Section 2.1 can also be evaluated in the rank-based observations. Therefore,

$$\hat{\xi}_i(\boldsymbol{w}) = \bigwedge_{j=1}^{p} \frac{-\log \hat{U}_{i,j}}{w_j}.$$

for $\boldsymbol{w} \in \Delta_{p-1}$, with $\hat{\boldsymbol{U}}_{i,j}$ as in (4.5) and the Pickands and CFG-estimators are again

$$\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} = \frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_i(\boldsymbol{w}) \qquad \text{and} \qquad \log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) = -\frac{1}{n}\sum_{i=1}^{n} \log \hat{\xi}_i(\boldsymbol{w}) - \gamma,$$

with $\gamma = 0.5772\ldots$ the Euler–Mascheroni constant.

Opposite to the situation in Chapter 3 where the marginal distributions are known, the endpoint corrections for these estimators are asymptotically irrelevant [Genest and Segers, 2009], since

$$\begin{aligned}
\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{e}_j)} - 1 &= \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{n+1}{i}\right) - 1 = O\left(\frac{\log n}{n}\right), \\
\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{e}_j) &= -\frac{1}{n}\sum_{i=1}^{n} \log\left(\log\left(\frac{n+1}{i}\right)\right) + \int_0^1 \log\left(\log\left(\frac{1}{x}\right)\right) \mathrm{d}x \\
&= O\left(\frac{(\log n)^2}{n}\right).
\end{aligned}$$

as $n \to \infty$. Nevertheless, in finite samples, the simple choice $\lambda_j(\boldsymbol{w}) = w_j$ can make quite a difference. Similarly, for unknown margins, the multivariate extension of the estimator in Hall and Tajvidi [2000] simplifies to $\hat{A}_n^{\mathrm{HT}}(\boldsymbol{w}) = \hat{A}_n^{\mathrm{P}}(\boldsymbol{w})/\hat{A}_n^{\mathrm{P}}(\boldsymbol{e}_j) = \hat{A}_n^{\mathrm{P}}(\boldsymbol{w})\{1 + O(n^{-1}\log n)\}$.

The next lemma establishes a functional relationship between $\hat{A}_n^{\mathrm{P}}$ and $\hat{A}_n^{\mathrm{CFG}}$ on the one hand and the empirical copula $\hat{C}_n$ on the other hand. Recall the empirical copula process $\mathbb{C}_n$ in (4.7).

Lemma 4.6. *For $\boldsymbol{w} \in \Delta_{p-1}$, we have*

$$n^{1/2}\left(\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} - \frac{1}{A(\boldsymbol{w})}\right) = \int_0^1 \mathbb{C}_n(u^{w_1}, \ldots, u^{w_p})\, \frac{\mathrm{d}u}{u}, \tag{4.13}$$

$$n^{1/2}\big(\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - \log A(\boldsymbol{w})\big) = \int_0^1 \mathbb{C}_n(u^{w_1}, \ldots, u^{w_p}) \frac{\mathrm{d}u}{u \log u}, \tag{4.14}$$

with $\mathbb{C}_n$ defined in (4.7)

The proof is not different from the one in dimension two done in Genest and Segers [2009, Lemma 3.1] and can be found in Section 4.3.

Equations (4.13) and (4.14) are instrumental for proving the weak convergence of the processes

$$\mathbb{A}_n^{\mathrm{P}} = n^{1/2}(\hat{A}_n^{\mathrm{P}} - A) \quad \text{and} \quad \mathbb{A}_n^{\mathrm{CFG}} = n^{1/2}(\hat{A}_n^{\mathrm{CFG}} - A).$$

Theorem 4.7. *Under Assumption 4.2 above,* $\mathbb{A}_n^{\mathrm{P}} \rightsquigarrow \mathbb{A}^{\mathrm{P}}$ *and* $\mathbb{A}_n^{\mathrm{CFG}} \rightsquigarrow \mathbb{A}^{\mathrm{CFG}}$ *with*

$$\begin{aligned} \mathbb{A}^{\mathrm{P}}(\boldsymbol{w}) &= -A^2(\boldsymbol{w}) \int_0^1 \mathbb{C}(u^{w_1}, \dots, u^{w_p}) \, \frac{\mathrm{d}u}{u} \\ \mathbb{A}^{\mathrm{CFG}}(\boldsymbol{w}) &= A(\boldsymbol{w}) \int_0^1 \mathbb{C}(u^{w_1}, \dots, u^{w_p}) \frac{\mathrm{d}u}{u \, \log u}, \end{aligned}$$

as $n \to \infty$ *in the space* $\mathscr{C}(\Delta_{p-1})$ *equipped with the topology of uniform convergence and* $\mathbb{C}$*, defined in* (4.11)*, being the weak limit of the process in* (4.7) *.*

The main idea of the proof consists in substituting $\mathbb{C}_n$ in (4.13) and (4.14) by Stute's expansion and to conclude using a refined version of the continuous mapping theorem. The proof follows the same lines as the one in Genest and Segers [2009].

4.3 Proofs

4.3.1 Proof of Proposition 4.4

If $\boldsymbol{u} \in (0,1]^p$, then $-\log \boldsymbol{u} \in W_{p,j}$ and

$$\dot{C}_j(\boldsymbol{u}) = \frac{C(\boldsymbol{u})}{u_j} \, \dot{\ell}_j(-\log \boldsymbol{u}). \tag{4.15}$$

The assumptions on ℓ imply continuity of $\dot{C}_j$ on the set $(0,1]^p$. If $\boldsymbol{u} \in [0,1]^p$ with $u_j > 0$ and $u_i = 0$ for some $i \in \{1, \dots, p\} \backslash \{j\}$, then $\dot{C}_j(u) = 0$ and continuity of $\dot{C}_j$ at such $\boldsymbol{u}$ follows from the fact that $0 \leqslant \dot{\ell}_j \leqslant 1$ and $0 \leqslant C(\boldsymbol{v}) \leqslant \min(\boldsymbol{v})$.

If (L2) holds, then also

$$\sup_{\boldsymbol{x} \in W_{p,i} \cap W_{p,j}} \max(x_i, x_j) \, |\ddot{\ell}_{i,j}(\boldsymbol{x})| < \infty,$$

that is, without the condition $x_1 + \dots + x_p = 1$. This result is based on the fact that the function ℓ is homogeneous of order one: $\ell(s\boldsymbol{x}) = s \, \ell(\boldsymbol{x})$

for all $s \in [0, \infty)$ and $\boldsymbol{x} \in [0, \infty)^p$. Hence if $\dot{\ell}_j$ exists on $W_{p,j}$, then for all $s \in (0, \infty)$ and $\boldsymbol{x} \in W_{p,j}$ we have

$$s\,\dot{\ell}_j(s\boldsymbol{x}) = \frac{\partial}{\partial x_j}\ell(s\boldsymbol{x}) = \frac{\partial}{\partial x_j} s\,\ell(\boldsymbol{x}) = s\,\dot{\ell}_j(\boldsymbol{x})$$

and thus

$$\dot{\ell}_j(s\boldsymbol{x}) = \dot{\ell}_j(\boldsymbol{x}).$$

Taking partial derivatives again, we find for all $s \in (0, \infty)$ and $\boldsymbol{x} \in W_{p,i} \cap W_{p,j}$ that

$$s\,\ddot{\ell}_{i,j}(s\boldsymbol{x}) = \frac{\partial}{\partial x_i}\dot{\ell}_j(s\boldsymbol{x}) = \frac{\partial}{\partial x_i}\dot{\ell}_j(\boldsymbol{x}) = \ddot{\ell}_{i,j}(\boldsymbol{x})$$

and thus

$$\ddot{\ell}_{i,j}(s\boldsymbol{x}) = s^{-1}\,\ddot{\ell}_{i,j}(\boldsymbol{x}).$$

It follows that

$$\max(sx_i, sx_j)\,\ddot{\ell}_{i,j}(s\boldsymbol{x}) = \max(x_i, x_j)\,\ddot{\ell}_{i,j}(\boldsymbol{x}),$$

that is, the map $\boldsymbol{x} \mapsto \max(x_i, x_j)\,\ddot{\ell}_{i,j}(\boldsymbol{x})$ is constant on rays through the origin.

Next, we show the equivalence of (L2) and (C2). Fix $i, j \in \{1, \ldots, p\}$, not necessarily distinct and let $\boldsymbol{u} \in V_{p,i} \cap V_{p,j}$. On the one hand, if $\boldsymbol{u} \in V_{p,i} \cap V_{p,j} \cap (0, 1]^p$ (meaning that every component of $\boldsymbol{u}$ is different from 0), then $-\log \boldsymbol{u} \in W_{p,i} \cap W_{p,j}$ and

$$\ddot{C}_{i,j}(\boldsymbol{u}) = \begin{cases} \dfrac{C(\boldsymbol{u})}{u_i u_j}\left(\dot{\ell}_i\dot{\ell}_j - \ddot{\ell}_{i,j}\right) & \text{if } i \neq j, \\[2ex] \dfrac{C(\boldsymbol{u})}{u_j^2}\left(\dot{\ell}_j^2 - \dot{\ell}_j - \ddot{\ell}_{j,j}\right) & \text{if } i = j, \end{cases}$$

with the convention that the partial derivatives of ℓ are evaluated in $-\log \boldsymbol{u}$. On the other hand, if $\boldsymbol{u} \in (V_{p,i} \cap V_{p,j}) \setminus (0, 1]^p$ (i.e. at least one coordinate of $\boldsymbol{u}$ vanishes), then $\ddot{C}_{i,j}(\boldsymbol{u}) = 0$.

We have to verify two things: first, the continuity of $\ddot{C}_{,ij}$ at points in the set $(V_{p,i} \cap V_{p,j}) \setminus (0, 1]^p$; secondly, the finiteness of the supremum in (C2).

First, let $\boldsymbol{u} \in V_{p,i} \cap V_{p,j} \cap (0,1]^p$. Let K be a positive constant not smaller than the supremum in (L2). By assumption (L2) and the fact that $0 \leqslant \dot{\ell}_j \leqslant 1$, we have

$$|\ddot{C}_{i,j}(\boldsymbol{u})| \leqslant \frac{\min(\boldsymbol{u})}{u_i u_j} \left(1 + \frac{K}{\max(-\log u_i, -\log u_j)} \right).$$

Continuity of $\ddot{C}_{i,j}$ at points in the set $V_{p,i} \cap V_{p,j} \setminus (0,1]^p$ follows.

Secondly, as $\min(\boldsymbol{u})/(u_i u_j) \leqslant \min(1/u_i, 1/u_j)$ and $-\log x \geqslant 1 - x$ for all positive x,

$$\begin{aligned} |\ddot{C}_{i,j}(\boldsymbol{u})| &\leqslant \min\left(\frac{1}{u_i}, \frac{1}{u_j}\right) \left\{ 1 + K \, \min\left(\frac{1}{1-u_i}, \frac{1}{1-u_j}\right) \right\} \\ &\leqslant (1+K) \, \min\left(\frac{1}{u_i}, \frac{1}{u_j}\right) \, \min\left(\frac{1}{1-u_i}, \frac{1}{1-u_j}\right) \\ &\leqslant (1+K) \, \min\left(\frac{1}{u_i(1-u_i)}, \frac{1}{u_j(1-u_j)}\right), \end{aligned}$$

which is equivalent to condition (C2).

4.3.2 Technical results

The proof of Theorem 4.7 will require the following preliminary result on weighted empirical copula processes. Recall the process α_n in (4.2). Define $q_\theta(t) = t^\theta (1-t)^\theta$ for $t \in (0,1)$ and a fixed value $\theta \in (0, 1/2)$. Write $\mathbb{E} = (0,1]^p \setminus \{(1, \ldots, 1)\}$. Define the process $\mathbb{G}_{n,\theta}$ on $[0,1]^p$ by

$$\mathbb{G}_{n,\theta}(\boldsymbol{u}) = \begin{cases} \frac{\alpha_n(\boldsymbol{u})}{q_\theta(\min(\boldsymbol{u}))} & \text{if } \boldsymbol{u} \in \mathbb{E}, \\ 0 & \text{if } \boldsymbol{u} \in [0,1]^p \setminus \mathbb{E}. \end{cases} \tag{4.16}$$

Similarly, define the process $\mathbb{G}_\theta$ on $[0,1]^p$ by replacing α_n in (4.16) by its weak limit α, see (4.10). The following result generalizes Theorem G.1 in Genest and Segers [2009].

Lemma 4.8. *For every $\theta \in (0, 1/2)$, the trajectories of $\mathbb{G}_\theta$ are continuous almost surely and $\mathbb{G}_{n,\theta} \rightsquigarrow \mathbb{G}_\theta$ in $\ell^\infty([0,1]^p)$.*

Proof. Fix $\boldsymbol{u} \in \mathbb{E}$ and define the mapping $f_{\boldsymbol{u}} : \mathbb{E} \to \mathbb{R}$ by

$$f_{\boldsymbol{u}}(\boldsymbol{s}) = \frac{\mathbf{1}_{(\mathbf{0}, \boldsymbol{u}]}(\boldsymbol{s}) - C(\boldsymbol{u})}{q_\theta(\min(\boldsymbol{u}))}, \qquad \boldsymbol{s} \in \mathbb{E},$$

and consider the class

$$\mathscr{F} = \{f_{\boldsymbol{u}} : \boldsymbol{u} \in \mathbb{E}\} \cup \{0\},$$

where 0 of course stands for the zero function. The space $\mathscr{F}$ will be endowed with the metric

$$\rho^2(f,g) = \mathrm{P}(f-g)^2 \qquad f,g \in \mathscr{F}. \tag{4.17}$$

Here, we adopt the notations of van der Vaart and Wellner [1996]: P denotes the probability distribution on $\mathbb{E}$ corresponding to C and $\mathbb{P}_{\mathrm{n}}$ denotes the empirical measure of the sample $(U_{i1},\ldots,U_{ip})$ for $i \in \{1,\ldots,n\}$, that is

$$\mathrm{P}\,f = \int f\,\mathrm{d}C, \qquad \mathbb{P}_{\mathrm{n}}\,f = \frac{1}{n}\sum_{i=1}^{n} f(U_{i,1},\ldots,U_{i,p}).$$

Moreover, put $\mathbb{G}_n = n^{1/2}(\mathbb{P}_{\mathrm{n}} - \mathrm{P})$, viewed as a random function on $\mathscr{F}$, i.e.

$$\begin{array}{llll} \mathbb{G}_n : & (\mathscr{F},\rho) & \to & (\mathbb{R}, |\cdot|) \\ & f_{(u_1,\ldots,u_p)} & \mapsto & \mathbb{G}_n f_{(u_1,\ldots,u_p)} \\ & & & = n^{1/2}(\mathbb{P}_{\mathrm{n}} - \mathrm{P}) f_{(u_1,\ldots,u_p)}. \end{array}$$

Proof. It is sufficient to verify the conditions of Theorem 2.6.14 of van der Vaart and Wellner [1996].

a) $\mathscr{F}$ is a VC-major class.

b) There exists $F : \mathbb{E} \to \mathbb{R}$ such that $|f| \leqslant F$ pointwise for all $f \in \mathscr{F}$ and
$$\int_0^{+\infty} \{\mathrm{P}(F > x)\}^{1/2} dx < \infty,$$
where $\mathrm{P}(F > x) = \mathrm{P}\Big(\boldsymbol{s} = (s_1,\ldots,s_p) \in \mathbb{E} : F(\boldsymbol{s}) > x\Big)$.

c) $\mathscr{F}$ is pointwise separable.

To prove a), we need to check that the class of subsets of $\mathbb{E}$ given by

$$\Big\{ \{\boldsymbol{s} \in \mathbb{E} : f_{\boldsymbol{u}}(\boldsymbol{s}) > t\} : f_{\boldsymbol{u}} \in \mathscr{F}, t \in \mathbb{R} \Big\} \tag{4.18}$$

is a VC-class (Vapnik-Cervonenkis class); see page 145 of van der Vaart and Wellner [1996].

In fact each $f_{\boldsymbol{u}}$ can take only 2 values

$$f_{\boldsymbol{u}}(\boldsymbol{s}) = \begin{cases} \frac{1-C(u_1,\ldots,u_p)}{q_\theta(u_1\wedge\ldots\wedge u_p)} > 0 & s_1 < u_1, \ldots, s_p < u_p \\ -\frac{C(u_1,\ldots,u_p)}{q_\theta(u_1\wedge\ldots\wedge u_p)} < 0 & \text{elsewhere.} \end{cases}$$

This implies that the class (4.18) corresponds to the family of the open intervals $(0, u_1] \times \ldots \times (0, u_p]$, which is indeed a VC-class; see page 135 of van der Vaart and Wellner [1996].

For point b), we will apply the Fréchet-Hoeffding bounds to show

$$C(\boldsymbol{u}) \leqslant u_1 \wedge \cdots \wedge u_p, \quad \forall \boldsymbol{u} = (u_1, \ldots, u_p) \in [0,1]^p \tag{4.19}$$

and

$$1 - C(\boldsymbol{u}) \leqslant p(1 - u_1 \wedge \ldots \wedge u_p). \tag{4.20}$$

The last inequality is established as follows

$$1 - C(\boldsymbol{u}) \leqslant 1 - \max\{1 - p + \sum_{i=1}^{p} u_i, 0\} \leqslant \begin{cases} 1, & \sum_{i=1}^p u_i \leqslant p - 1 \\ p - \sum_{i=1}^p u_i & \text{elsewhere,} \end{cases}$$

such that

$$1 - C(\boldsymbol{u}) \leqslant \sum_{i=1}^{p} |1 - u_i| \leqslant p(1 - u_1 \wedge \ldots \wedge u_p).$$

In the sequel we distinguish between two cases:

Case 1: $\{s_1 \leqslant u_1, \ldots, s_p \leqslant u_p\}$

$$\begin{aligned} f_{\boldsymbol{u}}(s_1, \ldots, s_p) &= \frac{1 - C(\boldsymbol{u})}{q_\theta(u_1 \wedge \cdots \wedge u_p)} \\ &\leqslant \frac{p(1 - u_1 \wedge \cdots \wedge u_p)}{(u_1 \wedge \cdots \wedge u_p)^\theta (1 - u_1 \wedge \cdots \wedge u_p)^\theta} \\ &= p(u_1 \wedge \cdots \wedge u_p)^{-\theta}(1 - u_1 \wedge \cdots \wedge u_p)^{1-\theta} \\ &\leqslant p(u_1 \wedge \cdots \wedge u_p)^{-\theta} \\ &= p(u_1^{-\theta} \vee \cdots \vee u_p^{-\theta}) \\ &\leqslant p(s_1^{-\theta} \vee \cdots \vee s_p^{-\theta}). \end{aligned}$$

Case 2: $\exists s_j$ s.t. $s_j > u_j$

$$\begin{aligned}
|f_{\boldsymbol{u}}(s_1,\ldots,s_p)| &= \frac{C(\boldsymbol{u})}{q_\theta(u_1\wedge\cdots\wedge u_p)} \\
&\leqslant \frac{u_1\wedge\cdots\wedge u_p}{(u_1\wedge\cdots\wedge u_p)^\theta(1-u_1\wedge\cdots\wedge u_p)^\theta} \\
&= (u_1\wedge\cdots\wedge u_p)^{1-\theta}(1-u_1\wedge\cdots\wedge u_p)^{-\theta} \\
&\leqslant (1-u_1\wedge\cdots\wedge u_p)^{-\theta}.
\end{aligned}$$

By assumption we have

$$u_1\wedge\cdots\wedge u_p \leqslant s_1\vee\cdots\vee s_p$$

implying that

$$\begin{aligned}
(1-u_1\wedge\cdots\wedge u_p)^{-\theta} &\leqslant (1-s_1\vee\cdots\vee s_P)^{-\theta} \\
&= (1-s_1)^{-\theta}\vee\cdots\vee(1-s_p)^{-\theta} \\
&\leqslant p\left[(1-s_1)^{-\theta}\vee\cdots\vee(1-s_p)^{-\theta}\right].
\end{aligned}$$

Finally, a candidate for the envelope function is found by regrouping the previous inequalities

$$F(s_1,\ldots,s_p)\leqslant p\left[s_1^{-\theta}\vee\cdots\vee s_p^{-\theta}\vee(1-s_1)^{-\theta}\vee\cdots\vee(1-s_p)^{-\theta}\right]$$

verifying the integral condition in point b) as

$$\begin{aligned}
& \mathrm{P}((s_1,\ldots,s_p):F(s_1,\ldots,s_p)>x) \\
\leqslant\ & \mathrm{P}\left((s_1,\ldots,s_p):p\left[s_1^{-\theta}\vee\cdots\vee s_p^{-\theta}\vee(1-s_1)^{-\theta}\vee\cdots\vee(1-s_p)^{-\theta}\right]>x\right) \\
\leqslant\ & \left[\mathrm{P}\left(s_1<\left(\frac{x}{p}\right)^{-1/\theta}\right)+\cdots+\mathrm{P}\left(s_p<\left(\frac{x}{p}\right)^{-1/\theta}\right)\right. \\
& \left.+\mathrm{P}\left((1-s_1)<\left(\frac{x}{p}\right)^{-1/\theta}\right)+\cdots+\mathrm{P}\left((1-s_p)<\left(\frac{x}{p}\right)^{-1/\theta}\right)\right]\wedge 1 \\
\leqslant\ & 2p\left(\frac{x}{p}\right)^{-1/\theta}\wedge 1.
\end{aligned}$$

We can conclude that for large values of x the probability given above will be proportional to $x^{-1/\theta}$ with $\theta \in (0, 1/2)$ and hence integrable on $(0, +\infty)$.

For point c), let $\mathbb{E}_0$ be a countable subset of $\mathbb{E}$ and $\mathscr{F}_{\mathbb{E}_0}$ be the subset of $\mathscr{F}$ generated by $\mathbb{E}_0$. Clearly every $f_{\boldsymbol{u}}$ in $\mathscr{F}$ can be written as a pointwise limit of elements $f_{\boldsymbol{u}_n}$ of elements in $\mathscr{F}_{\mathbb{E}_0}$. As there exists a P-square integrable envelope function, the dominated convergence theorem guarantees convergence in $L^2(\mathrm{P})$. By the definition on page 116 of van der Vaart and Wellner [1996], $\mathscr{F}$ is a pointwise separable class. $\square$

For the moment $\mathbb{G}_n$ is defined on $\mathscr{F}$ with the metric ρ though we would prefer to consider $\mathbb{G}_n$ as a function on $[0,1]^p$ with the more familiar Euclidean norm $\|\cdot\|_2$. This shift will be done by the map:

$$\begin{array}{rlcl} \phi: & ([0,1]^p, \|\cdot\|_2) & \to & (\mathscr{F}, \rho) \\ & \boldsymbol{u} & \mapsto & f_{\boldsymbol{u}} \end{array}$$

and the empirical process

$$\begin{array}{rlcl} \mathbb{G}_{n,\theta}: & ([0,1]^p, \|\cdot\|_2) & \to & (\mathbb{R}, |\cdot|) \\ & \boldsymbol{u} & \mapsto & \mathbb{G}_n \circ \phi(\boldsymbol{u}) \\ & & & = n^{1/2}\,(\mathbb{P}_\mathrm{n} - \mathbb{P})\, f_{\boldsymbol{u}}. \end{array}$$

The subtle difference between $\mathbb{G}_n$ and $\mathbb{G}_{n,\theta}$ consists in their different domains of definitions with their respective topologies. Let us define

$$\begin{array}{rlcl} T: & (\ell^\infty(\mathscr{F}), \|\cdot\|_{\ell^\infty(\mathscr{F})}) & \to & (\ell^\infty([0,1]^p), \|\cdot\|_\infty) \\ & g & \mapsto & g \circ \phi, \end{array} \tag{4.21}$$

where in particular $\mathbb{G}_{n,\theta} = T(\mathbb{G}_n)$. $\square$

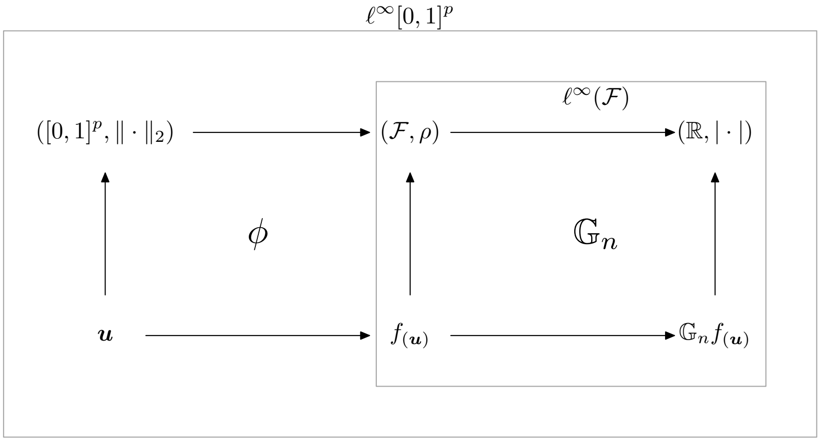
Lemma 4.9.

$$\mathbb{G}_{n,\theta} \rightsquigarrow \mathbb{G}_\theta \quad in \quad \mathscr{C}([0,1]^p).$$

Proof. From the previous lemma we know that

$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \quad \text{in} \quad \ell^\infty(\mathscr{F}),$$

and by definition T defined in (4.21) is a linear map and bounded (with constant 1), hence T is continuous.



The Continuous Mapping theorem permits to conclude that

$$\mathbb{G}_{n,\theta} \rightsquigarrow \mathbb{G}_\theta \quad \text{in} \quad \ell^\infty([0,1]^p).$$

It remains to show that the sample paths:

$$(u_1, \ldots, u_p) \mapsto G_\theta(u_1, \ldots, u_p)$$

belong to $\mathscr{C}([0,1]^p)$ almost surely.

We know that the map

$$\begin{array}{rccc} \mathbb{G}: & (\mathscr{F}, \rho) & \to & (\mathbb{R}, |\cdot|) \\ & f_{\boldsymbol{u}} & \mapsto & \mathbb{G} f_{\boldsymbol{u}} \end{array}$$

has continuous sample paths w.r.t. to the metric ρ defined on $\mathscr{F}$, see page 41 of van der Vaart and Wellner [1996].

We will establish that ϕ as defined in (4.3.2) is a continuous map, i.e.

$$\rho(f_{\boldsymbol{u}}, f_{\boldsymbol{v}})^2 = \mathrm{P}\, f_{\boldsymbol{u}}^2 + \mathrm{P}\, f_{\boldsymbol{v}}^2 - 2\,\mathrm{P}\, f_{\boldsymbol{u}} f_{\boldsymbol{v}}$$

tends to 0 as $\|\boldsymbol{u}-\boldsymbol{v}\|_2$ tends to 0 for every $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{E}_0$. Explicit expressions for $\mathrm{P}\, f_{\boldsymbol{u}}^2$ and $\mathrm{P}\, f_{\boldsymbol{u}} f_{\boldsymbol{v}}$ are given below.

Fix $\boldsymbol{u} = (u_1, \ldots, u_p)$ and $\boldsymbol{v} = (v_1, \ldots, v_p)$ in $\mathbb{E}$, then

$$\begin{aligned}
\mathrm{P}\, f_{\boldsymbol{u}} f_{\boldsymbol{v}} &= \int \cdots \int_{[0,1]^p} \frac{\mathbf{1}_{(0,u_1]\times\cdots\times(0,u_p]}(\boldsymbol{s}) - C(\boldsymbol{u})}{q_\theta(u_1 \wedge \cdots \wedge u_p)} \\
&\quad \frac{\mathbf{1}_{(0,v_1]\times\cdots\times(0,v_p]}(\boldsymbol{s}) - C(\boldsymbol{v})}{q_\theta(v_1 \wedge \cdots \wedge v_p)} \,\mathrm{d}C(\boldsymbol{s}) \\
&= \frac{C(u_1 \wedge v_1, \ldots, u_p \wedge v_p) - C(\boldsymbol{v})\, C(\boldsymbol{u})}{q_\theta(u_1 \wedge \cdots \wedge u_p)\, q_\theta(v_1 \wedge \cdots \wedge v_p)}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{P}\, f_{\boldsymbol{u}}^2 &= \int \cdots \int_{[0,1]^p} \left(\frac{\mathbf{1}_{(0,u_1]\times\cdots\times(0,u_p]}(\boldsymbol{s}) - C(\boldsymbol{u})}{q_\theta(u_1 \wedge \cdots \wedge u_p)} \right)^2 \mathrm{d}C(\boldsymbol{s}) \\
&= \int \cdots \int_{(0,u_1]\times\cdots\times(0,u_p]} \left(\frac{1 - C(\boldsymbol{u})}{q_\theta(u_1 \wedge \cdots \wedge u_p)} \right)^2 \mathrm{d}C(\boldsymbol{s}) \\
&+ \int \cdots \int_{[0,1]^p \setminus [0,\boldsymbol{u}]} \left(\frac{C(\boldsymbol{u})}{q_\theta(u_1 \wedge \cdots \wedge u_p)} \right)^2 \mathrm{d}C(\boldsymbol{s}) \\
&= \frac{C(\boldsymbol{u})^2 (1 - C(\boldsymbol{u})) + C(\boldsymbol{u})(1 - C(\boldsymbol{u}))^2}{q_\theta(u_1 \wedge \cdots \wedge u_p)^2} \\
&= \frac{C(\boldsymbol{u})(1 - C(\boldsymbol{u}))}{q_\theta(u_1 \wedge \cdots \wedge u_p)^2}.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\rho(f_{\boldsymbol{u}}, f_{\boldsymbol{v}})^2 \\
=& \left(\frac{C(u_1, \ldots, u_p)}{q_\theta(u_1 \wedge \cdots \wedge u_p)} - \frac{C(u_1 \wedge v_1, \ldots, u_p \wedge v_p)}{q_\theta(v_1 \wedge \cdots \wedge v_p)} \right) \frac{1}{q_\theta(u_1 \wedge \cdots \wedge u_p)} \\
+& \left(\frac{C(v_1, \ldots, v_p)}{q_\theta(v_1 \wedge \cdots \wedge v_p)} - \frac{C(u_1 \wedge v_1, \ldots, u_p \wedge v_p)}{q_\theta(v_1 \wedge \cdots \wedge v_p)} \right) \frac{1}{q_\theta(v_1 \wedge \cdots \wedge v_p)} \\
-& \left(\frac{C(u_1, \ldots, u_p)}{q_\theta(u_1 \wedge \cdots \wedge u_p)} - \frac{C(v_1, \ldots, v_p)}{q_\theta(v_1 \wedge \cdots \wedge v_p)} \right)^2,
\end{aligned}$$

which tends to 0 as $\boldsymbol{u} \to \boldsymbol{v}$ in $\mathbb{E}$. The continuity of ϕ in the cases where $u_1 \wedge \ldots \wedge u_p$ tends to either 0 or 1, is assured by the inequalities defined in (4.19) and (4.20)

$$\begin{aligned}
\rho^2(f_{\boldsymbol{u}}, \mathbf{0}) &= \mathrm{P}\, f_{\boldsymbol{u}}^2 \\
&= \frac{C(u_1, \ldots, u_p)(1 - C(u_1, \ldots, u_p))}{q_\theta(u_1 \wedge \cdots \wedge u_p)^2} \\
&\leqslant (u_1 \wedge \cdots \wedge u_p)^{1-2\theta}\, (1 - u_1 \wedge \cdots \wedge u_p)^{1-2\theta},
\end{aligned}$$

which tends to 0 in any of the two cases. □

4.3.3 Proof of Lemma 4.6

Proof. For a fixed value of $\boldsymbol{w} \in \Delta_{p-1}$ and a sample of rank-based observation $\hat{\boldsymbol{U}}_{\boldsymbol{i}} = (\hat{U}_{i,1}, \ldots, \hat{U}_{i,p})$ for $i \in \{1, \ldots, n\}$ constructed as explained in (4.5), we have

$$\hat{\xi}_i(\boldsymbol{w}) = \int_0^{\hat{\xi}_i(\boldsymbol{w})} dx = \int_0^{+\infty} \mathbf{1}(\hat{\xi}_i(\boldsymbol{w}) \geqslant x)\, dx$$

The Pickands estimator can then be written as follows

$$\begin{aligned}
\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} &= \frac{1}{n} \sum_{i=1}^n \hat{\xi}_i(\boldsymbol{w}) \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^{+\infty} \mathbf{1}\left(\hat{\xi}_i(\boldsymbol{w}) \geqslant x\right)\, dx \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^{+\infty} \mathbf{1}\left(-\frac{\log \hat{U}_{i,1}}{w_j} \geqslant x, \ldots, -\frac{\log \hat{U}_{i,p}}{w_j} \geqslant x\right)\, dx \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^{+\infty} \mathbf{1}\left(\hat{U}_{i,1} \leq e^{-w_1 x}, \ldots, \hat{U}_{i,p} \leq e^{-w_p x}\right)\, dx.
\end{aligned}$$

Applying the change of variables $u = e^{-x}$, we have

$$\begin{aligned}
\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} &= \frac{1}{n} \sum_{i=1}^n \int_0^1 \mathbf{1}\left(\hat{U}_{i,1} \leq u^{w_1}, \ldots, \hat{U}_{i,p} \leq u^{w_p}\right)\, \frac{du}{u} \\
&= \int_0^1 \hat{C}_n\left(u^{w_1}, \ldots, u^{w_p}\right)\, \frac{du}{u}.
\end{aligned}$$

For the CFG-estimator, the main trick consists in rewriting the logarithm function in the following way

$$\begin{aligned} \log(z) &= \int_0^{+\infty} \{\mathbf{1}(x \leq z) - \mathbf{1}(x \leq 1)\} \frac{dx}{x} \\ &= -\int_0^1 \{\mathbf{1}(-\log u \leq z) - \mathbf{1}(u \geqslant e^{-1})\} \frac{du}{u \log u}. \end{aligned}$$

Based on this integral representation of the logarithm, the CFG-estimator can be reexpressed as follows

$$\begin{aligned} &-\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - \gamma \\ &= \frac{1}{n} \sum_{i=1}^n \log \xi_i(\boldsymbol{w}) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^1 \left[\mathbf{1}(u \geqslant e^{-1}) - \mathbf{1}(U_{i1} \leq u^{w_1}, \dots, U_{ip} \leq u^{w_p})\right] \frac{du}{u \log u} \\ &= \int_0^1 \left[\mathbf{1}(u \geqslant e^{-1}) - \hat{C}_n(u^{w_1}, \dots, u^{w_p})\right] \frac{du}{u \log u} \end{aligned}$$

Rearranging the previous expression, we obtain the stated result (4.14)

$$\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) = \int_0^1 \left[\hat{C}_n(u^{w_1}, \dots, u^{w_p}) - \mathbf{1}(u \geqslant e^{-1})\right] \frac{du}{u \log u} - \gamma.$$

□

4.3.4 Proof of Theorem 4.7

We now proceed with the proof of Theorem 4.7. Define

$$\begin{aligned} B_n^{\mathrm{P}}(\boldsymbol{w}) &= n^{1/2} \left(\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} - \frac{1}{A(\boldsymbol{w})} \right), \\ B_n^{\mathrm{CFG}}(\boldsymbol{w}) &= n^{1/2} \big(\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - \log A(\boldsymbol{w})\big), \end{aligned}$$

for $\boldsymbol{w} \in \Delta_{p-1}$. Applying the change of variables $u = e^{-s}$ in Lemma 4.6, we find that the processes B_n^{P} and B_n^{CFG} can be written as

$$B_n(\boldsymbol{w}) = \int_0^\infty \mathbb{C}_n(e^{-w_1 s}, \dots, e^{-w_p s})\, h(s)\, \mathrm{d}s, \tag{4.22}$$

in terms of a function h on $(0,\infty)$ which is $h^{\mathrm{P}}(s) = 1$ for the Pickands estimator and $h^{\mathrm{CFG}}(s) = 1/s$ for the CFG estimator. In what follows, the function h denotes either h^{P} or h^{CFG}.

Put $l_n = 1/(n+1)$ and $k_n = p\log(n+1)$ and split the integral on the right-hand side of (4.22) into three parts:

$$B_n(\boldsymbol{w}) = \int_0^{l_n} + \int_{l_n}^{k_n} + \int_{k_n}^{\infty} = I_{1,n}(\boldsymbol{w}) + I_{2,n}(\boldsymbol{w}) + I_{3,n}(\boldsymbol{w}). \tag{4.23}$$

We will first prove that with probability one, the first and the third term on the right-hand side converge to zero uniformly in $\boldsymbol{w}$.

- If $s \in [0, l_n]$, then $C_n(e^{-w_1 s}, \ldots, e^{-w_p s}) = 1$ since for all $\boldsymbol{w} = (w_1, \ldots, w_p) \in \Delta_{p-1}$

  $$e^{-w_1 s} \wedge \ldots \wedge e^{-w_p s} \geq e^{-s} \geq 1 - s$$

  by the convexity of the function $x \mapsto e^{-x}$ and

  $$1 - s \geq \frac{n}{n+1} \quad \text{iff} \quad s \leq \frac{1}{n+1}.$$

  This implies

  $$\begin{aligned} 0 \leqslant I_{1,n}(\boldsymbol{w}) &= \int_0^{l_n} n^{1/2}\bigl(1 - e^{-s\,A(\boldsymbol{w})}\bigr)\, h(s)\,\mathrm{d}s \\ &\leqslant n^{1/2} \int_0^{l_n} s\,h(s)\,ds \leqslant n^{1/2} l_n \leqslant n^{-1/2}. \end{aligned}$$

- If $s \in (k_n, \infty)$, then $w_j \geqslant 1/p$ and thus $e^{-w_j s} < 1/(n+1)$ for at least one $j \in \{1, \ldots, p\}$, so that $C_n(e^{-w_1 s}, \ldots, e^{-w_p s}) = 0$, which implies

  $$\begin{aligned} |I_{3,n}(\boldsymbol{w})| &\leqslant \int_{k_n}^{\infty} n^{1/2} e^{-s\,A(\boldsymbol{w})}\, h(s)\,\mathrm{d}s \\ &\leqslant \frac{n^{1/2}}{A(\boldsymbol{w})} e^{-k_n\,A(\boldsymbol{w})} \leqslant p\,n^{-1/2}, \end{aligned}$$

  where we used the fact that $A(\boldsymbol{w}) \geqslant \max(\boldsymbol{w}) \geqslant 1/p$.

As a consequence, the only non-negligible term in (4.23) is $I_{2,n}$. By Assumption 4.2 and by Proposition 4.2 in Segers [2012], Stute's expansion (4.8)–(4.9) is valid, so that we can write

$$I_{n,2}(\boldsymbol{w}) = J_{0,n}(\boldsymbol{w}) - \sum_{j=1}^{p} J_{j,n}(\boldsymbol{w}) + J_{p+1,n}(\boldsymbol{w})$$

where

$$\begin{aligned} J_{0,n}(\boldsymbol{w}) &= \int_{l_n}^{k_n} \alpha_n(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\,\mathrm{d}s, \\ J_{j,n}(\boldsymbol{w}) &= \int_{l_n}^{k_n} \alpha_{n,j}(e^{-w_j s})\, \dot{C}_j(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\,\mathrm{d}s, \qquad j \in \{1, \ldots, p\}, \\ J_{p+1,n}(\boldsymbol{w}) &= \int_{l_n}^{k_n} R_n(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\,\mathrm{d}s. \end{aligned}$$

In view of the bound (4.9) on R_n, the term $J_{p+1,n}$ is negligible: as $n \to \infty$,

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}} |J_{p+1,n}(\boldsymbol{w})| = O\big(n^{-1/4} \log(n) \textstyle\int_{l_n}^{k_n} h(s)\,\mathrm{d}s\big) \to 0, \qquad \text{almost surely.}$$

Fix $\theta \in (0, 1/2)$ and recall the process $\mathbb{G}_{n,\theta}$ in (4.16). We have

$$\begin{aligned} J_{0,n}(\boldsymbol{w}) &= \int_{l_n}^{k_n} \mathbb{G}_{n,\theta}(e^{-sw_1}, \ldots, e^{-sw_p})\, K_0(s, \boldsymbol{w})\, h(s)\,\mathrm{d}s, \\ J_{j,n}(\boldsymbol{w}) &= \int_{l_n}^{k_n} \mathbb{G}_{n,\theta}(1, \ldots, 1, e^{-sw_j}, 1, \ldots, 1)\, K_j(s, \boldsymbol{w})\, h(s)\,\mathrm{d}s, \end{aligned}$$

for $j \in \{1, \ldots, p\}$, with

$$\begin{aligned} K_0(s, \boldsymbol{w}) &= q_\theta\big(\min(e^{-sw_1}, \ldots, e^{-sw_p})\big), \\ K_j(s, \boldsymbol{w}) &= q_\theta(e^{-sw_j})\, \dot{C}_j(e^{-sw_1}, \ldots, e^{-sw_p}), \qquad j \in \{1, \ldots, p\}. \end{aligned}$$

The functions $K_0, \ldots, K_p$ satisfy the bounds

$$\begin{aligned} 0 \leqslant K_j(s, \boldsymbol{w}) \leqslant K(s) = s^\theta\, \mathbf{1}_{(0,1]}(s) + e^{-(\theta/p)s}\, \mathbf{1}_{(1,\infty)}(s), \\ j \in \{0, \ldots, p\}, \qquad s \in (0, \infty), \qquad \boldsymbol{w} \in \Delta_{p-1}. \quad (4.24) \end{aligned}$$

To prove these bounds, use equation (4.15), the fact that $0 \leqslant \dot{\ell}_j \leqslant 1$ and $0 \leqslant C(\boldsymbol{v}) \leqslant \min(\boldsymbol{v})$ for $\boldsymbol{v} \in [0,1]^p$ and the fact that $\max(\boldsymbol{w}) \geqslant 1/p$ for $\boldsymbol{w} \in \Delta_{p-1}$.

In fact, if $s \in (1, +\infty)$,

$$\begin{aligned} & q_\theta(e^{-s\,w_1} \wedge \ldots \wedge e^{-s\,w_p}) \\ = \; & (e^{-s\,w_1} \wedge \ldots \wedge e^{-s\,w_p})^\theta \,(1 - e^{-s\,w_1} \wedge \ldots \wedge e^{-s\,w_p})^\theta \\ \leqslant \; & (e^{-s\,w_1} \wedge \ldots \wedge e^{-s\,w_p})^\theta \\ \leqslant \; & e^{-s\theta/p} \end{aligned}$$

and if $s \in (0,1)$,

$$\begin{aligned} q_\theta(e^{-s\,w_1} \wedge \ldots \wedge e^{-s\,w_p}) \; & \leqslant \; (1 - e^{-s\,w_1} \wedge \ldots \wedge e^{-s\,w_p})^\theta \\ & \leqslant \; (sw_1 \vee \ldots \vee sw_p)^\theta \leqslant s^\theta. \end{aligned}$$

Therefore, we define for each $j \in \{1, \ldots, p\}$ the map $T_{j,n}$:

$$\begin{array}{rcccl} T_{j,n} & : & (\ell^\infty(\Delta_{p-1}), \|\cdot\|_\infty) & \to & (\mathbb{R}, |\cdot|) \\ & & f & \mapsto & \int_0^{p\log(n+1)} f(e^{-sw_1}, \ldots, e^{-sw_p})\, K_j^P(s, \boldsymbol{w})\, \mathrm{d}s \end{array}$$

As $T_{j,n}$ is a linear application, it is sufficient to show that it is also bounded to establish continuity. We have that

$$|T_{j,n} f| \quad \leqslant \quad \|f\|_\infty \int_0^{p\log(n+1)} K_j(s, \boldsymbol{w})\, \mathrm{d}s$$

Letting n tend to infinity, this reduces to show that $\int_0^\infty K_j^P(s, \boldsymbol{w})\, \mathrm{d}s$ can be bounded by a constant for every $j \in \{1, \ldots, p\}$, which is guaranteed by the finiteness of the integral $\int_0^\infty K(s)\, h(s)\, \mathrm{d}s$ for K defined in (4.24)
.

By Lemma 4.8 and the extended continuous mapping theorem [van der Vaart and Wellner, 1996, Theorem 1.11.1], we find

$$B_n \rightsquigarrow B = J_0 - \sum_{j=1}^p J_j, \qquad n \to \infty$$

in $\ell^\infty(\Delta_{p-1})$, where

$$\begin{aligned} J_0(\boldsymbol{w}) &= \int_0^\infty \mathbb{G}_\theta(e^{-sw_1}, \ldots, e^{-sw_p})\, K_0(s, \boldsymbol{w})\, h(s)\, \mathrm{d}s, \\ J_j(\boldsymbol{w}) &= \int_0^\infty \mathbb{G}_\theta(1, \ldots, 1, e^{-sw_j}, 1, \ldots, 1)\, K_j(s, \boldsymbol{w})\, h(s)\, \mathrm{d}s \end{aligned}$$

with $j \in \{1, \ldots, p\}$.

From the definitions of the process $\mathbb{G}_\theta$ and the functions $K_0, \ldots, K_p$, we obtain

$$B(\boldsymbol{w}) = \int_0^\infty \mathbb{C}(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\, \mathrm{d}s, \qquad \boldsymbol{w} \in \Delta_{p-1}.$$

An application of the functional delta method [van der Vaart and Wellner, 1996, Theorem 3.9.4] now yields the result. From Lemma 4.9, we know that $\mathbb{G}_{n,\theta} \rightsquigarrow \mathbb{G}_\theta$ in $\mathscr{C}([0,1]^p)$. As the integration domain depends on n, we will refer to extended continuous mapping theorem stated at page 67 in van der Vaart and Wellner [1996].

4.4 Additional Remark

Let $E_{n,j}$ denote the j-th empirical distribution normalized with respect to n as opposed to $F_{n,j}$ defined in (4.4), i.e. for a real x

$$E_{n,j}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\left(X_{i,1} \leqslant x\right), \quad j \in \{1, \ldots, p\}. \tag{4.25}$$

In Deheuvels [1979], the empirical copula is defined as

$$\begin{aligned} \hat{C}_n^D(\boldsymbol{u}) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}\left(X_{i,1} \leqslant E_{n,1}^{\leftarrow}(u_1), \ldots, X_{i,p} \leqslant E_{n,p}^{\leftarrow}(u_p)\right) \\ &= F_n\big(E_{n,1}^{\leftarrow}(u_1), \ldots, E_{n,p}^{\leftarrow}(u_p)\big), \end{aligned}$$

with $E_{n,j}^{\leftarrow}(u_j) = \inf\{x_j \in \mathbb{R} : E_{n,j}(x_j) \geqslant u_j\}$.

Straightforward calculus shows that, in the absence of ties,

$$\sup_{\boldsymbol{u} \in [0,1]^p} |\hat{C}_n^D(\boldsymbol{u}) - \hat{C}_n(\boldsymbol{u})| \leqslant \frac{2p}{n}. \tag{4.26}$$

First, we remark that by the discontinuity of $E_{n,j}$ and $E_{n,j}^{\leftarrow}$, we have that:

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \mathbf{1}\Big(X_{i1} \leqslant E_{n,1}^{\leftarrow}(u_1), \ldots, X_{ip} \leqslant E_{n,p}^{\leftarrow}(u_p)\Big) \\ \leqslant &\frac{1}{n} \sum_{i=1}^n \mathbf{1}\Big(E_{n,1}(X_{i1}) \leqslant (u_1), \ldots, E_{n,p}(X_{ip}) \leqslant u_p\Big) + \frac{p}{n}. \end{aligned}$$

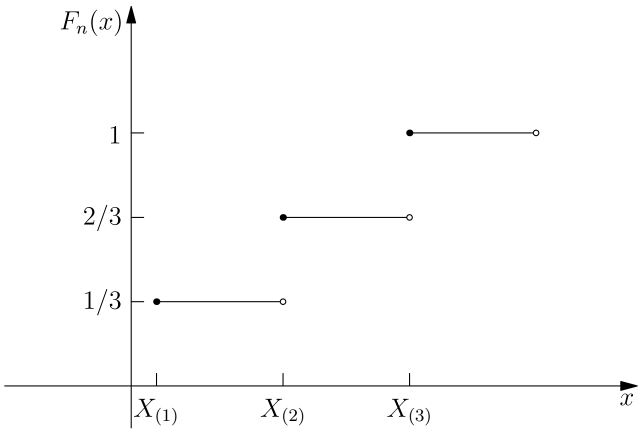


Figure 4.2: We observe that $F_n(X_1) \leqslant 1/2$ but $F_n(X_{(2)}) \nleqslant 1/2$.

Then, the difference $|\hat{C}_n(u_1, \ldots, u_p) - \hat{C}_n^D(u_1, \ldots, u_p)|$ can be maximized by

$$\Big| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1} \left(E_{n,1}(X_{i,1}) \leqslant \frac{n+1}{n} u_1, \ldots, E_{n,p}(X_{i,p}) \leqslant \frac{n+1}{n} u_p \right) - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1} \left(E_{n,1}(X_{i,1}) \leqslant u_1, \ldots, E_{n,p}(X_{i,p}) \leqslant u_p \right) \Big| + \frac{p}{n}$$

The result follows from a telescopic development and the fact that $u_i \leqslant (1 + 1/n)u_i \leqslant u_i + 1/n$.

As a consequence, Stute's expansion (4.8) is valid for $\hat{C}_n$ if and only if it is valid for $\hat{C}_n^D$.

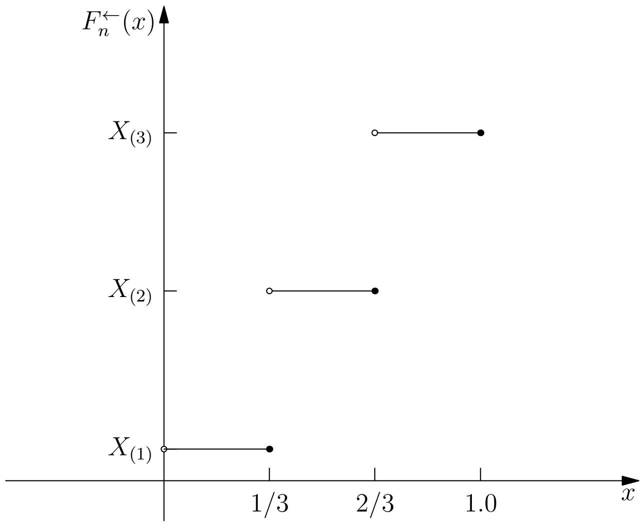


Figure 4.3: However, we have that $X_{(2)} = F_n^{\leftarrow}(1/2)$ such that $X_{(1)}, X_{(2)} \leqslant F_n^{\leftarrow}(1/2)$.

5 Projection estimator

In practice, the nonparametric estimators presented in the previous chapters do not provide valid Pickands dependence functions defined as the functions belonging to the set $\mathscr{A}$ given in (1.12). In this chapter the projection methodology developed in Fils-Villetard et al. [2008], consisting in transforming an initial (nonparametric) estimate $\hat{A}_n$ of A into a valid Pickands function, will be adapted to the higher-dimensional case. The methodology will be introduced in Section 5.1 followed by a simulation study in Section 5.2. The proofs can be found in Section 5.3. This chapter is based on the second part of Gudendorf and Segers [2011b].

5.1 Discrete finite-dimensional approximations

Apart from the bivariate case, it is not sufficient to project the initial estimates on the set of convex functions such that $\max(w_1, \ldots, w_p) \leqslant A(\boldsymbol{w}) \leqslant 1$ for $\boldsymbol{w} = (w_1, \ldots, w_p) \in \Delta_{p-1}$. From the definition of $\mathscr{A}$ given in (1.12) one can see that any Pickands dependence function is fully characterized by its associated spectral measure H. From this observation it follows that the goal will be to determine a spectral measure such that its associated Pickands function minimizes the distance to the initial estimate $\hat{A}_n$.

In the context of this chapter we view $\mathscr{A}$ as a closed and convex subset of the real Hilbert space $L^2(\Delta_{p-1})$ with Δ_{p-1} equipped with $(p-1)$-dimensional Lebesgue measure when viewed as a subset of $\mathbb{R}^{p-1}$. The inner product and the norm on $L^2(\Delta_{p-1})$ are denoted by $\langle f, g\rangle = \int fg$ and $\|f\|_2 = (\langle f, f\rangle)^{1/2}$ respectively.

The orthogonal projection of an initial estimator $\hat{A}_n$ for A, for example the Pickands or the CFG estimator, onto $\mathscr{A}$ is then defined as

$$\hat{A}^{\mathrm{pr}} = \Pi(\hat{A}_n|\mathscr{A}) = \operatorname*{arg\,min}_{A \in \mathscr{A}} \|A - \hat{A}_n\|_2.$$

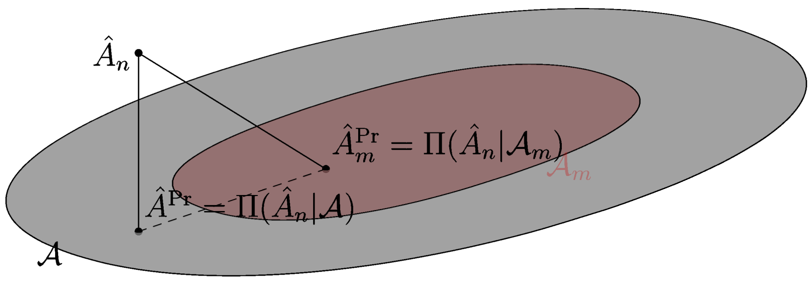


Figure 5.1: Projection framework.

Projections being contractions, it follows that $\|\hat{A}^{\mathrm{pr}} - A\|_2 \leqslant \|\hat{A}_n - A\|_2$ for all $A \in \mathscr{A}$: the L^2-risk of the projection estimator is bounded by the one of the initial estimator.

The projection space $\mathscr{A}$ being infinite-dimensional we prefer to refer to finite-dimensional subclasses $\mathscr{A}_m \subset \mathscr{A}$ for practical computations, yielding the approximate projection estimator

$$\hat{A}^{\mathrm{pr}}_m = \Pi(\hat{A}_n|\mathscr{A}_m) = \operatorname*{arg\,min}_{A \in \mathscr{A}_m} \|A - \hat{A}_n\|_2. \tag{5.1}$$

For each positive integer m, the class $\mathscr{A}_m$ will be defined as the set of Pickands dependence functions characterized by discrete spectral measures H with fixed and finite support depending on m. Specifically, let $\mathscr{V}_{p,m}$ be the (finite) set of points $\boldsymbol{v} = (v_1, \dots, v_p) \in \Delta_{p-1}$ such that $k_j = mv_j$ is integer for every $j \in \{1, \dots, p\}$, so that in fact $\boldsymbol{v} = (k_1/m, \dots, k_p/m)$ where $k_j \in \{0, \dots, m\}$ and $k_1 + \dots + k_p = m$. The cardinality of $\mathscr{V}_{p,m}$ is of the order $O(m^{p-1})$ as $m \to \infty$. Let $\mathscr{H}_{p,m}$ be the set of (discrete) spectral measures $H \in \mathscr{H}_p$ supported on $\mathscr{V}_{p,m}$, that is, $H = \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} h_{\boldsymbol{v}}\, \delta_{\boldsymbol{v}}$, with $\delta_{\boldsymbol{v}}$ the Dirac measure at $\boldsymbol{v}$ and where $h_{\boldsymbol{v}} = H(\{\boldsymbol{v}\})$ is the spectral mass of the atom $\boldsymbol{v}$. The vector $\boldsymbol{h} = (h_{\boldsymbol{v}})_{\boldsymbol{v} \in \mathscr{V}_{p,m}}$ satisfies the constraints

$$\begin{cases} h_{\boldsymbol{v}} \geqslant 0, & \forall \boldsymbol{v} \in \mathscr{V}_{p,m}, \\ \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} h_{\boldsymbol{v}}\, v_j = 1, & \forall j \in \{1, \dots, p\}, \end{cases} \tag{5.2}$$

the second constraint stemming from (1.10).

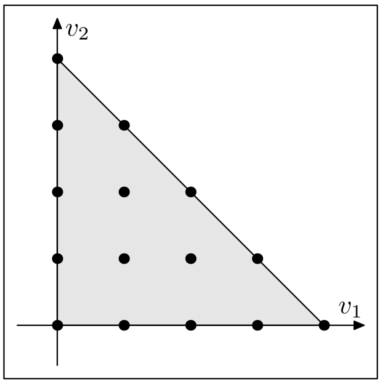


Figure 5.2: $\mathscr{V}_{3,5}$: Set of masspoints on the unit-simplex Δ_2 with $m = 5$.

The Pickands dependence function A of a spectral measure H in $\mathscr{H}_{p,m}$ can be written as

$$A_{\boldsymbol{h}}(\boldsymbol{w}) = \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} h_{\boldsymbol{v}} \max\{w_1 v_1, \dots, w_p v_p\}, \qquad \boldsymbol{w} \in \Delta_{p-1}. \tag{5.3}$$

Being a linear combination of piecewise linear functions, the function A in (5.3) is itself piecewise linear. All Pickands dependence function of the form (5.3) will be collected in the class $\mathscr{A}_m$. The next result can be seen as the equivalent of Lemma 2 in Fils-Villetard et al. [2008] stating the denseness of the piecewise linear Pickands dependence functions.

Lemma 5.1. *For every $H \in \mathscr{H}_p$ and every positive integer m, there exists $H_m \in \mathscr{H}_{p,m}$ such that the Pickands dependence functions A and A_m of H and H_m respectively satisfy*

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}} |A_m(\boldsymbol{w}) - A(\boldsymbol{w})| \leqslant \frac{p^2}{m}. \tag{5.4}$$

The bound in (5.4) implies that $\sup_{A \in \mathscr{A}} \inf_{\tilde{A} \in \mathscr{A}_m} \|\tilde{A} - A\|_2 = O(m^{-1})$ as $m \to \infty$. This rate is perhaps not sharp, for in case $p = 2$, Lemma 2 in Fils-Villetard et al. [2008] states the rate $O(m^{-3/2})$. It remains an open problem whether the latter rate can also be achieved in general dimension p.

In practice, the task is to compute the vector $\hat{\boldsymbol{h}}$ such that the function

$$\hat{A}_m^{\mathrm{pr}}(\boldsymbol{w}) = A_{\hat{\boldsymbol{h}}}(\boldsymbol{w}) = \sum_{\boldsymbol{v}\in\mathscr{V}_{p,m}} \hat{h}_{\boldsymbol{v}} \max\{w_1 v_1, \ldots, w_p v_p\}, \quad \boldsymbol{w} \in \Delta_{p-1},$$

solves (5.1). The vector $\hat{\boldsymbol{h}}$ is given as the solution to the least-squares problem

$$\hat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \|A_{\boldsymbol{h}} - \hat{A}_n\|_2^2 = \arg\min_{\boldsymbol{h}} \bigl(\|A_{\boldsymbol{h}}\|_2^2 - 2\langle A_{\boldsymbol{h}}, \hat{A}_n\rangle\bigr), \tag{5.5}$$

with $\boldsymbol{h}$ subject to the constraints (5.2). The optimisation problem in (5.5) is a quadratic program with linear constraints, which in matrix notation reads

$$\hat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \left(\frac{1}{2}\boldsymbol{h}^\top \boldsymbol{D}\boldsymbol{h} - \boldsymbol{d}^\top \boldsymbol{h}\right), \quad \text{subject to} \quad \begin{cases} \boldsymbol{C}\boldsymbol{h} = \boldsymbol{c}, \\ \boldsymbol{h} \geqslant \boldsymbol{0}. \end{cases} \tag{5.6}$$

The matrix $\boldsymbol{D}$ and the vector $\boldsymbol{d}$ regroup all the scalar products of the form $\int_{\Delta_{p-1}} \max(\boldsymbol{w}\,\boldsymbol{v}) \max(\boldsymbol{w}\,\boldsymbol{v}')\,\mathrm{d}\boldsymbol{w}$ and $\int_{\Delta_{p-1}} \max(\boldsymbol{w}\,\boldsymbol{v})\hat{A}_n(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}$ respectively, for $\boldsymbol{v}, \boldsymbol{v}' \in \mathscr{V}_{p,m}$. The p equality constraints $\sum_{\boldsymbol{v}\in\mathscr{V}_{p,m}} h_{\boldsymbol{v}}\, v_j = 1$ are encoded by means of the matrix $\boldsymbol{C}$ and the vector $\boldsymbol{c}$.

For implementation, we used the R-package quadprog by Weingessel [2010] for solving quadratic programs under linear constraints. Although there exist multiple packages for numerical multivariate integration, we preferred to compute all the integrals appearing in $\boldsymbol{D}$ and $\boldsymbol{d}$ using Riemann sums on the same fine grid. By so doing we reduce the risk of numerical problems as we impose $\boldsymbol{D}$ to be positive definite.

The derivation of the asymptotics of the projection estimator follows the same lines as in Fils-Villetard et al. [2008]. Assume that $\varepsilon_n^{-1}(\hat{A}_n - A) \rightsquigarrow \zeta$ in $L^2(\Delta_{p-1})$ where ζ is a random process in $L^2(\Delta_{p-1})$ and $0 < \varepsilon_n \to 0$; for the Pickands and CFG estimators, we have $\varepsilon_n = n^{-1/2}$ and we have weak convergence with respect to the uniform topology, which implies convergence with respect to the L^2-topology.

By Lemma 1 in Fils-Villetard et al. [2008], we have

$$\|\hat{A}_m^{\mathrm{pr}} - \hat{A}^{\mathrm{pr}}\|_2 \leqslant [\delta_m\{2\|\hat{A}_n - \hat{A}^{\mathrm{pr}}\|_2 + \delta_m\}]^{1/2},$$

with $\delta_m = \|\hat{A}^{\mathrm{pr}} - \Pi(\hat{A}^{\mathrm{pr}}|\mathscr{A}_m)\|_2$. From Lemma 5.1 above, we have $\delta_m = O(1/m)$ as $m \to \infty$. As a consequence, if $m = m_n$ is such that $1/m_n = o(\varepsilon_n)$ as $n \to \infty$, then $\|\hat{A}_m^{\mathrm{pr}} - \hat{A}^{\mathrm{pr}}\|_2 = o_{\mathrm{P}}(\epsilon_n)$.

Consider a sequence of real-valued functions $(g_n)_{n\in\mathbb{N}}$ defined on Δ_{p-1} converging to g in $L^2(\Delta_{p-1})$. Albeit under a more general setting, it was proved in Lemma 3 of Fils-Villetard et al. [2008] that

$$\lim_{n\to\infty} \frac{\Pi\big(A+\epsilon_n\, g_n|\mathscr{A}\big)-A}{\epsilon_n} = \Pi\big(g|\mathrm{T}_{\mathscr{A}}(A)\big),$$

where $\mathrm{T}_{\mathscr{A}}(A)$ is the tangent cone of $\mathscr{A}$ at A, defined as the L^2-closure of $\{\lambda(\tilde{A}-A) : \lambda \geqslant 0, \tilde{A} \in \mathscr{A}\}$. As $A = \Pi\big(A|\mathscr{A}\big)$, the idea behind the proof of this lemma consists in proving some kind of Hadamard-differentiability of the operator $\Pi\big(\cdot|\mathscr{A}\big)$.

From this we can deduce that

$$\varepsilon_n^{-1}(\hat{A}^{\mathrm{pr}}_m - A) = \varepsilon_n^{-1}(\hat{A}^{\mathrm{pr}} - A) + o_{\mathrm{P}}(1) \rightsquigarrow \Pi\big(\zeta|\mathrm{T}_{\mathscr{A}}(A)\big) \qquad (n\to\infty) \quad (5.7)$$

in the space $L^2(\Delta_{p-1})$.

From the properties of projection estimators and the fact that $0 \in \mathrm{T}_{\mathscr{A}}(A)$, one has the following inequality

$$\|\zeta\|_2^2 \geqslant \|\zeta - \Pi\big(\zeta|\mathrm{T}_{\mathscr{A}}(A)\big)\|_2^2 + \|\Pi\big(\zeta|\mathrm{T}_{\mathscr{A}}(A)\big)\|_2^2.$$

It follows that

$$\int \zeta^2(\boldsymbol{w})\,\mathrm{d}w_1 \ldots \mathrm{d}w_{p-1} \geqslant \int \Pi\big(\zeta(\boldsymbol{w})|\mathrm{T}_{\mathscr{A}}(A)\big)^2\,\mathrm{d}w_1 \ldots \mathrm{d}w_{p-1},$$

meaning a decrease of the asymptotic variance of the estimator due to the projection.

Interestingly, equation (5.7) implies that the choice of m is not to be seen as a bias-variance trade-off problem but rather as a discretization problem. As soon as $m = m_n$ converges to infinity faster than $1/\varepsilon_n$, the finite-dimensional projection estimator $\hat{A}^{\mathrm{pr}}_m$ has the same limit behaviour as the 'ideal' projection estimator $\hat{A}^{\mathrm{pr}}$. In practice, we will choose m sufficiently large so that any further increase of m does not make any significant difference, of course subject to constraints on computing time and numerical stability. The grid of equidistant masspoints $\mathscr{V}_{p,m}$ has been chosen for purely pragmatic reasons and it might be interesting to investigate the influence of the distribution of the mass points on the quality of the projection estimator. For example, one could imagine that a curvature-sensitive grid of masspoints could help to improve the overall fit.

Finally, note that the convergence in (5.7) is with respect to the L^2-topology only, even if originally the weak convergence of $\varepsilon_n^{-1}(\hat{A}_n - A)$ took place in the stronger ℓ^∞-topology. The asymptotic distribution of the projection estimator under the ℓ^∞-topology remains an open problem.

5.2 Simulation study

5.2.1 Piecewise linear projections

A simulation experiment was conducted to compare the finite-sample performance of the following four estimators:

PD – the endpoint-corrected Pickands estimator in (2.4) with $\lambda_j(\boldsymbol{w}) = w_j$, in the spirit of Deheuvels [1991];

PD-pr – the projection estimator in (5.1) with the previous end-point corrected Pickands estimator as initial estimator;

CFG – the endpoint-corrected CFG estimator in (2.9) with $\lambda_j(\boldsymbol{w}) = w_j$;

CFG-pr – the projection estimator in (5.1) with the previous end-point corrected CFG estimator as initial estimator.

The set-up of the experiment was as follows. Following Zhang et al. [2008] and Gudendorf and Segers [2011a], random samples were generated from a trivariate extreme-value distribution with asymmetric logistic dependence function A [Tawn, 1990]:

$$\begin{aligned} A(\boldsymbol{w}) = {} & (\theta^{1/\alpha} w_1^{1/\alpha} + \phi^{1/\alpha} w_2^{1/\alpha})^\alpha + (\theta^{1/\alpha} w_2^{1/\alpha} + \phi^{1/\alpha} w_3^{1/\alpha})^\alpha \\ & + (\theta^{1/\alpha} w_3^{1/\alpha} + \phi^{1/\alpha} w_1^{1/\alpha})^\alpha + \psi (w_1^{1/\alpha} + w_2^{1/\alpha} + w_3^{1/\alpha})^\alpha \\ & + 1 - \theta - \phi - \psi, \qquad \boldsymbol{w} \in \Delta_2, \quad (5.8) \end{aligned}$$

with parameter vector $(\alpha, \theta, \phi, \psi) \in (0, 1] \times [0, 1]^3$. For this model, Assumption 4.3 can be verified by direct calculation. The dependence parameter α ranged from 0.3 (high dependence) to 1 (independence, $A \equiv 1$) and the vector (ϕ, ψ, θ) was set equal to either $(0, 1, 0)$ (symmetric logistic copula or Gumbel copula) and $(0.3, 0, 0.6)$ (an asymmetric logistic copula). For each distribution, 1000 samples were generated of size $n \in \{50, 100, 200\}$. Simulations were performed using the R-package evd [Stephenson, 2002], which implements the algorithms presented in Stephenson [2003]. The discretization parameter m was set to 20, at which value the grid $\mathscr{V}_{3,20}$ contains 231 points.

Monte-Carlo approximations for the mean integrated squared error (MISE) $\mathrm{E}[\int(\hat{A}-A)^2]$ for the four estimators considered above are reported in the tables below. The three main findings are the following:

1- The projection step yields a gain in efficiency, especially in case of weak dependence.
2- Without projection step, the CFG estimator outperforms the PD estimator.
3- After the projection step, the PD-pr estimator is more efficient than the CFG-pr estimator in case of independence and weak dependence ($\alpha \geqslant 0.9$), but less efficient otherwise ($\alpha \leqslant 0.7$).

Further, we find that as the dependence increases, all estimators tend to perform better. In accordance with asymptotic theory, the MISE is roughly proportional to $1/n$.

5.3 Proofs

5.3.1 Proof of Lemma 5.1

The proof is constructive and consists of the following steps:

1. Construction of the spectral measure H_m.
   (a) Discretisation of H yielding a measure G_m on Δ_{p-1}, which is not necessarily a spectral measure.
   (b) Modification of G_m into a genuine spectral measure H_m .
2. Proof of the inequality (5.4).

1. Construction of the spectral measure H_m. For $\boldsymbol{v} \in \mathscr{V}_{p,m}$, consider the set $\Delta_{p-1,\boldsymbol{v},m}$ of points $\boldsymbol{t} \in \Delta_{p-1}$ such that $v_j \leqslant t_j < v_j + 1/m$ for every $j \in \{1, \dots, p-1\}$; recall that $t_p = 1 - t_1 - \cdots - t_{p-1}$, so that necessarily $v_p - (p-1)/m < t_p \leqslant v_p$. The collection of sets $\{\Delta_{p-1,\boldsymbol{v},m} : \boldsymbol{v} \in \mathscr{V}_{p,m}\}$ constitutes a partition of Δ_{p-1}. Indeed, for every point $\boldsymbol{t} \in \Delta_{p-1}$ there is a unique point $\boldsymbol{v} \in \mathscr{V}_{p,m}$ such that $\boldsymbol{t} \in \Delta_{p-1,\boldsymbol{v},m}$: Let v_j be the integer part of mt_j for $j \in \{1, \dots, p-1\}$ and put $v_p = 1 - v_1 - \cdots - v_{p-1}$.

(a) Discretisation of H, yielding G_m. Define a discrete measure G_m on Δ_{p-1} with support contained in $\mathscr{V}_{p,m}$ by $G_m(\{\boldsymbol{v}\}) = H(\Delta_{p-1,\boldsymbol{v},m})$ for $\boldsymbol{v} \in \mathscr{V}_{p,m}$. In words, the mass assigned by the spectral measure H on the set $\Delta_{p-1,\boldsymbol{v},m}$ is relocated to the corner point $\boldsymbol{v}$.

Since the sets $\Delta_{p-1,\boldsymbol{v},m}$ constitute a partition of Δ_{p-1}, the total mass of G_m is still $G_m(\Delta_{p-1}) = H(\Delta_{p-1}) = p$. However, G_m does not need to verify the moment constraints. For $j \in \{1, \dots, p\}$ we have

$$\int_{\Delta_{p-1}} t_j \, dG_m(\boldsymbol{t}) = \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} v_j \, G_m(\{\boldsymbol{v}\}) = \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} v_j \, H(\Delta_{p-1,\boldsymbol{v},m}),$$

which in general is not equal to unity.

Still, the moment constraints are not far from being verified. For $\boldsymbol{t} \in \Delta_{p-1,\boldsymbol{v},m}$ and $j \in \{1, \dots, p-1\}$ we have $v_j \leqslant t_j < v_j + 1/m$. Integrating these inequalities over $\boldsymbol{t} \in \Delta_{p-1,\boldsymbol{v},m}$ with respect to H and summing them over $\boldsymbol{v} \in \mathscr{V}_{p,m}$ yields

$$\int_{\Delta_{p-1}} t_j \, dG_m(\boldsymbol{t}) \leqslant 1 < \int_{\Delta_{p-1}} t_j \, dG_m(\boldsymbol{t}) + \frac{1}{m}, \qquad j \in \{1, \dots, p-1\}.$$

As a consequence, there exist numbers $c_j \in [0, 1)$ such that

$$\int_{\Delta_{p-1}} t_j \, dG_m(\boldsymbol{t}) = 1 - \frac{c_j}{m}, \qquad j \in \{1, \dots, p-1\}.$$

(b) Modification of G_m into a spectral measure H_m. We will modify G_m into a genuine spectral measure H_m by (slightly) increasing the masses at the vertices $e_1, \dots, e_{p-1}$, where e_j is the jth coordinate vector in $\mathbb{R}^p$. Specifically, we set

$$H_m = (1 - a_0) \, G_m + a_1 \, \delta_{e_1} + \dots + a_{p-1} \, \delta_{e_{p-1}}$$

for some nonnegative numbers $a_0, \dots, a_{p-1}$ to be determined by the moment constraints. For $j \in \{1, \dots, p-1\}$, we must have

$$1 = \int_{\Delta_{p-1}} t_j \, dH_m(\boldsymbol{t}) = (1 - a_0)(1 - c_j/m) + a_j.$$

In addition, the total mass must be equal to

$$p = H_m(\Delta_{p-1}) = (1 - a_0) \, p + a_1 + \dots + a_{p-1}.$$

Substituting $a_j = 1 - (1 - a_0)(1 - c_j/m)$ into this equation and solving for a_0 yields, after some algebra,

$$a_0 = \frac{\sum_{i=1}^{p-1} c_i}{m + \sum_{i=1}^{p-1} c_i},$$

$$a_j = \frac{c_j + \sum_{i=1}^{p-1} c_i}{m + \sum_{i=1}^{p-1} c_i}, \qquad j \in \{1, \dots, p-1\}.$$

This concludes the construction of the spectral measure H_m. Note that $0 \leqslant a_j < p/m$ for every $j \in \{0, \ldots, p-1\}$.

2. Proof of the inequality (5.4). For $\boldsymbol{w}, \boldsymbol{t} \in \Delta_{p-1}$, write

$$f(\boldsymbol{w}, \boldsymbol{t}) = \max\{w_1 t_1, \ldots, w_p t_p\}.$$

The Pickands dependence function A_m of the spectral measure H_m constructed above is given by

$$\begin{aligned} A_m(\boldsymbol{w}) &= \int_{\Delta_{p-1}} f(\boldsymbol{w}, \boldsymbol{t})\, H_m(\mathrm{d}\boldsymbol{t}) \\ &= (1 - a_0) \int_{\Delta_{p-1}} f(\boldsymbol{w}, \boldsymbol{t})\, G_m(\mathrm{d}\boldsymbol{t}) + a_1 w_1 + \cdots + a_{p-1} w_{p-1}. \end{aligned}$$

Put $B_m(\boldsymbol{w}) = \int f(\boldsymbol{w}, \boldsymbol{t})\, G_m(\mathrm{d}\boldsymbol{t})$, the "Pickands transform" of G_m. Clearly $B_m \geqslant 0$ and B_m is convex, being a weighted average (over $\boldsymbol{t}$) of the convex functions $\boldsymbol{w} \mapsto f(\boldsymbol{w}, \boldsymbol{t})$. As a consequence, $B_m(\boldsymbol{w}) \leqslant \max\{B_m(\boldsymbol{e_1}), \ldots, B_m(\boldsymbol{e_p})\}$. Now $B_m(\boldsymbol{e_j}) = \int t_j\, G_m(\mathrm{d}\boldsymbol{t})$, which is equal to $1 - c_j/m$ if $j \in \{1, \ldots, p-1\}$ and which is equal to $p - \sum_{i=1}^{p-1}(1 - c_i/m) = 1 + \sum_{i=1}^{p-1} c_i/m = 1/(1 - a_0)$ if $j = p$. It follows that $B_m(\boldsymbol{w}) \leqslant 1/(1 - a_0)$ for all $\boldsymbol{w} \in \Delta_{p-1}$.

We obtain, on the one hand,

$$\begin{aligned} A_m(\boldsymbol{w}) &\leqslant B_m(\boldsymbol{w}) + a_1 w_1 + \cdots + a_{p-1} w_{p-1} \\ &\leqslant B_m(\boldsymbol{w}) + \max(a_1, \ldots, a_{p-1}) < B_m(\boldsymbol{w}) + \frac{p}{m} \end{aligned}$$

and, on the other hand,

$$\begin{aligned} A_m(\boldsymbol{w}) \geqslant (1 - a_0) B_m(\boldsymbol{w}) \geqslant B_m(\boldsymbol{w}) - \frac{a_0}{1 - a_0} &= B_m(\boldsymbol{w}) - \frac{1}{m} \sum_{i=1}^{p-1} c_i \\ &> B_m(\boldsymbol{w}) - \frac{p}{m}. \end{aligned}$$

Therefore,

$$|A(\boldsymbol{w}) - A_m^{\mathrm{P}}(\boldsymbol{w})| \leqslant |A(\boldsymbol{w}) - B_m(\boldsymbol{w})| + |B_m(\boldsymbol{w}) - A_m(\boldsymbol{w})| < |A(\boldsymbol{w}) - B_m(\boldsymbol{w})| + \frac{p}{m}.$$

Furthermore,

$$\begin{aligned}
|A(\boldsymbol{w}) - B_m(\boldsymbol{w})| &\leqslant \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \left| \int_{\Delta_{p,\boldsymbol{v},m}} f(\boldsymbol{w}, \boldsymbol{t})\, H(\mathrm{d}\boldsymbol{t}) - \int_{\Delta_{p,\boldsymbol{v},m}} f(\boldsymbol{w}, \boldsymbol{t})\, G_m(\mathrm{d}\boldsymbol{t}) \right| \\
&= \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \left| \int_{\Delta_{p,\boldsymbol{v},m}} f(\boldsymbol{w}, \boldsymbol{t})\, H(\mathrm{d}\boldsymbol{t}) - f(\boldsymbol{w}, \boldsymbol{v})\, H(\Delta_{p,\boldsymbol{v},m}) \right| \\
&= \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \left| \int_{\Delta_{p,\boldsymbol{v},m}} \big(f(\boldsymbol{w}, \boldsymbol{t}) - f(\boldsymbol{w}, \boldsymbol{v}) \big)\, H(\mathrm{d}\boldsymbol{t}) \right| \\
&\leqslant \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \int_{\Delta_{p,\boldsymbol{v},m}} \big| f(\boldsymbol{w}, \boldsymbol{t}) - f(\boldsymbol{w}, \boldsymbol{v}) \big|\, H(\mathrm{d}\boldsymbol{t}).
\end{aligned}$$

By checking all possible cases one verifies that $|\max(a_1, a_2) - \max(b_1, b_2)| \leqslant \max(|a_1 - b_1|, |a_2 - b_2|)$ for all real a_1, a_2, b_1, b_2. An induction argument then yields $|\max(a_1, \ldots, a_k) - \max(b_1, \ldots, b_k)| \leqslant \max(|a_1 - b_1|, \ldots, |a_k - b_k|)$. It follows that $|f(\boldsymbol{w}, \boldsymbol{t}) - f(\boldsymbol{w}, \boldsymbol{v})| \leqslant \max(|t_1 - v_1|, \ldots, |t_p - v_p|)$. As a consequence,

$$|A(\boldsymbol{w}) - B_m(\boldsymbol{w})| < \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \int_{\Delta_{p,\boldsymbol{v},m}} \frac{p-1}{m}\, H(\mathrm{d}\boldsymbol{t}) = \frac{(p-1)p}{m}.$$

Inequality (5.4) follows.

5.4 Appendix

		$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$	$\alpha = 1$
$n = 50$	PD	$1.40 \cdot 10^{-4}$	$5.44 \cdot 10^{-4}$	$1.36 \cdot 10^{-3}$	$2.68 \cdot 10^{-3}$	$3.44 \cdot 10^{-3}$
	PD-pr	$1.37 \cdot 10^{-4}$	$5.14 \cdot 10^{-4}$	$1.21 \cdot 10^{-3}$	$2.08 \cdot 10^{-3}$	$2.44 \cdot 10^{-3}$
	CFG	$9.77 \cdot 10^{-5}$	$4.27 \cdot 10^{-4}$	$1.26 \cdot 10^{-3}$	$2.54 \cdot 10^{-3}$	$3.48 \cdot 10^{-3}$
	CFG-pr	$9.69 \cdot 10^{-5}$	$4.22 \cdot 10^{-4}$	$1.22 \cdot 10^{-3}$	$2.43 \cdot 10^{-3}$	$3.31 \cdot 10^{-3}$
$n = 100$	PD	$7.08 \cdot 10^{-5}$	$2.84 \cdot 10^{-4}$	$7.06 \cdot 10^{-4}$	$1.34 \cdot 10^{-3}$	$1.69 \cdot 10^{-3}$
	PD-pr	$6.99 \cdot 10^{-5}$	$2.74 \cdot 10^{-4}$	$6.53 \cdot 10^{-4}$	$1.03 \cdot 10^{-3}$	$1.08 \cdot 10^{-3}$
	CFG	$5.03 \cdot 10^{-5}$	$2.39 \cdot 10^{-4}$	$6.56 \cdot 10^{-4}$	$1.23 \cdot 10^{-3}$	$1.48 \cdot 10^{-3}$
	CFG-pr	$5.01 \cdot 10^{-5}$	$2.37 \cdot 10^{-4}$	$6.47 \cdot 10^{-4}$	$1.18 \cdot 10^{-3}$	$1.36 \cdot 10^{-3}$
$n = 200$	PD	$3.31 \cdot 10^{-5}$	$1.43 \cdot 10^{-4}$	$3.92 \cdot 10^{-4}$	$7.02 \cdot 10^{-4}$	$8.71 \cdot 10^{-4}$
	PD-pr	$3.29 \cdot 10^{-5}$	$1.40 \cdot 10^{-4}$	$3.73 \cdot 10^{-4}$	$5.72 \cdot 10^{-4}$	$5.14 \cdot 10^{-4}$
	CFG	$2.45 \cdot 10^{-5}$	$1.23 \cdot 10^{-4}$	$3.39 \cdot 10^{-4}$	$6.40 \cdot 10^{-4}$	$6.56 \cdot 10^{-4}$
	CFG-pr	$2.45 \cdot 10^{-5}$	$1.23 \cdot 10^{-4}$	$3.36 \cdot 10^{-4}$	$6.19 \cdot 10^{-4}$	$5.78 \cdot 10^{-4}$

Table 5.1: Symmetric logistic dependence function, $(\phi, \psi, \theta) = (0, 1, 0)$: Monte-Carlo approximation of the MISE of four estimators of A based on 1000 random samples

		$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
$n = 50$	PD	$1.42 \cdot 10^{-3}$	$1.72 \cdot 10^{-3}$	$2.20 \cdot 10^{-3}$	$2.88 \cdot 10^{-3}$
	PD-pr	$1.22 \cdot 10^{-3}$	$1.45 \cdot 10^{-3}$	$1.74 \cdot 10{-}3$	$2.01 \cdot 10^{-3}$
	CFG	$1.15 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$	$1.84 \cdot 10^{-3}$	$2.77 \cdot 10^{-3}$
	CFG-pr	$1.10 \cdot 10^{-3}$	$1.35 \cdot 10^{-3}$	$1.75 \cdot 10^{-3}$	$2.60 \cdot 10^{-3}$
$n = 100$	PD	$7.67 \cdot 10^{-4}$	$8.76 \cdot 10^{-4}$	$1.10 \cdot 10^{-3}$	$1.51 \cdot 10^{-3}$
	PD-pr	$6.77 \cdot 10^{-4}$	$7.70 \cdot 10^{-4}$	$9.00 \cdot 10^{-4}$	$1.08 \cdot 10^{-3}$
	CFG	$5.90 \cdot 10^{-4}$	$7.06 \cdot 10^{-4}$	$9.05 \cdot 10^{-4}$	$1.20 \cdot 10^{-3}$
	CFG-pr	$5.70 \cdot 10^{-4}$	$6.85 \cdot 10^{-4}$	$8.68 \cdot 10^{-4}$	$1.10 \cdot 10^{-3}$
$n = 200$	PD	$3.92 \cdot 10^{-4}$	$4.72 \cdot 10^{-4}$	$5.84 \cdot 10^{-4}$	$7.60 \cdot 10^{-4}$
	PD-pr	$3.52 \cdot 10^{-4}$	$4.30 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$	$5.19 \cdot 10^{-4}$
	CFG	$3.01 \cdot 10^{-4}$	$3.31 \cdot 10^{-4}$	$4.43 \cdot 10^{-4}$	$5.81 \cdot 10^{-4}$
	CFG-pr	$2.92 \cdot 10^{-4}$	$3.22 \cdot 10^{-4}$	$4.29 \cdot 10^{-4}$	$5.36 \cdot 10^{-4}$

Table 5.2: Asymmetric logistic dependence function, $(\phi, \psi, \theta) = (0.3, 1, 0.6)$: Monte-Carlo approximation of the MISE of four estimators of A based on 1000 random samples

6 An extreme-value dependence test for multivariate extreme-value copulas

There exist two different but equivalent representations for the class of extreme-value copulas, one based on max-stability and the other one based on the Pickands representation. The max-stable characterization of extreme-value copulas is exploited in Kojadinovic et al. [2011] to test extreme-value dependence for multivariate copulas. However, the nature of this characterization makes it hard to develop a consistent test in practice. For this reason it might be preferable to build a test for extreme-value dependence on the Pickands representation as done in Section 6.2. The newly presented test is implemented in combination with the multiplier method to obtain p-values; see Section 6.3. The finite-sample properties of the test are investigated through extensive Monte-Carlo simulations in Section 6.4. All the proofs have been relegated to Section 6.5. The results in this chapter have been regrouped in a paper recently submitted for publication in a special edition of the Journal of the French Statistical Society dedicated on copulas.

6.1 Introduction

To fix notations, let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,p})$, $i \in \{1, \ldots, n\}$ be an independent random sample from a p-variate, continuous distribution function F with margins $F_1, \ldots, F_p$ and copula C, that is,

$$F(\boldsymbol{x}) = C\bigl(F_1(x_1), \ldots, F_p(x_p)\bigr), \qquad \boldsymbol{x} \in \mathbb{R}^p,$$

where $F(\boldsymbol{x}) = \mathrm{P}(\boldsymbol{X} \leqslant \boldsymbol{x})$ (componentwise inequalities), $F_j(x_j) = \mathrm{P}(X_j \leqslant x_j)$ and C is the joint distribution function of $(F_1(X_1), \ldots, F_p(X_p))$.

Kojadinovic et al. [2011] used the max-stable characterization for extreme-value copulas given in Definition 1.4 in connection with the multiplier resampling method presented in Rémillard and Scaillet [2009] to implement a test for extreme-value dependence. At least from a theoretical point of view, a drawback of adopting this perspective lies in the fact that the max-stability relationship should be verified for every positive real m, which cannot be realized in practice. Hence, the test cannot be consistent.

Following the same objective to implement a test of extreme-value dependence it might be preferable, at least theoretically, to adopt the perspective of Pickands representation for extreme-value copulas given in (1.13).

Nonparametric estimation of higher-dimensional extreme-value copulas in the realistic case of unknown margins has been treated recently in Gudendorf and Segers [2011b] and can be found back in Chapter 4 of this work. Based on these results in connection with the multiplier resampling method, a test for extreme-value dependence, measuring the distance between an extreme-value copula estimate and the empirical copula estimate, will be introduced. The same test exists already for dimension $p = 2$ in Kojadinovic and Yan [2010b], of which the present paper can thus be considered a generalization to arbitrary dimensions.

Based on Kendall's distribution function for bivariate copulas, the first test for extreme-value dependence was proposed by Ghoudi et al. [1998]. A refined version was suggested recently in Ghorbal et al. [2009] and Quessy [2011]. This approach has not yet been generalized to the higher-dimensional case.

6.2 Extreme-value copula goodness-of-fit test

The null hypothesis H_0 consists in the assumption that C belongs to the class of extreme-value copulas. As we are not willing to make any assumptions on the marginal distributions, we will work again with the rank-based estimates $\hat{\boldsymbol{U}}_i = (\hat{U}_{i,1}, \ldots, \hat{U}_{i,p})$ with entries

$$\hat{U}_{i,j} = \frac{1}{n+1} \sum_{l=1}^{n} \mathbf{1}(X_{l,j} \leqslant X_{i,j})$$

for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$. As already indicated in the introduction, we are going to measure the distance between the empirical copula $\hat{C}_n$

$$\hat{C}_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\hat{U}_{i,1} \leqslant u_1, \ldots, \hat{U}_{i,p} \leqslant u_p\right), \qquad \boldsymbol{u} \in [0,1]^p \tag{6.1}$$

and the constrained extreme-value copula estimate

$$\hat{C}_n^A(\boldsymbol{u}) = \exp\left(\sum_{j=1}^{p} \log(u_j)\, \hat{A}_n(\boldsymbol{w_u})\right), \quad \boldsymbol{u} \in \mathbb{E}_p = (0,1]^p \setminus \{(1,\ldots,1)\}, \tag{6.2}$$

with $\hat{A}_n$ being a nonparametric estimator for A evaluated in $\boldsymbol{w_u} = (w_{\boldsymbol{u},1}, \ldots, w_{\boldsymbol{u},p-1})$, where $w_{\boldsymbol{u},i} = \log(u_i)/\sum_{j=1}^{p}\log(u_j)$. In this work, $\hat{A}_n$ will be chosen either as the Pickands estimator presented in Subsection 2.1.1 or as the CFG estimator presented in Subsection 2.1.2.

As we will rely hereafter on the asymptotic results in Chapter 4, we will consider only extreme-value copulas verifying the smoothness conditions given in Assumption 4.3.

Analogously to what has been done in Kojadinovic and Yan [2010b], define

$$\mathbb{D}_n(\boldsymbol{u}) = n^{1/2}(\hat{C}_n(\boldsymbol{u}) - \hat{C}_n^A(\boldsymbol{u})). \tag{6.3}$$

This leads to the following Cramér-von Mises test statistic

$$\mathbb{T}_n = \int_{[0,1]^p} \mathbb{D}_n(\boldsymbol{u})^2 \, \mathrm{d}\hat{C}_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{D}_n(\hat{\boldsymbol{U}}_i)^2. \tag{6.4}$$

The asymptotic distribution of $\mathbb{T}_n$ being difficult to establish, we will rely on the multiplier method to construct random samples

$$\mathbb{T}_n^{(k)} = \int_{[0,1]^p} \mathbb{D}_n^{(k)}(\boldsymbol{u})^2 \, \mathrm{d}\hat{C}_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{D}_n^{(k)}(\hat{\boldsymbol{U}}_i)^2,$$

with $\mathbb{T}_n^{(k)}$ and $\mathbb{D}_n^{(k)}$ being asymptotically independent distributional copies of $\mathbb{T}_n$ and $\mathbb{D}_n$ respectively. In general, terms with the superscript (k) correspond to the k-th bootstrap replication of the underlying process.

More precisely, for the processes $\mathbb{A}_n = n^{1/2}\big(\hat{A}_n - A\big)$ and $\mathbb{C}_n = n^{1/2}\big(\hat{C}_n - C\big)$, the replications $\mathbb{D}_n^{(k)}$ can be defined as follows

$$\mathbb{D}_n^{(k)}(\boldsymbol{u}) = \mathbb{C}_n^{(k)}(\boldsymbol{u}) - \exp\left(\sum_{j=1}^{p} \log(u_j)\, \hat{A}_n\left(\boldsymbol{w}_{\boldsymbol{u}}\right)\right) \sum_{j=1}^{p} \log(u_j)\, \mathbb{A}_n^{(k)}\left(\boldsymbol{w}_{\boldsymbol{u}}\right). \tag{6.5}$$

Under the null hypothesis, it can be shown that $\left(\mathbb{D}_n, \mathbb{D}_n^{(1)}, \dots, \mathbb{D}_n^{(N)}\right)$ converges weakly to $\big(\mathbb{D}, \mathbb{D}^{(1)}, \dots, \mathbb{D}^{(N)}\big)$ in $\ell^\infty([0,1]^p)^{\otimes(N+1)}$ implying the weak convergence of $\left(\mathbb{T}_n, \mathbb{T}_n^{(1)}, \dots, \mathbb{T}_n^{(N)}\right)$ to $\big(\mathbb{T}, \mathbb{T}^{(1)}, \dots, \mathbb{T}^{(N)}\big)$ in $[0,\infty)^{\otimes(N+1)}$, where

$$\mathbb{T} = \int_{[0,1]^p} \{\mathbb{D}(\boldsymbol{u})\}^2 \mathrm{d}C(\boldsymbol{u}). \tag{6.6}$$

Approximate p-values for $\mathbb{T}_n$ are computed via $N^{-1}\sum_{k=1}^{N} \mathbf{1}(\mathbb{T}_n^{(k)} \geqslant \mathbb{T}_n)$. Details about the resampling method, particularly how to obtain the resampled versions $\mathbb{A}_n^{(k)}$ and $\mathbb{C}_n^{(k)}$, will be given Section 6.3.

The proof of the asymptotic results follows the same lines as the one related to Proposition 3 in Kojadinovic et al. [2011].

The test statistic in (6.4) could be generalized by the introduction of a weight function with the intention to attribute more weight to the tails of the distribution and hence hopefully increasing the power of the test, an idea on which we will not insist further in this work.

6.3 Resampling

In this section, details on the resampling procedure will be given. This means that we will see how to obtain the resampled versions of $\mathbb{C}_n^{(k)}$ and $\mathbb{A}_n^{(k)}$ to be plugged into $\mathbb{D}_n^{(k)}$ and hence $\mathbb{T}_n^{(k)}$.

The construction of $\mathbb{C}_n^{(k)}$ will require estimates of the partial derivatives $\dot{C}_j$ of C. Although such estimates could be obtained for instance via finite differencing, the partial derivatives will be estimated using Pickands representation of C as this approach has already lead to more powerful tests in Kojadinovic and Yan [2010b].

For $j \neq p$,

$$\dot{C}_j(\boldsymbol{u}) = \frac{C(\boldsymbol{u})}{u_j}\Big(A(\boldsymbol{w_u}) - \dot{A}_1(\boldsymbol{w_u})\, w_{\boldsymbol{u},1} \\ - \cdots + \dot{A}_j(\boldsymbol{w_u})\,(1 - w_{\boldsymbol{u},j}) - \cdots - \dot{A}_{p-1}(\boldsymbol{w_u})\, w_{\boldsymbol{u},p-1}\Big),$$

and the partial derivative w.r.t. u_p is

$$\dot{C}_p(\boldsymbol{u}) = \frac{C(\boldsymbol{u})}{u_p}\Big(A(\boldsymbol{w_u}) - \dot{A}_1(\boldsymbol{w_u}) w_{\boldsymbol{u},1} - \cdots - \dot{A}_{p-1}(\boldsymbol{w_u})\, w_{\boldsymbol{u},p-1}\Big).$$

Then for $u \in (0,1)$ and $\boldsymbol{w} \in \Delta_{p-1}$, the partial derivatives of C evaluated in $u^{\boldsymbol{w}} = (u^{w_1}, \ldots, u^{w_p})$ simplify to

$$\dot{C}_j(u^{\boldsymbol{w}}) = u^{A(\boldsymbol{w})-w_j}(A(\boldsymbol{w}) - w_1\dot{A}_1(\boldsymbol{w}) \\ - \cdots + (1-w_j)\dot{A}_j(\boldsymbol{w}) - \cdots - w_{p-1}\dot{A}_{p-1}(\boldsymbol{w}))$$

and

$$\dot{C}_p(u^{\boldsymbol{w}}) = u^{A(\boldsymbol{w})-w_p}\Big(A(\boldsymbol{w}) - w_1\dot{A}_1(\boldsymbol{w}) - \cdots - w_{p-1}\dot{A}_{p-1}(\boldsymbol{w})\Big).$$

In practice, we will replace A by its estimate $\hat{A}_n$ and its partial derivatives by the finite-difference estimators

$$\hat{A}^{[j]}(\boldsymbol{w}) = \frac{\hat{A}_n(w_1, \ldots, w_{j,n}^+, \ldots, w_{p-1}) - \hat{A}_n(w_1, \ldots, w_{j,n}^-, \ldots, w_{p-1})}{w_{j,n}^+ - w_{j,n}^-},$$

where

$$w_{j,n}^+ = w_j + \epsilon_n^{\text{inc},+} \quad \text{and} \quad w_{j,n}^- = w_j - \epsilon_n^{\text{inc},-},$$

using the increments

$$\epsilon_n^{\text{inc},+} = \min\left(n^{-1/2}, 1 - \sum_{k=1}^{p-1} w_j\right) \quad \text{and} \quad \epsilon_n^{\text{inc},-} = \min\left(n^{-1/2}, w_j\right).$$

Just keep in mind that if you add a small increment to w_i in order to get $w_{i,n}^+$, you need to subtract the same amount from the last component w_p to ensure that $\boldsymbol{w}$ still belongs to the unit-simplex Δ_{p-1}.

Let N be a large integer and let $Z_i^{(k)}$, $i \in \{1, \dots, n\}$, $k \in \{1, \dots, N\}$ be i.i.d. random variables independent of the data $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ with mean 0 and variance 1 and satisfying $\int_0^\infty P(|Z_i^{(k)}| > x)^{1/2} \, \mathrm{d}x < \infty$ (see Section 2.9 of van der Vaart and Wellner [1996]). In the simulations, the $Z_i^{(k)}$ are chosen to be standard Gaussian. For $k \in \{1, \dots, N\}$ and $\boldsymbol{u} \in [0,1]^p$, let

$$\begin{aligned} \hat{\alpha}_n^{(k)}(\boldsymbol{u}) &= n^{-1/2} \sum_{i=1}^n Z_i^{(k)} (\mathbf{1}(\hat{U}_i \leqslant \boldsymbol{u}) - \hat{C}_n(\boldsymbol{u})) \\ &= n^{-1/2} \sum_{i=1}^n (Z_i^{(k)} - \bar{Z}^{(k)}) \mathbf{1}(\hat{\boldsymbol{U}}_i \leqslant \boldsymbol{u}), \end{aligned}$$

with $\bar{Z}^{(k)} = n^{-1} \sum_{i=1}^n Z_i^{(k)}$.

In particular, set

$$\hat{\alpha}_{j,n}^{(k)}(u_j) = n^{-1/2} \sum_{i=1}^n (Z_i^{(k)} - \bar{Z}^{(k)}) \mathbf{1}(\hat{U}_{i,j} \leqslant u_j), \quad j \in \{1, \dots, p\}.$$

Finally, the replications

$$\mathbb{C}_n^{(k)}(\boldsymbol{u}) = \hat{\alpha}_n^{(k)}(\boldsymbol{u}) - \sum_{j=1}^p \hat{C}_n^{[j]}(\boldsymbol{u}) \hat{\alpha}_{j,n}^{(k)}(u_j). \tag{6.7}$$

will represent asymptotically equivalent copies of the process $\mathbb{C}_n = n^{1/2}(\hat{C}_n - C)$. The presence of the terms multiplied by the partial derivatives estimates is explained by the fact that the margins are unknown. Theoretical foundations can be found in Tsukahara [2005] and in Segers [2012].

In Section 6.5 it is shown that

$$\mathbb{A}_n^{\mathrm{P},(k)}(\boldsymbol{w}) = -(\hat{A}_n^{\mathrm{P}}(\boldsymbol{w}))^2 \int_0^1 \mathbb{C}_n^{(k)}(u^{w_1}, \dots, u^{w_p}) \, \frac{\mathrm{d}u}{u}$$

and

$$\mathbb{A}_n^{\mathrm{CFG},(k)}(\boldsymbol{w}) = (\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w})) \int_0^1 \mathbb{C}_n^{(k)}(u^{w_1}, \dots, u^{w_p}) \, \frac{\mathrm{d}u}{u \log u},$$

represent approximately independent random from samples of $\mathbb{A}^{\mathrm{P}}$ and $\mathbb{A}^{\mathrm{CFG}}$, the weak limits of $\mathbb{A}_n^{\mathrm{P}} = n^{1/2}(\hat{A}_n^{\mathrm{P}} - A)$ and $\mathbb{A}_n^{\mathrm{CFG}} = n^{1/2}(\hat{A}_n^{\mathrm{CFG}} - A)$.

6.4 Simulation results

In order to investigate the finite-sample performance of the test procedure presented in Section 6.2, we will generate under H_0 random samples from the Gumbel-Hougaard (logistic) copula (GH) and under the alternative hypothesis we consider random vectors the normal copula (N), the t-copula (t) with four degrees of freedom and the Clayton copula (C). Random sample generators for all these copulas can be found in the package copula presented in Kojadinovic and Yan [2010a]. The degree of dependence of the different copulas is measured hereafter by a multivariate generalization of Kendall's tau presented in Schmid et al. [2010]

$$\tau = \frac{1}{2^{p-1}-1}\left\{2^p \,\mathrm{E}[C(\boldsymbol{u})] - 1\right\}. \tag{6.8}$$

For all studies, we will generate 1000 datasets of size $n \in \{100, 200, 400\}$. The number of multiplier iterations N will be chosen equal to 1000 and the significance level for the test is set equal to 5 %.

The Pickands based test statistic introduced in this paper being too conservative for small samples, it is preferable to work with the modified, but asymptotically equivalent, test statistic $T_n^A = \frac{n+15}{n} T_n$. The test statistic T_n^{MS} based on the max-stable characterization of EV-copulas presented in Kojadinovic et al. [2011] is implemented in the package copula cited before.

The results are presented in Table [6.1,6.2,6.3]. In summary, both tests tend to be too conservative for copulas with high degree of dependence. Nevertheless, the newly proposed test is in most cases more powerful under the alternative hypothesis. Observe also that the power of the test increases with the dimension, which can be explained by the fact that the null hypothesis becomes more restrictive in higher dimensions. In fact, the constraint of being a multivariate extreme-value copula requires all bivariate marginal distributions to be also extreme-value.

6.5 Asymptotic results

6.5.1 Partial derivatives of A

In general, the partial derivatives of a Pickands dependence function of an extreme-value copula function might not exist in all vertices $\boldsymbol{e_j}$ of the unit-simplex Δ_{p-1}.

Example 6.1. To illustrate this, consider the asymmetric logistic dependence function introduced in Tawn [1990], of which we will use in this paper a more parsimonious version

$$A(\boldsymbol{w}) = (\theta^{1/\alpha} w_1^{1/\alpha} + \phi^{1/\alpha} w_2^{1/\alpha})^\alpha + (\theta^{1/\alpha} w_2^{1/\alpha} + \phi^{1/\alpha} w_3^{1/\alpha})^\alpha + (\theta^{1/\alpha} w_3^{1/\alpha} + \phi^{1/\alpha} w_1^{1/\alpha})^\alpha + \psi(w_1^{1/\alpha} + w_2^{1/\alpha} + w_3^{1/\alpha})^\alpha + 1 - \theta - \phi - \psi, \qquad \boldsymbol{w} \in \Delta_2 \quad (6.9)$$

with parameter vector $(\alpha, \theta, \phi, \psi) \in (0,1] \times [0,1]^3$. Let us focus on the second term

$$A_2^{asy,2}(\boldsymbol{w}) = (\theta^{1/\alpha} w_2^{1/\alpha} + \phi^{1/\alpha} w_3^{1/\alpha})^\alpha$$

whose partial derivative w.r.t. w_1

$$\dot{A}_1^{asy,2}(\boldsymbol{w}) = -\frac{\phi}{\left\{1 + \left(\frac{\theta}{\phi}\frac{w_2}{(1-w_1-w_2)}\right)^{1/\alpha}\right\}^{1-\alpha}}, \tag{6.10}$$

for $\boldsymbol{w}$ in the interior of the unit-simplex Δ_2. To see that the partial derivative of $A_{asy,2}$ does not exist in $\boldsymbol{e_1}$, approach the vertex $\boldsymbol{e_1}$ in (6.10) from two different paths $(1-2\epsilon, \epsilon, \epsilon)$ and $(1-3\epsilon, 2\epsilon, \epsilon)$ for $\epsilon > 0$ sufficiently small. □

Remark that as A represents the restriction of ℓ to the unit-simplex Δ_{p-1}, one can see that $\dot{A}_1 = \dot{\ell}_1 - \dot{\ell}_3$ is bounded as the partial derivatives of ℓ are bounded between 0 and 1. The latter property is a consequence of the fact that

$$\ell(\boldsymbol{x}) = \lim_{s \downarrow 0} \frac{1}{s}\{1 - C(1 - sx_1, \dots, 1 - sx_p)\}, \qquad \boldsymbol{x} \in [0, \infty)^p,$$

implying that ℓ is nondecreasing and Lipschitz in each of its arguments with Lipschitz constant equal to one.

All the other terms in (6.9) do not present anomalies in $\boldsymbol{e_1}$.

For a fixed value $\delta > 0$, define

$$\Delta_{p-1}^\delta = \Delta_{p-1} \setminus \left\{ \bigcup_{j=1}^p B_2(\boldsymbol{e_j}, \delta) \right\},$$

where $B_2(\boldsymbol{e_j}, \delta) = \{\boldsymbol{v} \in \Delta_{p-1} : \|\boldsymbol{v} - \boldsymbol{e_j}\|_2 \leqslant \delta\}$ with $j \in \{1, \dots, p\}$.

Theorem 6.2. *Under Assumption 4.3, we have for* $j \in \{1, \ldots, p-1\}$

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}^{\delta}} \left| \hat{A}^{[j]}(\boldsymbol{w}) - \dot{A}_j(\boldsymbol{w}) \right| \xrightarrow{\mathrm{P}} 0.$$

Proof. Consider $\boldsymbol{w} \in \Delta_{p-1}^{\delta}$ and without loss of generality take $j = 1$ and set $\boldsymbol{w_{-1}} = (w_2, \ldots, w_{p-1})$. Then,

$$\begin{aligned}
\hat{A}_n^{[1]}(\boldsymbol{w}) - \dot{A}_1(\boldsymbol{w}) = & \frac{\hat{A}_n(w_{1,n}^{+}, \boldsymbol{w_{-1}}) - A(w_{1,n}^{+}, \boldsymbol{w_{-1}})}{w_{1,n}^{+} - w_{1,n}^{-}} \\
& - \frac{\hat{A}_n(w_{1,n}^{-}, \boldsymbol{w_{-1}}) - A(w_{1,n}^{-}, \boldsymbol{w_{-1}})}{w_{1,n}^{+} - w_{1,n}^{-}} \\
& + \frac{A(w_{1,n}^{+}, \boldsymbol{w_{-1}}) - A(w_{1,n}^{-}, \boldsymbol{w_{-1}})}{w_{1,n}^{+} - w_{1,n}^{-}} - \dot{A}_1(\boldsymbol{w}) \qquad (6.11)
\end{aligned}$$

If we suppose that A is continuously differentiable w.r.t. w_1 on Δ_{p-1}^{δ}, the mean-value theorem guarantees the existence of $w_{1,n}^{*} \in (w_{1,n}^{-}, w_{1,n}^{+})$ such that

$$A(w_{1,n}^{+}, \boldsymbol{w_{-1}}) - A(w_{1,n}^{-}, \boldsymbol{w_{-1}}) = \dot{A}_j(w_{1,n}^{*}, \boldsymbol{w_{-1}})(w_{1,n}^{+} - w_{1,n}^{-}),$$

and

$$\begin{aligned}
& \sup_{\boldsymbol{w} \in \Delta_{p-1}^{\delta}} \left| \dot{A}_1(w_{1,n}^{*}, \boldsymbol{w_{-1}}) - \dot{A}^1(\boldsymbol{w}) \right| \\
\leqslant & \sup_{\substack{\boldsymbol{w}, \boldsymbol{w}' \in \Delta_{p-1}^{\delta} \\ |w_1' - w_1| \leqslant 2n^{-1/2}}} \left| \dot{A}_1(\boldsymbol{w}) - \dot{A}_1(\boldsymbol{w}') \right|,
\end{aligned}$$

which tends to 0 from the differentiability assumptions on A. Furthermore, as $n^{-1/2} \leqslant w_{1,n}^{+} - w_{1,n}^{-} \leqslant 2n^{-1/2}$, the first two terms in (6.11) can be bounded by

$$\sup_{\boldsymbol{w} \in \Delta_{p-1}^{\delta}} \left| \mathbb{A}_n(w_{1,n}^{+}, \boldsymbol{w_{-1}}) - \mathbb{A}_n(w_{1,n}^{-}, \boldsymbol{w_{-1}}) \right|.$$

This quantity tends to 0 from the asymptotic results given in Theorem 4.7. □

6.5.2 Resampling

For technical reasons, we impose on estimates of A and the estimates of the first-order partial derivatives of A the following upper and lower limits induced by the properties of A, i.e. $\forall \boldsymbol{w} \in \Delta_{p-1}$

$$\hat{A}_n(\boldsymbol{w}) = (\hat{A}_n^{\mathrm{CFG,P}}(\boldsymbol{w}) \wedge 1) \vee \max(w_1, \ldots, w_p) \tag{6.12}$$

and

$$-1 \leqslant \hat{A}_n^{[j]}(\boldsymbol{w}) \leqslant 1, \quad \forall j \in \{1, \ldots, p\}, \tag{6.13}$$

where $\hat{A}_n$ will stand for both (CFG and Pickands) constrained estimators from now on.

Theorem 6.3. *Under Assumption 4.3,*

$$(\mathbb{A}_n^{\mathrm{P}}, \mathbb{A}_n^{\mathrm{P},(1)}, \ldots, \mathbb{A}_n^{\mathrm{P},(N)}) \rightsquigarrow (\mathbb{A}^{\mathrm{P}}, \mathbb{A}^{\mathrm{P},(1)}, \ldots, \mathbb{A}^{\mathrm{P},(N)}) \tag{6.14}$$

and

$$(\mathbb{A}_n^{\mathrm{CFG}}, \mathbb{A}_n^{\mathrm{CFG},(1)}, \ldots, \mathbb{A}_n^{\mathrm{CFG},(N)}) \rightsquigarrow (\mathbb{A}^{\mathrm{CFG}}, \mathbb{A}^{\mathrm{CFG},(1)}, \ldots, \mathbb{A}^{\mathrm{CFG},(N)}) \tag{6.15}$$

in $\mathscr{C}(\Delta_{p-1})^{\otimes(N+1)}$*, where for all* $k \in \{1, \ldots, N\}$*, we have* $\mathbb{A}^{\mathrm{P},(k)}$ *and* $\mathbb{A}^{\mathrm{CFG},(k)}$ *are independent copies of* $\mathbb{A}^{\mathrm{P}}$ *respectively* $\mathbb{A}^{\mathrm{CFG}}$.

Proof. Recall the definition of the process $\mathbb{G}_{n,\theta}$ on $[0,1]^p$ given in (4.16)

$$\mathbb{G}_{n,\theta}(\boldsymbol{u}) = \begin{cases} \frac{\alpha_n(\boldsymbol{u})}{q_\theta(\min(\boldsymbol{u}))} = \frac{1}{n^{1/2}} \sum_{i=1}^n f_{\boldsymbol{u},\theta}(U_{i,1}, \ldots, U_{i,p}) & \text{if } \boldsymbol{u} \in \mathbb{E}, \\ 0 & \text{if } \boldsymbol{u} \in [0,1]^p \setminus \mathbb{E}. \end{cases}$$

In Lemma 4.8 it was proved that $\mathbb{G}_{n,\theta} \rightsquigarrow \mathbb{G}_\theta$ in $\ell^\infty([0,1]^p)$ with the trajectories of $\mathbb{G}_\theta$ being continuous almost surely.

Based on Theorem 10.1 in Kosorok [2008] one can show that

$$(\mathbb{G}_{n,\theta}, \mathbb{G}_{n,\theta}^{(1)}, \ldots, \mathbb{G}_{n,\theta}^{(N)}) \rightsquigarrow (\mathbb{G}_\theta, \mathbb{G}_\theta^{(1)}, \ldots, \mathbb{G}_\theta^{(N)}) \quad \text{in} \quad \ell^\infty([0,1]^p)^{\otimes(N+1)}.$$

For $\boldsymbol{w} \in \Delta_{p-1}$, let us define

$$\begin{aligned} \phi_{n,1}^{A}(\mathbb{G}_{n,\theta}^{(k)}) &= \mu(\boldsymbol{w}) \int_0^{p\log(n+1)} \mathbb{G}_{n,\theta,1}^{(k)}(e^{-sw_1}) q_\theta(e^{-sw_1}) e^{-s(A(\boldsymbol{w})-w_1)}\, h(s)\, \mathrm{d}s \\ \phi_{n,1}^{\hat{A}_n}(\mathbb{G}_{n,\theta}^{(k)}) &= \hat{\mu}(\boldsymbol{w}) \int_0^{p\log(n+1)} \mathbb{G}_{n,\theta,1}^{(k)}(e^{-sw_1}) q_\theta(e^{-sw_1}) e^{-s(\hat{A}_n(\boldsymbol{w})-w_1)}\, h(s)\, \mathrm{d}s, \end{aligned}$$

where $\mathbb{G}_{n,\theta,1}(u_1) = \mathbb{G}_{n,\theta}(u_1, 1, \dots, 1)$ for $u_1 \in [0,1]$ with weak limit $G_{\theta,1}$ (the notation with superscript (k) representing the bootstrap replications) and

$$\mu(\boldsymbol{w}) = A(\boldsymbol{w}) + (1 - w_1)\dot{A}_1(\boldsymbol{w}) - \sum_{i=2}^{p-1} w_i \dot{A}_i(\boldsymbol{w}) \tag{6.16}$$

and

$$\hat{\mu}(\boldsymbol{w}) = \hat{A}_n(\boldsymbol{w}) + (1 - w_1)\hat{A}_n^{[1]}(\boldsymbol{w}) - \sum_{i=2}^{p-1} w_i \hat{A}_n^{[i]}(\boldsymbol{w}). \tag{6.17}$$

The function $h(\cdot)$ is defined on $(0, \infty)$ equal to $h^{\mathrm{P}}(s) = 1$ for the Pickands estimator and $h^{\mathrm{CFG}}(s) = \frac{1}{s}$ for the CFG-estimator.

Let us show that the difference between $\phi_{n,1}^{\hat{A}_n}$ and $\phi_{n,1}^{A}$ is asymptotically negligible using the following decomposition

$$\begin{aligned}
& \mu(\boldsymbol{w}) \int_0^{p\log(n+1)} \mathbb{G}_{n,\theta,1}^{(k)}(e^{-sw_1})\, q_\theta(e^{-sw_1})\, e^{-s(A(\boldsymbol{w})-w_1)}\, h(s)\,\mathrm{d}s \\
- \;& \mu(\boldsymbol{w}) \int_0^{p\log(n+1)} \mathbb{G}_{n,\theta,1}^{(k)}(e^{-sw_1})\, q_\theta(e^{-sw_1})\, e^{-s(\hat{A}_n(\boldsymbol{w})-w_1)}\, h(s)\,\mathrm{d}s \\
+ \;& \mu(\boldsymbol{w}) \int_0^{p\log(n+1)} \mathbb{G}_{n,\theta,1}^{(k)}(e^{-sw_1})\, q_\theta(e^{-sw_1})\, e^{-s(\hat{A}_n(\boldsymbol{w})-w_1)}\, h(s)\,\mathrm{d}s \\
- \;& \hat{\mu}(\boldsymbol{w}) \int_0^{p\log(n+1)} \mathbb{G}_{n,\theta,1}^{(k)}(e^{-sw_1})\, q_\theta(e^{-sw_1})\, e^{-s(\hat{A}_n(\boldsymbol{w})-w_1)}\, h(s)\,\mathrm{d}s.
\end{aligned}$$

The difference between the first two terms can be bounded by

$$\|\mu\|_\infty \, \|\mathbb{G}_{n,\theta}^{(k)}\|_\infty \int_0^{p\log(n+1)} q_\theta(e^{-sw_1})\, e^{sw_1} \left| e^{-s\hat{A}_n(\boldsymbol{w})} - e^{-sA(\boldsymbol{w})} \right| h(s)\,\mathrm{d}s.$$

Using the inequality

$$\left| e^{-x} - e^{-y} \right| \leqslant e^{-x \wedge y} \left| x - y \right|,$$

the following upper bound can be found

$$\begin{aligned}
& \|\mu\|_\infty \, \|\mathbb{G}_{n,\theta}^{(k)}\|_\infty \left| \hat{A}_n(\boldsymbol{w}) - A(\boldsymbol{w}) \right| \\
& \qquad \int_0^{p\log(n+1)} s\, q_\theta(e^{-sw_1})\, e^{-s\{\hat{A}_n(\boldsymbol{w}) \wedge A(\boldsymbol{w}) - w_1\}}\, h(s)\,\mathrm{d}s.
\end{aligned} \tag{6.18}$$

The integral part in (6.18) is finite as $\hat{A}_n(\boldsymbol{w}) \geqslant w_1 \vee \ldots \vee w_p$ by construction and hence can be bounded by $\int_0^\infty s\, K_\theta(s)\, h(s)\, \mathrm{d}s$, where

$$K_\theta(s) = s^\theta \mathbf{1}_{(0,1]}(s) + e^{-(\theta/p)s} \mathbf{1}_{(1,+\infty)}(s), \quad s \in (0, +\infty).$$

The uniform convergence in probability of $\hat{A}_n$ is sufficient to show that the difference converges uniformly to 0 in probability on the unit-simplex.

The difference between the third and fourth term can be bounded by

$$\left|\hat{A}_n(\boldsymbol{w}) - A(\boldsymbol{w})\right| + (1-w_1)\left|\hat{A}_n^{[1]}(\boldsymbol{w}) - \dot{A}_1(\boldsymbol{w})\right| + \sum_{i=2}^{p-1} w_i \left|\hat{A}_n^{[i]}(\boldsymbol{w}) - \dot{A}_i(\boldsymbol{w})\right|$$
$$\|\mathbb{G}_{n,\theta}^{(k)}\|_\infty \int_0^{p\log(n+1)} q_\theta(e^{-sw_1})\, e^{-s(\hat{A}_n(\boldsymbol{w})-w_1)}\, h(s)\, \mathrm{d}s. \quad (6.19)$$

If $\boldsymbol{w} \in \Delta_{p-1}^\delta$, the convergence properties of $\hat{A}_n$ and $\hat{A}_n^{[j]}$ permit to conclude as the integral part in Expression (6.19) can be uniformly bounded.

In the case where $\boldsymbol{w} \in \Delta_{p-1} \cap B_2(\boldsymbol{e_1}, \delta)$, it can be shown using the bounds given in (6.12) and (6.13) that the term

$$\sup_{\boldsymbol{w} \in \Delta_{p-1} \cap B_2(\boldsymbol{e_1}, \delta)} \left|\hat{A}_n(\boldsymbol{w}) - A(\boldsymbol{w})\right| + (1 - w_1)$$
$$\left|\hat{A}_n^{[1]}(\boldsymbol{w}) - \dot{A}_1(\boldsymbol{w})\right| + \sum_{i=2}^{p-1} w_i \left|\hat{A}_n^{[i]}(\boldsymbol{w}) - \dot{A}_i(\boldsymbol{w})\right|$$

can be made arbitrarily small for a sufficiently small δ.

Alternatively, let us focus on the situation where $\boldsymbol{w} \in \Delta_{p-1} \cap B_2(\boldsymbol{e_2}, \delta)$. Fix $\epsilon > 0$, then the lim sup of the probability

$$\mathrm{P}\left(\sup_{\boldsymbol{w} \in \Delta_{p-1} \cap B_2(\boldsymbol{e_2}, \delta)} \int_0^{+\infty} |\mathbb{G}_{n,\theta,1}(e^{-sw_1})|\, K_\theta(s)\, h(s)\mathrm{d}s > \epsilon\right)$$

can be bounded by inovking the Portmanteau lemma

$$\mathrm{P}\left(\sup_{\boldsymbol{w} \in \Delta_{p-1} \cap B_2(\boldsymbol{e_2}, \delta)} \int_0^{+\infty} |\mathbb{G}_{\theta,1}(e^{-sw_1})|\, K_\theta(s)\, h(s)\mathrm{d}s \geqslant \epsilon\right)$$

which can again be made arbitrarily small by choosing δ. This can be seen from the continuity properties of the function

$$f_g(x) = \int_0^\infty g\left(e^{-s\,x}\right)\, K_\theta(s)\, h(s)\, \mathrm{d}s, \quad x \in [0, 1]$$

where $g \in \mathscr{C}([0,1])$ with $g(1) = 0$.

The same kind of reasoning can be repeated for $(\phi_{n,2}^{\hat{A}_n} - \phi_{n,2}^{A}, \ldots, \phi_{n,p-1}^{\hat{A}_n} - \phi_{n,p-1}^{A})$, the respective definitions being obvious. To be complete, let us define

$$\phi_{n,0}^{A}(\mathbb{G}_{n,\theta}) = \phi_{n,0}^{\hat{A}_n}(\mathbb{G}_{n,\theta}) = \int_0^\infty \mathbb{G}_{n,\theta}\left(e^{-sw_1}, \ldots, e^{-sw_p}\right) h(s)\,\mathrm{d}s.$$

Using the extended continuous mapping theorem (van der Vaart and Wellner [1996], Theorem 1.11.1), it can be shown that for $\phi_n^A = \sum_{i=0}^{p-1} \phi_{n,i}^A$, we have

$$\left(\phi_n^A(\mathbb{G}_{n,\theta}), \phi_n^A(\mathbb{G}_{n,\theta}^{(1)}), \ldots, \phi_n^A(\mathbb{G}_{n,\theta}^{(N)})\right)$$
$$\rightsquigarrow \left(\int_0^1 \mathbb{C}(u^{\boldsymbol{w}})\frac{\mathrm{d}u}{u}, \int_0^1 \mathbb{C}^{(1)}(u^{\boldsymbol{w}})\frac{\mathrm{d}u}{u}, \ldots, \int_0^1 \mathbb{C}^{(N)}(u^{\boldsymbol{w}})\frac{\mathrm{d}u}{u}\right),$$

in $\mathscr{C}(\Delta_{p-1})^{\otimes(N+1)}$, where $(\mathbb{C}^{(1)}, \ldots, \mathbb{C}^{(N)})$ are independent and asymptotic equivalent copies of $\mathbb{C}$ defined in (4.11). From the previous development it is clear that the convergence holds also for

$$\left(\phi_n^{\hat{A}_n}(\mathbb{G}_{n,\theta}), \phi_n^{\hat{A}_n}(\mathbb{G}_{n,\theta}^{(1)}), \ldots, \phi_n^{\hat{A}_n}(\mathbb{G}_{n,\theta}^{(N)})\right),$$

where $\phi_n^{\hat{A}_n} = \sum_{i=0}^{p-1} \phi_{n,i}^{\hat{A}_n}$.

The final statements in the theorem can be deduced from further application of the continuous mapping theorem. □

Theorem 6.4. *Suppose Assumption 4.3 is verified. Then under H_0*

$$\left(\mathbb{T}_n, \mathbb{T}_n^{(1)}, \ldots, \mathbb{T}_n^{(N)}\right) \rightsquigarrow \left(\mathbb{T}, \mathbb{T}^{(1)}, \ldots, \mathbb{T}^{(N)}\right)$$

in $[0,\infty)^{\otimes(N+1)}$, where

$$\mathbb{T} = \int_{[0,1]^p} \{\mathbb{D}(\boldsymbol{u})\}^2 \mathrm{d}C(\boldsymbol{u}).$$

Proof. In Chapter 4, it was proved that $n^{1/2}(\mathbb{C} - C, \hat{A}_n - A) \rightsquigarrow (\mathbb{C}, \mathbb{A})$ in $\ell^\infty([0,1]^p) \times \mathscr{C}(\Delta_{p-1})$. Define the map $\phi : \mathscr{C}(\Delta_{p-1}) \to \ell^\infty([a,b]^p)$, such that for $f \in \mathscr{C}(\Delta_{p-1})$ the map $\phi(f)$ admits the following expression

$$\phi(f)(\boldsymbol{u}) = \exp\left\{\sum_{j=1}^{p} \log(u_j)\, f\left(\frac{\log u_1}{\sum_{j=1}^{p}\log(u_j)}, \dots, \frac{\log u_{p-1}}{\sum_{j=1}^{p}\log(u_j)}\right)\right\},$$

$\forall \boldsymbol{u} \in \mathbb{E}_p$. The Hadamard derivative of ϕ in A results into a linear functional $\phi_A^{'}$

$$\phi_A^{'}(f) = \exp\left(\sum_{j=1}^{p} \log(u_j)\, A\left(\boldsymbol{w_u}\right)\right)\left(\sum_{j=1}^{p} \log u_j\right) f(\boldsymbol{w_u}), \quad f \in \mathscr{C}(\Delta_{p-1}).$$

More explanations on this follow in Lemma 6.6. Invoking the continuous mapping theorem, one can see

$$n^{1/2}(\mathbb{C} - C, \phi(\hat{A}_n) - \phi(A)) \rightsquigarrow (\mathbb{C}, \phi_A^{'}(\mathbb{A})) \text{ in } \ell^\infty[0,1]^p \times \mathscr{C}(\mathbb{E}_p).$$

The continuous mapping theorem allows to conclude

$$\begin{aligned} n^{1/2}\left(\mathbb{C} - C\right) &- n^{1/2}\left(\phi(\hat{A}_n) - \phi(A)\right) \\ &\rightsquigarrow \mathbb{C} - \exp\left\{\left(\sum_{j=1}^{p}\log u_j\right) A(\boldsymbol{w_u})\right\}\left(\sum_{j=1}^{p}\log u_j\right)\mathbb{A}(\boldsymbol{w_u}) \end{aligned} \tag{6.20}$$

in $\ell^\infty(\mathbb{E}_p)$. Application of the continuous mapping theorem permits to conclude. □

The inequality in Lemma 6.5, suggested by Johan Segers, plays a key role in the proof of the Hadamard-differentiability of ϕ.

Lemma 6.5. *For all $x \in \mathbb{R}\setminus\{0\}$,*

$$\left|\frac{e^x - 1}{x} - 1\right| \leqslant |x|\, e^{|x|}. \tag{6.21}$$

As a consequence, for all $a \in \mathbb{R}$ and all $x \in \mathbb{R}\setminus\{0\}$,

$$\left|\frac{e^{ax} - 1}{x} - a\right| \leqslant a^2\, |x|\, e^{|ax|}. \tag{6.22}$$

Proof. We have

$$\begin{aligned}
\frac{e^x - 1}{x} - 1 &= \frac{1}{x}\int_0^x e^y \, dy - 1 \\
&= \frac{1}{x}\int_0^x (e^y - 1) \, dy \\
&= \frac{1}{x}\int_{y=0}^x \int_{z=0}^y e^z \, dz \, dy \\
&= \frac{1}{x}\int_{z=0}^x e^z \int_{y=z}^x dy \, dz \\
&= \frac{1}{x}\int_0^x (x - z) e^z \, dz.
\end{aligned}$$

If $x > 0$, it follows that, as $0 \leqslant x - z \leqslant x$ for all $z \in [0, x]$,

$$\left| \frac{e^x - 1}{x} - 1 \right| \leqslant \int_0^x e^z \, dz \leqslant x \, e^x.$$

If $x < 0$, we find, as $0 \leqslant z - x \leqslant |x|$ for all $z \in [x, 0]$,

$$\left| \frac{e^x - 1}{x} - 1 \right| \leqslant \frac{1}{|x|} \int_x^0 (z - x) e^z \, dz \leqslant |x|.$$

The two previous displays together constitute equation (6.21). Since

$$\left| \frac{e^{ax} - 1}{x} - a \right| = |a| \left| \frac{e^{ax} - 1}{ax} - 1 \right|$$

we arrive at equation (6.22). □

Lemma 6.6. *The functional ϕ defined as*

$$\phi(h)(\boldsymbol{u}) = \exp\left\{ \sum_{j=1}^p \log(u_j) \, h\left(\frac{\log u_1}{\sum_{j=1}^p \log(u_j)}, \ldots, \frac{\log u_{p-1}}{\sum_{j=1}^p \log(u_j)} \right) \right\},$$

for $h \in \mathscr{C}(\Delta_{p-1})$ and $\boldsymbol{u} \in \mathbb{E}_p$. is Hadamard differentiable tangentially to $\mathscr{C}(\Delta_{p-1})$. In particular, the Hadamard derivative of ϕ in the Pickands dependence function A results into the linear function $\phi'_A : \mathscr{C}(\Delta_{p-1}) \to \ell^\infty(\mathbb{E}_p)$ given by

$$\phi'_A(h)(\boldsymbol{u}) = \exp\left(\sum_{j=1}^p \log(u_j) \, A\left(\boldsymbol{w_u}\right) \right) \left(\sum_{j=1}^p \log u_j \right) h(\boldsymbol{w_u}), \qquad (6.23)$$

for $h \in \mathscr{C}(\Delta_{p-1})$ and $\boldsymbol{u} \in \mathbb{E}_p$.

Proof. Consider the sequences $(t_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ and $(h_n)_{n\in\mathbb{N}}$ in $\mathscr{C}(\Delta_{p-1})$ endowed with the uniform topology with $t_n \to 0$ and $\|h_n - h\|_\infty \to 0$ as n tends to infinity.

We have

$$\begin{aligned}\frac{\phi(A + t_n h_n)(\boldsymbol{u}) - \phi(A)(\boldsymbol{u})}{t_n}&\\ &= \phi(A)(\boldsymbol{u})\,\frac{1}{t_n}\left[\exp\left\{\sum_{j=1}^{p}\log(u_j)\,t_n\,h_n(\boldsymbol{w_u})\right\} - 1\right]. \end{aligned} \tag{6.24}$$

We now apply equation (6.22) to the right-hand side of equation (6.24), with $x = t_n$ and $a = h_n(\boldsymbol{w_u})\sum_{j=1}^{p}\log(u_j)$:

$$\begin{aligned}&\left|\frac{\phi(A + t_n h_n)(\boldsymbol{u}) - \phi(A)(\boldsymbol{u})}{t_n} - \phi(A)(\boldsymbol{u})\,h_n(\boldsymbol{w_u})\sum_{j=1}^{p}\log(u_j)\right|\\ &= \phi(A)(\boldsymbol{u})\left(h_n(\boldsymbol{w_u})\sum_{j=1}^{p}\log(u_j)\right)^2 |t_n|\,\exp\left\{|t_n|\,|h_n(\boldsymbol{w_u})|\sum_{j=1}^{p}|\log(u_j)|\right\}.\end{aligned} \tag{6.25}$$

Fix $0 < \varepsilon < 1/p$. For n sufficiently large, we have

$$|t_n|\,|h_n(\boldsymbol{w_u})| \leqslant \varepsilon, \qquad \boldsymbol{u} \in \mathbb{E}_p,$$

and thus

$$\exp\left\{|t_n|\,|h_n(\boldsymbol{w_u})|\sum_{j=1}^{p}|\log(u_j)|\right\} \leqslant (u_1 \cdots u_p)^{-\varepsilon}, \qquad \boldsymbol{u} \in \mathbb{E}_p. \tag{6.26}$$

Since A is a Pickands dependence function, we have $A \geqslant 1/p$ and therefore

$$\phi(A)(\boldsymbol{u}) = (u_1 \cdots u_p)^{A(\boldsymbol{w_u})} \leqslant (u_1 \cdots u_p)^{1/p}. \tag{6.27}$$

From equations (6.25), (6.26) and (6.27), it follows that, as $n \to \infty$,

$$\sup_{u\in\mathbb{E}_p}\left|\frac{\phi(A + t_n h_n)(\boldsymbol{u}) - \phi(A)(\boldsymbol{u})}{t_n} - \phi(A)(\boldsymbol{u})\,h_n(\boldsymbol{w_u})\sum_{j=1}^{p}\log(u_j)\right| \to 0. \tag{6.28}$$

Moreover, since

$$\sup_{\boldsymbol{u}\in\mathbb{E}_p}\left|\phi(A)(\boldsymbol{u})\sum_{j=1}^{p}\log(u_j)\right| \leqslant \sup_{\boldsymbol{u}\in\mathbb{E}_p}\left|(u_1\cdots u_p)^{1/p}\sum_{j=1}^{p}\log(u_j)\right| < \infty,$$

it follows, by the triangle inequality and the fact that $h_n \to h$ uniformly, that

$$\sup_{u\in\mathbb{E}_p}\left|\frac{\phi(A+t_nh_n)(\boldsymbol{u})-\phi(A)(\boldsymbol{u})}{t_n} - \phi(A)(\boldsymbol{u})\,h(\boldsymbol{w_u})\sum_{j=1}^{p}\log(u_j)\right| \to 0. \tag{6.29}$$

□

		$n = 100$		$n = 200$		$n = 400$	
		T^A	T^{MS}	T^A	T^{MS}	T^A	T^{MS}
GH	$\tau = 0.25$	5.8	3.6	5.6	5.2	4.1	4.0
	$\tau = 0.50$	3.2	2.5	4.3	2.7	4.3	3.2
	$\tau = 0.75$	2.1	0.3	1.8	0.5	2.7	0.7
C	$\tau = 0.25$	92.9	91.8	99.9	99.9	100	100
	$\tau = 0.50$	100	100	100	100	100	100
	$\tau = 0.75$	100	100	100	100	100	100
N	$\tau = 0.25$	33.6	35.8	61.1	61.6	90.7	88.3
	$\tau = 0.50$	48.9	32.7	78.6	62.9	98.0	93.7
	$\tau = 0.75$	33.0	10.3	72.0	31.4	97.3	77.4
t	$\tau = 0.25$	21.1	16.5	32.3	24.4	53.0	38.9
	$\tau = 0.50$	34.7	19.5	61.5	40.4	88.0	69.5
	$\tau = 0.75$	22.4	5.7	54.8	15.1	88.5	50.0

Table 6.1: Empirical rejection rates for dimension $p = 3$.

		$n = 100$		$n = 200$		$n = 400$	
		T^A	T^{MS}	T^A	T^{MS}	T^A	T^{MS}
GH	$\tau = 0.25$	6.7	5.1	4.9	3.6	5.1	4.3
	$\tau = 0.50$	2.9	2.8	3.2	1.9	3.1	1.7
	$\tau = 0.75$	0.4	0.2	0.6	0.1	1.7	0.2
C	$\tau = 0.25$	98.2	98.2	100	100	100	100
	$\tau = 0.50$	100	100	100	100	100	100
	$\tau = 0.75$	100	100	100	100	100	100
N	$\tau = 0.25$	44.1	45.1	77.4	75.4	98.5	97.5
	$\tau = 0.50$	53.2	36.7	87.3	69.8	99.5	96.5
	$\tau = 0.75$	6.7	5.1	69.8	21.4	98.3	70.8
t	$\tau = 0.25$	24.8	17.6	44.0	31.7	68.1	52.8
	$\tau = 0.50$	40.5	18.1	67.8	38.5	95.1	79.5
	$\tau = 0.75$	10.6	2.1	48.6	6.0	90.9	41.6

Table 6.2: Empirical rejection rates for dimension $p = 4$.

		$n = 100$		$n = 200$		$n = 400$	
		T^A	T^{MS}	T^A	T^{MS}	T^A	T^{MS}
GH	$\tau = 0.25$	6.3	5.0	5.4	4.2	5.7	4.8
	$\tau = 0.50$	2.2	1.6	3.9	2.3	3.3	1.8
	$\tau = 0.75$	0.3	0.0	0.5	0.0	2.0	0.1
C	$\tau = 0.25$	99.4	99.4	100	100	100	100
	$\tau = 0.50$	100	100	100	100	100	100
	$\tau = 0.75$	100	99.8	100	100	100	100
N	$\tau = 0.25$	55.6	53.3	90.0	86.8	99.6	99.4
	$\tau = 0.50$	56.6	35.2	91.9	73.8	99.9	98.0
	$\tau = 0.75$	9.2	1.9	65.4	10.7	97.9	53.8
t	$\tau = 0.25$	29.6	17.6	52.5	36.3	81.4	69.4
	$\tau = 0.50$	39.5	16.0	71.8	39.4	97.0	80.6
	$\tau = 0.75$	4.6	0.8	38.6	2.1	56.5	9.2

Table 6.3: Empirical rejection rates for dimension $p = 5$.

Conclusion

The conclusion part aims to enumerate and identify potential areas for further research and improvement in the domain of extreme-value copulas.

In Chapter 3 we obtained the lowest-variance estimator in the class of CFG-estimators given in (2.9) under the unrealistic assumption of known marginal distributions. However, up to my knowledge, the development of an efficient estimator for extreme-value copulas remains still an open question. As the class of Pickands dependence functions $\mathscr{A}$ is infinite-dimensional, we would need to rely on a (quasi-)maximum likelihood approach for reasons of feasibility under the necessary smoothness condtions. For example a B-spline basis could appear as a convenient choice to model the elements in $\mathscr{A}$ in the case $p = 2$.

The same methodology could also be extended to the case of higher dimensions. However, the first challenge consists in finding an adequate basis to model the spectral density h as one cannot simply use the discrete spectral measure from Chapter 4 for this purpose.

Alternatively, it might also be interesting to adopt the Bayesian perspective in this framework as the strength of a Bayesian approach lies in situations where many parameters are involved. For instance, one might get some inspiration in the paper of Lambert [2007] on Archimedean copulas.

In recent years, research on time-varying copulas has reached some popularity, mostly driven by the fact that a constant degree of dependence over time in financial time series might be considered as unrealistic; see for example Manner and Reznikova [2012] for a survey. In this context it might also be of interest to investigate a new class of time-varying extreme-value copulas and identify their possible domains of applications.

One can also be surprised by the relatively modest number of existing parametric models for higher-dimensional extreme-value copulas, hence leaving some room for innovation in this domain.

The different codes created in the context of this thesis will be regrouped in an R-package.

Bibliography

Abdous, B. and K. Ghoudi (2005). Non-parametric estimators of multivariate extreme dependence functions. *Journal of Nonparametric Statistics 17*(8), 915–935.

Asimit, A. V. and B. L. Jones (2007). Extreme behavior of bivariate elliptical distributions. *Insurance: Mathematics and Economics 41*, 53–61.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.

Berman, S. M. (1961/1962). Convergence to bivariate limiting extreme value distributions. *Annals of the Institute of Statistical Mathematics 13*(3), 217–223.

Boldi, M.-O. and A. C. Davison (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society, Series B 69*(2), 217–229.

Bücher, A., H. Dette, and S. Volgushev (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics (to appear).*

Capéraà, P. and A.-L. Fougères (2000). Estimation of a bivariate extreme value distribution. *Extremes 3*, 311–329.

Capéraà, P., A.-L. Fougères, and C. Genest (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika 84*, 567–577.

Capéraà, P., A.-L. Fougères, and C. Genest (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis 72*, 30–49.

Cebrián, A., M. Denuit, and P. Lambert (2003). Analysis of bivariate tail dependence using extreme values copulas: An application to the SOA medical large claims database. *Belgian Actuarial Journal 3*(1), 33–41.

Charpentier, A. and J. Segers (2008). Convergence of Archimedean copulas. *Statistics & Probability Letters 78*, 412–419.

Charpentier, A. and J. Segers (2009). Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis 100*, 1521–1537.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*(1), 141–151.

Coles, S. (2001). *An introduction to statistical modeling of extreme values.* Springer Series in Statistics. London: Springer-Verlag London Ltd.

Coles, S. G. and J. A. Tawn (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B 53*(2), 377–392.

Crowder, M. (1989). A multivariate distribution with Weibull connections. *J. Roy. Statist. Soc. Ser. B 51*(1), 93–107.

de Haan, L. and A. Ferreira (2006). *Extreme value theory.* Springer Series in Operations Research and Financial Engineering. New York: Springer.

de Haan, L., C. Neves, and L. Peng (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis 99*(6), 1260–1275.

de Haan, L. and S. I. Resnick (1977). Limit theorem for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 40*, 317–337.

Deheuvels, P. (1979). La fonction de dpendance empirique et ses proprits, un test non paramtrique d'indpendance. *Bull.Acad.Royale.Belg., Classe Sciences 65*, 274–292.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In J. Tiago de Oliveira (Ed.), *Statistical extremes and applications*, pp. 117–130. Reidel.

Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters 12*(5), 429–439.

Demarta, S. and A. McNeil (2005). The t-copula and related copulas. *International Statistical Review 73*, 111–129.

Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimates of the stable tail dependence function. *Journal of Multivariate Analysis 64*, 25–47.

Dupuis, D. J. and S. Morgenthaler (2002). Robust weighted likelihood estimators with an application to bivariate extreme value problems. *The Canadian Journal of Statistics 30*(1), 17–36.

Dupuis, D. J. and J. A. Tawn (2001). Effects of mis-specification in bivariate extreme value problems. *Extremes 4*, 315–330.

Einmahl, J. H. J., L. de Haan, and D. Li (2006). Weighted approximations to tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics 34*(4), 1987–2014.

Einmahl, J. H. J., L. de Haan, and V. I. Piterbarg (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics 29*(5), 1401–1423.

Einmahl, J. H. J., A. Krajina, and J. Segers (2008). A method of moments estimator of tail dependence. *Bernoulli 14*(4), 1003–1026.

Einmahl, J. H. J. and J. Segers (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics 37*(5B), 2953–2989.

Fang, K., S. Kotz, and K. W. Ng (1990). *Symmetric multivariate and related distributions.* Chapman and Hall (London and New York).

Fils-Villetard, A., A. Guillou, and J. Segers (2008). Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics 36*(3), 369–382.

Finkelstein, B. V. (1953). On the limiting distributions of the extreme terms of a variational series of a two-dimensional random quantity. *Dokladi Akademia SSSR 91*(2), 209–211. In Russian.

Fougères, A.-L., J. P. Nolan, and H. Rootzén (2009). Models for dependent extremes using stable mixtures. *Scandinavian Journal of Statistics 36*, 42–59.

Frank, M. (1979). On the simultaneous associativity of f(x,y) and x+y-f(x,y). *Aequationes Mathematicae 19*, 194–226.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *J. Amer. Statist. Assoc. 70*(351, part 1), 674–680.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics.* John Wiley & Sons, New York-Chichester-Brisbane. Wiley Series in Probability and Mathematical Statistics.

Galambos, J. (1987). *The asymptotic theory of extreme order statistics* (Second ed.). Melbourne, FL: Robert E. Krieger Publishing Co. Inc.

Geffroy, J. (1958). Contributions a la théorie des valeurs extrêmes. *Publ. Instit. Stat. Univ. Paris 7*, 37–121.

Geffroy, J. (1959). Contributions a la théorie des valeurs extrêmes. *Publ. Instit. Stat. Univ. Paris 8*, 123–184.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82*(3), 543–552.

Genest, C., J. Nelehov, and J. Ziegel (2011). Inference in multivariate archimedean copula models. *TEST 20*, 223–256.

Genest, C. and L.-P. Rivest (1989). A characterization of Gumbel's family of extreme value distributions. *Statistics & Probability Letters 8*(3), 207–211.

Genest, C. and L.-P. Rivest (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association 88*, 1034–1043.

Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *Annals of Statistics 37*(5B), 2990–3022.

Ghorbal, N. B., C. Genest, and J. Nešlehová (2009). On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. *The Canadian Journal of Statistics*.

Ghoudi, B., A.-L. Fougères, and K. Ghoudi (2005). Extreme behavior for bivariate elliptical distributions. *The Canadian Journal of Statistics 33*(3), 317–334.

Ghoudi, K., A. Khoudraji, and L.-P. Rivest (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist. 26*(1), 187–197.

Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In W. H. P. Jaworski, F. Durante and T. Rychlik (Eds.), *Proceedings of the Workshop on Copula Theory and its Applications*, pp. 127–146. Springer.

Gudendorf, G. and J. Segers (2011a). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *J. Multivariate Analysis 102*(1), 37–47.

Gudendorf, G. and J. Segers (2011b). Nonparametric estimation of multivariate extreme-value copulas. Technical report, Université Catholique de Louvain.

Guillem, A. G. (2000). Structure de dépendance des lois de valeurs extrêmes bivariées. *C. R. Acad. Sci. Paris, Série I*, 593–596.

Gumbel, E. J. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc. 55*, 698–707.

Gumbel, E. J. (1961). Bivariate logistic distributions. *J. Amer. Statist. Assoc. 56*, 335–349.

Gumbel, E. J. (1962). Multivariate extremal distributions. *Bull. Inst. Internat. Statist. 39*(livraison 2), 471–475.

Gumbel, E. J. and N. Goldstein (1964). Analysis of empirical bivariate extremal distributions. *Journal of the American Statistical Association 59*(307), 794–816.

Hall, P. and N. Tajvidi (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli 6*(5), 835–844.

Hashorva, E. (2005). Extremes of asymptotically spherical and elliptical random vectors. *Insurance: Mathematics and Economics 36*, 285–302.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika 73*(3), 671–678.

Hürlimann, W. (2003). Hutchinson–Lai's conjecture for bivariate extreme value copulas. *Statistics & Probability Letters 61*, 191–198.

Hüsler, J. and R. Reiss (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters 7*, 283–286.

Hutchinson, T. and C. Lai (1990). *Continuous Bivariate Distributions, Emphasizing Applications.* Adelaide: Rumbsy Scientific.

Jiménez, J. R., E. Villa-Diharce, and M. Flores (2001a). Nonparametric estimation of the dependence function in bivariate extreme value distributions. *Journal of Multivariate Analysis 76*(2), 159–191.

Jiménez, J. R., E. Villa-Diharce, and M. Flores (2001b). Nonparametric estimation of the dependence function in bivariate extreme value distributions. *Journal of Multivariate Analysis 76*(2), 159–191.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett. 9*(1), 75–81.

Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *Canad. J. Statist. 22*(1).

Joe, H., R. L. Smith, and I. Weissman (1992). Threshold methods for extremes. *Journal of the Royal Statistical Society, Series B 54*, 171–183.

Klüppelberg, C., G. Kuhn, and L. Peng (2007). Estimating the tail dependence function of an elliptical distribution. *Bernoulli 13*(1), 229–251.

Klüppelberg, C., G. Kuhn, and L. Peng (2008). Semi-parametric models for the multivariate tail dependence function—the asymptotically dependent case. *Scandinavian Journal of Statistics 35*(4), 701–718.

Kojadinovic, I., J. Segers, and J. Yan (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *Canadian Journal of Statistics 39*(4), 703–720.

Kojadinovic, I. and J. Yan (2010a). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software 34*(9), 1–20.

Kojadinovic, I. and J. Yan (2010b). Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis 101*(9), 2234 – 2249.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference; electronic version.* Springer Series in Statistics. Dordrecht: Springer.

Kotz, S. and S. Nadarajah (2000). *Extreme value distributions.* London: Imperial College Press. Theory and applications.

Lambert, P. (2007, August). Archimedean copula estimation using bayesian splines smoothing techniques. *Comput. Stat. Data Anal. 51*, 6307–6320.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*(1), 169–187.

Li, D. (2000). On default correlation: A copula function approach. *Journal of Fixed Income 9*, 43–54.

Li, H. (2009). Orthant tail dependence of multivariate extreme value distributions. *Journal of Multivariate Analysis 100*, 243–256.

Longin, F. and B. Solnik (2001). Extreme correlation of international equity markets. *The Journal of Finance 56*(2), 649–676.

Mai, J.-F. and M. Scherer (2009). Lévy-frailty copulas. *Journal of Multivariate Analysis 100*(7), 1567–1585.

Manner, H. and O. Reznikova (2012). A survey on time-varying copulas: Specification, simulations, and application. *Econometric Reviews 31*(6), 654–687.

Marshall, A. W. and I. Olkin (1967). A multivariate exponential distribution. *Journal of the American Statistical Association 62*, 30–44.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlquist (Ed.), *Spatial interaction theory and planning models*, pp. 75–96. Amsterdam: North-Holland.

McNeil, A. and J. Nešlehová (2009). Multivariate archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *The Annals of Statistics 37*(5B), 3059–3097.

Nelsen, R. B. (2006). *An introduction to copulas* (Second ed.). Springer Series in Statistics. New York: Springer.

Oakes, D. and A. K. Manatunga (1992). Fisher information for a bivariate extreme value distribution. *Biometrika 79*(4), 827–832.

Obretenov, A. (1991). On the dependence function of sibuya in multivariate extreme value theory. *Journal of Multivariate Analysis 36*(1), 35–43.

Peng, L., L. Qian, and J. Yang (2011). Weighted estimation of the dependence function for an extreme-value distribution. *Bernoulli ??*, ??

Pickands, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, Volume 49, pp. 859–878, 894–902. With a discussion.

Quessy, J. (2011). Testing for bivariate extreme dependence using kendall's process. *Scandinavian Journal of Statistics Vol: to appear*, pp: to appear.

Rémillard, B. and O. Scaillet (2009). Testing for equality between two copulas. *J. Multivar. Anal. 100*(3), 377–386.

Resnick, S. I. (1987). *Extreme Values,Regular Variation and Point Processes*, Volume 4 of *Springer Series in Operations Research and Financial Engineering*. New York: Springer.

Resnick, S. I. (2007). *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. New York: Springer. Probabilistic and statistical modeling.

Schlather, M. and J. A. Tawn (2002). Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes 5*, 87–102.

Schmid, F., R. Schmidt, T. Blumentritt, S. Gaisser, and M. Ruppert (2010). Copula-based measures of multivariate association. In W. H. P. Jaworski, F. Durante and T. Rychlik (Eds.), *Proceedings of the Workshop on Copula Theory and its Applications*, pp. 127–146. Springer.

Schmidt, R. and U. Stadtmüller (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics 33*(2), 307–335.

Segers, J. (2007). Non-parametric inference for bivariate extreme-value copulas. In M. Ahsanulah and S. Kirmani (Eds.), *Extreme Value Distributions*, Chapter 9, pp. 181–203. Nova Science Publishers, Inc. Older version available as CentER DP 2004-91, Tilburg University.

Segers, J. (2012). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli 18*, 764–782.

Shi, D. (1995). Fisher information for a multivariate extreme value distribution. *Biometrika 82*(3), 644–649.

Sibuya, M. (1960). Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics 11*, 195–210.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

Smith, R. L., J. A. Tawn, and H. K. Yuen (1990). Statistics of multivariate extremes. *International Statistical Review / Revue Internationale de Statistique 58*(1), 47–58.

Stărică, C. (1999). Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance 6*(5), 515–553.

Stephenson, A. and E. Gilleland (2005). Software for the analysis of extreme events: The current state and future directions. *Extremes 8*(3), 87–109.

Stephenson, A. G. (2002). evd: Extreme Value Distributions. *R News 2*(2), June.

Stephenson, A. G. (2003). Simulating multivariate extreme value analysis of logistic type. *Extremes 6*, 49–59.

Stute, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *The Annals of Probability 12*(2), 361–379.

Tawn, J. A. (1988). Extreme value theory: Models and estimation. *Biometrika 75*, 397–415.

Tawn, J. A. (1990). Modelling multivariate extreme vlaue distributions. *Biometrika 77*(2), 245–253.

Tiago de Oliveira, J. (1958). Extremal distributions. *Revista Faculdade de Ciencias de Lisboa 7*, 219–227.

Tiago de Oliveira, J. (1974). Regression in the nondifferentiable bivariate extreme models. *Journal of the American Statistical Association 69*, 816–818.

Tiago de Oliveira, J. (1980). Bivariate extremes: foundations and statistics. In *Multivariate analysis, V (Proc. Fifth Internat. Sympos., Univ. Pittsburgh, Pittsburgh, Pa., 1978)*, pp. 349–366. Amsterdam: North-Holland.

Tiago de Oliveira, J. (1989). Statistical decision for bivariate extremes. In *Extreme value theory (Oberwolfach, 1987)*, Volume 51 of *Lecture Notes in Statistics*, pp. 246–261. New York: Springer.

Toulemonde, G., A. Guillou, P. Naveau, M. Vrac, and F. Chevalier (2009). Autoregressive models for maxima and their applications to CH_4 and N_2O. *Environmetrics*.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 33*(3), 357–375.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Cambridge Series in Statistical and Probabilistic Mathematics. New York: Springer.

Weingessel, A. (2010). *quadprog: Functions to solve Quadratic Programming Problems.* T.U. Wien.

Yan, J. (2007). Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software 21*(4), 1–21.

Zhang, D., M. T. Wells, and L. Peng (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *Journal of Multivariate Analysis 99*(4), 577–588.