TECHNICAL REPORT
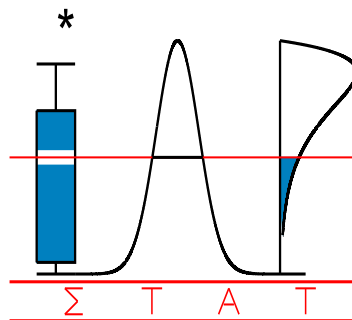
11020

Nonparametric estimation of multivariate extreme-value copulas

GUDENDORF , G. and J. SEGERS

IAP STATISTICS NETWORK

INTERUNIVERSITY ATTRACTION POLE

Nonparametric estimation of multivariate extreme-value copulas

Gordon Gudendorf · Johan Segers

July 12, 2011

Abstract Extreme-value copulas arise in the asymptotic theory for componentwise maxima of independent random samples. An extreme-value copula is determined by its Pickands dependence function, which is a function on the unit simplex subject to certain shape constraints that arise from an integral transform of an underlying measure called spectral measure. Multivariate extensions are provided of certain rank-based nonparametric estimators of the Pickands dependence function. The shape constraint that the estimator should itself be a Pickands dependence function is enforced by replacing an initial estimator by its best least-squares approximation in the set of Pickands dependence functions having a discrete spectral measure supported on a sufficiently fine grid. Weak convergence of the standardized estimators is demonstrated and the finite-sample performance of the estimators is investigated by means of a simulation experiment.

Keywords empirical copula · extreme-value copula · Pickands dependence function · simplex · shape constraints · spectral measure · weak convergence

Mathematics Subject Classification (2000) 62G05, 62G32, 62G20

1 Introduction

Extreme-value copulas arise in the asymptotic theory for componentwise maxima of independent random samples. They provide the dependence structures for the class of multivariate extreme-value or max-stable distributions. More

Gordon Gudendorf, Johan Segers

Université catholique de Louvain, Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. Funding was provided by IAP research network grant P6/03 of the Belgian government (Belgian Science Policy) and by "Projet d'actions de recherche concertées" number 07/12/002 of the Communauté française de Belgique, granted by the Académie universitaire de Louvain. E-mail: {gordon.gudendorf,johan.segers}@uclouvain.be

generally, they constitute a flexible class of models for describing positive association; see Gudendorf and Segers (2010) for a survey.

In this paper we will focus on the nonparametric estimation of extreme-value copulas in general dimensions. In particular, we aim at multivariate extensions of the rank-based estimators in Genest and Segers (2009) and the projection methodology in Fils-Villetard et al. (2008).

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,

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

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 $(F_1(X_1), \ldots, F_p(X_p))$. We are interested in nonparametric estimation of C in the model where the margins $F_1, \ldots, F_p$ are completely unknown (but continuous) and C is known to be an extreme-value copula.

A p-variate copula C is an *extreme-value copula* if there exists a finite Borel measure H on the unit simplex $\Delta_{p-1} = \{(w_1, \ldots, w_p) \in [0,1]^p : w_1 + \cdots + w_p = 1\}$, called *spectral measure*, such that

$$C(\boldsymbol{u}) = \exp\big(-\ell(-\log u_1, \ldots, -\log u_p)\big), \qquad \boldsymbol{u} \in (0,1]^p, \qquad (1.1)$$

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

$$\ell(\boldsymbol{x}) = \int_{\Delta_{p-1}} \max\{v_1 x_1, \ldots, v_p x_p\}\, H(\mathrm{d}\boldsymbol{v}). \qquad (1.2)$$

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

$$\int_{\Delta_{p-1}} v_j\, H(\mathrm{d}\boldsymbol{v}) = 1, \qquad j \in \{1, \ldots, p\}, \qquad (1.3)$$

which are equivalent to requiring that the margins of C be uniform on $(0,1)$.

The tail dependence function ℓ in (1.2) is convex, homogeneous of order one, that is $\ell(c\boldsymbol{x}) = c\,\ell(\boldsymbol{x})$ for $c > 0$, and satisfies $\max(x_1, \ldots, x_p) \leqslant \ell(\boldsymbol{x}) \leqslant x_1 + \cdots + x_p$ for all $\boldsymbol{x} \in [0,\infty)^p$. By homogeneity, ℓ 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: for $\boldsymbol{x} \in [0,\infty)^p \setminus \{\boldsymbol{0}\}$,

$$\ell(\boldsymbol{x}) = (x_1 + \cdots + x_p)\, A(w_1, \ldots, w_{p-1})$$
$$\text{where} \qquad w_j = \frac{x_j}{x_1 + \cdots + x_p}.$$

Here and further on, we frequently identify Δ_{p-1} with $\{(w_1, \ldots, w_{p-1}) \in [0,1]^{p-1} : w_1 + \cdots + w_{p-1} \leqslant 1\}$. The extreme-value copula C can be expressed in terms of A via

$$C(\boldsymbol{u}) = \exp\left\{ \Big(\textstyle\sum_{j=1}^p \log u_j\Big) A\left(\frac{\log u_1}{\sum_{j=1}^p \log u_j}, \ldots, \frac{\log u_{p-1}}{\sum_{j=1}^p \log u_j} \right) \right\} \qquad (1.4)$$

for $\boldsymbol{u} \in (0,1]^p \setminus \{(1,\ldots,1)\}$, with

$$A(\boldsymbol{w}) = \int_{\Delta_{p-1}} \max\{v_1 w_1, \ldots, v_p w_p\} H(\mathrm{d}\boldsymbol{v}), \qquad \boldsymbol{w} \in \Delta_{p-1}, \qquad (1.5)$$

see Pickands (1981) and Zhang et al. (2008). The function A is convex as well and satisfies $\max(w_1,\ldots,w_p) \leqslant A(\boldsymbol{w}) \leqslant 1$ for all $\boldsymbol{w} \in \Delta_{p-1}$.

Popular nonparametric estimators for A have initially been developed in Pickands (1981), with modifications in Deheuvels (1991) and Hall and Tajvidi (2000), and in Capéraà et al. (1997). These estimators will be referred to as the Pickands and CFG estimators, respectively; see Section 3 for definitions. In the previously cited papers, the marginal distributions were assumed to be known. The more realistic case of unknown margins has been treated in the bivariate case in Jiménez et al. (2001) for a submodel and in Genest and Segers (2009) for the general model. Multivariate extensions have been proposed in Zhang et al. (2008) and Gudendorf and Segers (2011) for the case of known margins. In Section 3, we will provide a proof for the convergence 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. As in Kojadinovic and Yan (2010) and Genest et al. (2011), the estimators could also be used as a starting point for goodness-of-fit tests, but for brevity we do not pursue this here. Finally, a new type of nonparametric estimator has been proposed in Bücher et al. (2011) for the bivariate case.

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 (2011), it was shown that the same expansion continues to hold under much weaker conditions on the partial derivatives. In Section 2, these issues are considered for multivariate extreme-value copulas.

As the estimators for A considered here fail to be Pickands dependence functions themselves, it is natural to ask how to enforce the shape constraints on such functions in the estimation procedure. In dimension $p = 2$, it is sufficient to ensure that A is convex and takes values in the range $\max(\boldsymbol{w}) \leqslant A(\boldsymbol{w}) \leqslant 1$, for instance by truncation and convexification (Deheuvels, 1991). In dimension $p \geqslant 3$, however, this procedure is no longer sufficient (Beirlant et al., 2004, page 257) and one needs to rely on the spectral representation in (1.5). In Section 4 we will apply an projection methodology (Fils-Villetard et al., 2008) to obtain valid estimates: an initial estimate is replaced by its best least-squares approximation in the set of Pickands dependence functions corresponding to discrete spectral measures supported on a fine grid.

The results of a simulation experiment aimed at investigating the finite-sample performance of the original and projected Pickands and CFG estimators are reported in Section 5. All proofs are relegated to the Appendix.

Throughout the article, we will apply the following notations. For a space $\mathscr{W}$, let $\ell^\infty(\mathscr{W})$ and $\mathscr{C}(\mathscr{W})$ denote the spaces of real-valued bounded and real-valued continuous functions respectively, where we endow both spaces with the uniform norm $\| \cdot \|_\infty : f \mapsto \sup_{x \in \mathscr{W}} |f(x)|$. Furthermore, we denote by $\mathbf{1}(E)$ the indicator function evaluated in a set E. The arrow '$\rightsquigarrow$' will stand for weak convergence. For any $p-$variate real-valued function f with values in $\mathbb{R}$, first and second-order partial derivatives will be denoted by $\dot{f}_i(\boldsymbol{x}) = \frac{\partial}{\partial x_i} f(x_1, \ldots, x_p)$ and $\dot{f}_{ij}(\boldsymbol{x}) = \frac{\partial^2}{\partial x_i \partial x_j} f(x_1, \ldots, x_p)$.

2 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

$$C_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\left(U_{i,1} \leqslant u_1, \ldots, U_{i,p} \leqslant u_p \right), \qquad \boldsymbol{u} \in [0,1]^p, \tag{2.1}$$

with associated empirical process

$$\alpha_n = n^{1/2} \left(C_n - C \right). \tag{2.2}$$

For ease of notation, we will write

$$\alpha_{n,j}(u_j) = \alpha_n(1, \ldots, 1, u_j, 1, \ldots, 1) \quad \text{for } j \in \{1, \ldots, p\}. \tag{2.3}$$

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

$$F_{n,j}(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{1}(X_{i,j} \leqslant x), \qquad j \in \{1, \ldots, p\}, \tag{2.4}$$

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

$$\hat{U}_{i,j} = F_{n,j}(X_{i,j}) = \frac{1}{n+1} \sum_{l=1}^n \mathbf{1}(X_{l,j} \leqslant X_{i,j}) \tag{2.5}$$

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

$$\hat{C}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\left(\hat{U}_{i,1} \leqslant u_1, \ldots, \hat{U}_{i,p} \leqslant u_p \right), \qquad \boldsymbol{u} \in [0,1]^p \tag{2.6}$$

with associated empirical copula process

$$\mathbb{C}_n = n^{1/2}\big(\hat{C}_n - C\big). \tag{2.7}$$

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 (2.2) and $\mathbb{C}_n$ in (2.7) are related via

$$\mathbb{C}_n(\boldsymbol{u}) = \alpha_n(\boldsymbol{u}) - \sum_{j=1}^{p} \dot{C}_j(\boldsymbol{u})\,\alpha_{n,j}(u_j) + R_n(\boldsymbol{u}), \tag{2.8}$$

the remainder term R_n satisfying

$$\sup_{\boldsymbol{u}\in[0,1]^p} |R_n(\boldsymbol{u})| = O\big(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\big) \quad \text{almost surely.} \tag{2.9}$$

Let $\ell^\infty([0,1]^p)$ be the space of bounded real-valued functions on $\mathbb{R}$, equipped with the topology of uniform convergence. Weak convergence of maps taking values in $\ell^\infty([0,1]^p)$ will be understood as in van der Vaart and Wellner (1996, Definition 1.3.3) and will be denoted by '$\rightsquigarrow$'. 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

$$\mathrm{cov}\big(\alpha(\boldsymbol{u}), \alpha(\boldsymbol{v})\big) = C(\boldsymbol{u}\wedge\boldsymbol{v}) - C(\boldsymbol{u})\,C(\boldsymbol{v}), \qquad \boldsymbol{u},\boldsymbol{v}\in[0,1]^p, \tag{2.10}$$

where $(\boldsymbol{u}\wedge\boldsymbol{v})_j = \min(u_j, v_j)$. In view of the expansion (2.8), it then follows that in $\ell^\infty([0,1]^p)$, we have $\mathbb{C}_n \rightsquigarrow \mathbb{C}$ as $n \to \infty$, where

$$\mathbb{C}(u) = \alpha(\boldsymbol{u}) - \sum_{j=1}^{p} \dot{C}_j(\boldsymbol{u})\,\alpha_j(u_j) \tag{2.11}$$

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 (2011, 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 1 below, the expansion (2.8)–(2.9) of the empirical copula process remains valid.

Assumption 1 *(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), $\ddot{C}_{ij}$ exists and is continuous on $V_{p,i} \cap V_{p,j}$ and

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

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 1 to hold.

Assumption 2 *(L1) For every $j \in \{1, \ldots, p\}$, the first-order partial derivative $\dot{\ell}_j$ exists and is continuous on the set $W_{p,j} = \{x \in [0,\infty)^p : x_j > 0\}$.*
(L2) For every $i, j \in \{1, \ldots, p\}$ (i and j not necessarily distinct), $\ddot{\ell}_{ij}$ exists and is continuous on $W_{p,i} \cap W_{p,j}$ and

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

Proposition 1 *For p-variate extreme-value copulas, (L1) implies (C1). If in addition (L2) holds, then (C2) holds as well.*

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, 2011, Example 5.3).

For completeness, we want to mention that in the above references, the empirical copula is not defined as in (2.6) but rather as

$$\hat{C}_n^D(\boldsymbol{u}) = F_n\big(F_{n,1}^\leftarrow(u_1), \ldots, F_{n,p}^\leftarrow(u_p)\big),$$

with $F_{n,j}^\leftarrow(u_j) = \inf\{x_j \in \mathbb{R} : F_{n,j}(x_j) \geqslant u_j\}$. Straightforward calculus shows that, in the absence of ties,

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

As a consequence, Stute's expansion (2.8) is valid for $\hat{C}_n$ if and only if it is valid for $\hat{C}_n^D$.

3 Nonparametric estimation of the dependence function

Among the most popular nonparametric estimators for A figure the Pickands estimator $\hat{A}_n^{\mathrm{P}}$ (Pickands, 1981) and the estimator $\hat{A}_n^{\mathrm{CFG}}$ proposed by Capéraà et al. (1997), referred to as the CFG estimator from now on. Writing

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

for $\boldsymbol{w} \in \Delta_{p-1}$, with $\hat{\boldsymbol{U}}_{i,j}$ as in (2.5), these estimators are defined as

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

with $\gamma = 0.5772\ldots$ the Euler–Mascheroni constant. Explanations on the construction of these estimators are provided for instance in the original references given before, in Genest and Segers (2009) and in the survey paper Gudendorf and Segers (2010). The multivariate extension of the CFG estimator was presented in Zhang et al. (2008), albeit under a different but equivalent form.

In order to improve the small-sample properties of the above estimators, the endpoint constraints $A(\boldsymbol{e}_j) = 1$ for $j \in \{1, \ldots, p\}$ can be imposed as follows. 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 $j, k \in \{1, \ldots, p\}$, define

$$\frac{1}{\hat{A}^{\mathrm{P}}_{\boldsymbol{\lambda},n}(\boldsymbol{w})} = \frac{1}{\hat{A}^{\mathrm{P}}_n(\boldsymbol{w})} - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \left(\frac{1}{\hat{A}^{\mathrm{P}}_n(\boldsymbol{e}_j)} - 1 \right), \qquad (3.1)$$

$$\log \hat{A}^{\mathrm{CFG}}_{\boldsymbol{\lambda},n}(\boldsymbol{w}) = \log \hat{A}^{\mathrm{CFG}}_n(\boldsymbol{w}) - \sum_{j=1}^{p} \lambda_j(\boldsymbol{w}) \log \hat{A}^{\mathrm{CFG}}_n(\boldsymbol{e}_j). \qquad (3.2)$$

In case of known margins, variance-minimizing weight functions λ_j can be determined adaptively by ordinary least squares (Segers, 2007; Gudendorf and Segers, 2011). However, if the marginal distributions are unknown, these endpoint corrections are asymptotically irrelevant (Genest and Segers, 2009), since

$$\frac{1}{\hat{A}^{\mathrm{P}}_n(\boldsymbol{e}_j)} - 1 = \frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{n+1}{i} \right) - 1 = O\left(\frac{\log n}{n} \right),$$

$$\log \hat{A}^{\mathrm{CFG}}_n(\boldsymbol{e}_j) = -\frac{1}{n} \sum_{i=1}^{n} \log \left(\log \left(\frac{n+1}{i} \right) \right) + \int_0^1 \log \left(\log \left(\frac{1}{x} \right) \right) \mathrm{d}x$$

$$= O\left(\frac{(\log n)^2}{n} \right).$$

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 $\hat{A}^{\mathrm{HT}}_n(\boldsymbol{w}) = \hat{A}^{\mathrm{P}}_n(\boldsymbol{w})/\hat{A}^{\mathrm{P}}_n(\boldsymbol{e}_j) = \hat{A}^{\mathrm{P}}_n(\boldsymbol{w})\{1 + O(n^{-1}\log n)\}$.

The next lemma establishes a functional relationship between $\hat{A}^{\mathrm{P}}_n$ and $\hat{A}^{\mathrm{CFG}}_n$ on the one hand and the empirical copula $\hat{C}_n$ on the other hand. Recall the empirical copula process $\mathbb{C}_n$ in (2.7).

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

$$n^{1/2} \left(\frac{1}{\hat{A}^{\mathrm{P}}_n(\boldsymbol{w})} - \frac{1}{A(\boldsymbol{w})} \right) = \int_0^1 \mathbb{C}_n(u^{w_1}, \ldots, u^{w_p}) \frac{\mathrm{d}u}{u}, \qquad (3.3)$$

$$n^{1/2} \left(\log \hat{A}^{\mathrm{CFG}}_n(\boldsymbol{w}) - \log A(\boldsymbol{w}) \right) = \int_0^1 \mathbb{C}_n(u^{w_1}, \ldots, u^{w_p}) \frac{\mathrm{d}u}{u \log u}. \qquad (3.4)$$

The proof is not different from the one in dimension two and can be found in Genest and Segers (2009, Lemma 3.1). Equations (3.3) and (3.4) 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 1 *Under Assumption 1 above, $\mathbb{A}_n^{\mathrm{P}} \rightsquigarrow \mathbb{A}^{\mathrm{P}}$ and $\mathbb{A}_n^{\mathrm{CFG}} \rightsquigarrow \mathbb{A}^{\mathrm{CFG}}$ with*

$$\mathbb{A}^{\mathrm{P}}(\boldsymbol{w}) = -A^2(\boldsymbol{w}) \int_0^1 \mathbb{C}(u^{w_1}, \ldots, u^{w_p}) \, \frac{\mathrm{d}u}{u}$$

$$\mathbb{A}^{\mathrm{CFG}}(\boldsymbol{w}) = A(\boldsymbol{w}) \int_0^1 \mathbb{C}(u^{w_1}, \ldots, u^{w_p}) \frac{\mathrm{d}u}{u \, \log u},$$

as $n \to \infty$ in the space $\mathscr{C}(\Delta_{p-1})$ equipped with the topology of uniform convergence.

The main idea of the proof consists in substituting $\mathbb{C}_n$ in (3.3) and (3.4) by Stute's expansion and to conclude using a refined version of the continuous mapping theorem. As the proof follows the same lines as the one in Genest and Segers (2009), we will just point out the main adjustments.

4 Projection estimator

The estimators of the Pickands dependence functions considered so far are in general not valid Pickands dependence functions themselves. In this section, we adapt the methodology in Fils-Villetard et al. (2008) to project a pilot estimate $\hat{A}_n$ onto the set $\mathscr{A}$ of Pickands dependence functions of p-variate extreme-value copulas. To this end, 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

$$\hat{A}^{\mathrm{pr}} = \Pi(\hat{A}_n | \mathscr{A}) = \arg\min_{A \in \mathscr{A}} \|A - \hat{A}_n\|_2.$$

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.

For practical computations, we are obliged to refer to finite-dimensional subclasses $\mathscr{A}_m \subset \mathscr{A}$, yielding the approximate projection estimator

$$\hat{A}_m^{\mathrm{pr}} = \Pi(\hat{A}_n | \mathscr{A}_m) = \arg\min_{A \in \mathscr{A}_m} \|A - \hat{A}_n\|_2. \tag{4.1}$$

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 $\boldsymbol{v} = (v_1, \ldots, v_p) \in \Delta_{p-1}$ such that $k_j = mv_j$ is integer for every $j \in \{1, \ldots, p\}$, so that in fact $\boldsymbol{v} = (k_1/m, \ldots, k_p/m)$ where $k_j \in \{0, \ldots, m\}$ and $k_1 + \cdots + k_p = m$. 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

$$
\begin{cases}
h_{\boldsymbol{v}} \geqslant 0, & \forall \boldsymbol{v} \in \mathscr{V}_{p,m}, \\
\sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} h_{\boldsymbol{v}} \, v_j = 1, & \forall j \in \{1, \ldots, p\},
\end{cases}
\tag{4.2}
$$

the second constraint stemming from (1.3).

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

$$
A_{\boldsymbol{h}}(\boldsymbol{w}) = \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} h_{\boldsymbol{v}} \max\{w_1 v_1, \ldots, w_p v_p\}, \qquad \boldsymbol{w} \in \Delta_{p-1}. \tag{4.3}
$$

Being a linear combination of piecewise linear functions, the function A in (4.3) is itself piecewise linear. All Pickands dependence function of the form (4.3) 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 2 *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*

$$
\sup_{\boldsymbol{w} \in \Delta_{p-1}} |A_m(\boldsymbol{w}) - A(\boldsymbol{w})| \leqslant \frac{p^2}{m}. \tag{4.4}
$$

The bound in (4.4) 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 (4.1). The vector $\hat{\boldsymbol{h}}$ is given as the solution to the least-squares problem

$$
\hat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \|A_{\boldsymbol{h}} - \hat{A}_n\|_2^2 = \arg\min_{\boldsymbol{h}} \left(\|A_{\boldsymbol{h}}\|_2^2 - 2\langle A_{\boldsymbol{h}}, \hat{A}_n \rangle \right), \tag{4.5}
$$

with $\boldsymbol{h}$ subject to the constraints (4.2). The optimisation problem in (4.5) is a quadratic program with linear constraints, which in matrix notation reads

$$\hat{\boldsymbol{h}} = \underset{\boldsymbol{h}}{\arg\min}\left(\frac{1}{2}\boldsymbol{h}^\top \boldsymbol{D}\boldsymbol{h} - \boldsymbol{d}^\top \boldsymbol{h}\right), \quad \text{subject to} \quad \begin{cases} \boldsymbol{C}\boldsymbol{h} = \boldsymbol{c}, \\ \boldsymbol{h} \geqslant \boldsymbol{0}. \end{cases} \qquad (4.6)$$

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 (Turlach and 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 2 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 $\|\hat{A}_m^{\mathrm{pr}} - \hat{A}^{\mathrm{pr}}\|_2 = o_{\mathrm{P}}(\epsilon_n)$. From Fils-Villetard et al. (2008, Theorem 1), we deduce that

$$\varepsilon_n^{-1}(\hat{A}_m^{\mathrm{pr}} - A) = \varepsilon_n^{-1}(\hat{A}^{\mathrm{pr}} - A) + o_{\mathrm{P}}(1) \rightsquigarrow \Pi\big(\zeta|\mathrm{T}_{\mathscr{A}}(A)\big) \qquad (n \to \infty) \quad (4.7)$$

in the space $L^2(\Delta_{p-1})$, where $\mathrm{T}_{\mathscr{A}}(A)$ is the tangent cone of $\mathscr{A}$ at A, defined as the L^2-closure of $\{\lambda(\tilde{A} - A) : \lambda \geqslant 0, \tilde{A} \in \mathscr{A}\}$.

Interestingly, equation (4.7) 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 $\hat{A}_m^{\mathrm{pr}}$ 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.

Finally, note that the convergence in (4.7) 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 Simulation study

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

PD – the endpoint-corrected Pickands estimator in (3.1) with $\lambda_j(\boldsymbol{w}) = w_j$, in the spirit of Deheuvels (1991);

PD-pr – the projection estimator in (4.1) with the previous end-point corrected Pickands estimator as initial estimator;

CFG – the endpoint-corrected CFG estimator in (3.2) with $\lambda_j(\boldsymbol{w}) = w_j$;

CFG-pr – the projection estimator in (4.1) 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 (2011), random samples were generated from a trivariate extreme-value distribution with asymmetric logistic dependence function A (Tawn, 1990):

$$
\begin{aligned}
A(\boldsymbol{w}) = {} & (\theta^{1/\alpha} w_1^{1/\alpha} + \phi^{1/\alpha} w_2^{1/\alpha})^\alpha + (\theta^{1/\alpha} w_2^{1/\alpha} + \phi^{1/\alpha} w_3^{1/\alpha})^\alpha \\
& + (\theta^{1/\alpha} w_3^{1/\alpha} + \phi^{1/\alpha} w_1^{1/\alpha})^\alpha + \psi(w_1^{1/\alpha} + w_2^{1/\alpha} + w_3^{1/\alpha})^\alpha \\
& \qquad\qquad\qquad + 1 - \theta - \phi - \psi, \qquad \boldsymbol{w} \in \Delta_2, \quad (5.1)
\end{aligned}
$$

with parameter vector $(\alpha, \theta, \phi, \psi) \in (0,1] \times [0,1]^3$. For this model, Assumption 2 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 $1/n$.

		$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$	$\alpha = 1$
$n = 50$	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 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 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

References

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.

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., A.-L. Fougères, and C. Genest (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika 84*, 567–577.

Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters 12*(5), 429–439.

Fils-Villetard, A., A. Guillou, and J. Segers (2008). Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics 36*(3), 369–382.

Genest, C., I. Kojadinovic, J. Nešlehová, and J. Yan (2011). A goodness-of-fit test for extreme-value copulas. *Bernoulli 17*, 253–275.

Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *Annals of Statistics 37*(5B), 2990–3022.

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 (2011). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *J. Multivariate Analysis 102*(1), 37–47.

Hall, P. and N. Tajvidi (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli 6*(5), 835–844.

Jiménez, J. R., E. Villa-Diharce, and M. Flores (2001). Nonparametric estimation of the dependence function in bivariate extreme value distributions. *Journal of Multivariate Analysis 76*(2), 159–191.

Kojadinovic, I. and J. Yan (2010). Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis 101*(9), 2234 – 2249.

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.

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. (2011). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli (to appear)*. arXiv:1012.2133v2.

Stephenson, A. G. (2002). evd: Extreme Value Distributions. *R News 2*(2), June.

Stephenson, A. G. (2003). Simulating multivariate extrme 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. (1990). Modelling multivariate extreme value distributions. *Biometrika 77*(2), 245–253.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 33*(3), 357–375.

Turlach, B. A. and A. Weingessel (2010). *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.5-3. S original by Berwin A. Turlach, R port by Andreas Weingessel, http://CRAN.R-project.org/package=quadprog.

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.

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.

A Proofs

A.1 Proof of Proposition 1

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

$$\dot{C}_j(\boldsymbol{u}) = \frac{C(\boldsymbol{u})}{u_j}\, \dot{\ell}_j(-\log \boldsymbol{u}). \tag{A.1}$$

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 $i \in \{1,\ldots,p\} \setminus \{j\}$, 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 $x_1 + \cdots + x_p = 1$. 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

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

and thus

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

It follows that

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

that is, the map $\boldsymbol{x} \mapsto \max(x_i, x_j)\,\ddot{\ell}_{ij}(\boldsymbol{x})$ 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

$$\ddot{C}_{ij}(\boldsymbol{u}) = \begin{cases} \dfrac{C(\boldsymbol{u})}{u_i u_j}\left(\dot{\ell}_i \dot{\ell}_j - \ddot{\ell}_{ij}\right) & \text{if } i \neq j, \\[3mm] \dfrac{C(\boldsymbol{u})}{u_j^2}\left(\dot{\ell}_j^2 - \dot{\ell}_j - \ddot{\ell}_{jj}\right) & \text{if } i = j, \end{cases}$$

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 $\ddot{C}_{ij}(\boldsymbol{u}) = 0$.

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

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

Continuity of $\ddot{C}_{ij}$ 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,

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

which is equivalent to condition (C2).

A.2 Proof of Theorem 1

The proof of theorem 1 will require the following preliminary result on weighted empirical copula processes. Recall the process α_n in (2.2). 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

$$
\mathbb{G}_{n,\theta}(\boldsymbol{u}) = \begin{cases} \frac{\alpha_n(\boldsymbol{u})}{q_\theta(\min(\boldsymbol{u}))} & \text{if } \boldsymbol{u} \in \mathbb{E}, \\ 0 & \text{if } \boldsymbol{u} \in [0,1]^p \setminus \mathbb{E}. \end{cases} \tag{A.2}
$$

Similarly, define the process $\mathbb{G}_\theta$ on $[0,1]^p$ by replacing α_n in (A.2) by its weak limit α, see (2.10). The following result generalizes Theorem G.1 in Genest and Segers (2009).

Lemma 3 *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 (Lemma 3) The proof is entirely analogue as the one of Theorem G.1 in Genest and Segers (2009). For completeness, we sketch the main lines.

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

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

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

$$
\rho^2(f,g) = \mathrm{P}(f-g)^2 \qquad f, g \in \mathscr{F}. \tag{A.3}
$$

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

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

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

We will show that the collection $\mathscr{F}$ is a P-Donsker class, i.e. there exists a P-Brownian bridge $\mathbb{G}$ such that

$$
\mathbb{G}_n \rightsquigarrow \mathbb{G} \quad \text{in} \quad \ell^\infty(\mathscr{F}) \text{ as } n \to \infty.
$$

It is sufficient to verify the conditions of Theorem 2.6.14 of van der Vaart and Wellner (1996). The function F on $\mathbb{E}$ defined by

$$
F(s_1, \ldots, s_p) = p \max\left\{ s_1^{-\theta}, \ldots, s_p^{-\theta}, (1-s_1)^{-\theta}, \ldots, (1-s_p)^{-\theta} \right\}
$$

is a suitable envelope function for $\mathscr{F}$. The fact that $\mathscr{F}$ is a VC-major class and is pointwise separable follows from the same arguments as in Genest and Segers (2009).

For the moment $\mathbb{G}_n$ is defined on $\mathscr{F}$ with the metric ρ in (A.3). Consider the map $\phi : [0,1]^p \to \mathscr{F}$ defined by $\phi(\boldsymbol{u}) = f_{\boldsymbol{u}}$ if $\boldsymbol{u} \in \mathbb{E}$ and $\phi(\boldsymbol{u}) = 0$ if $\boldsymbol{u} \in [0,1]^p \setminus \mathbb{E}$. Then $\mathbb{G}_{n,\theta} = \mathbb{G}_n \circ \phi$ and $\mathbb{G}_{\theta} = \mathbb{G} \circ \phi$. The map $\ell^{\infty}(\mathscr{F}) \to \ell^{\infty}([0,1]^p) : z \mapsto z \circ \phi$ being continuous, the continuous mapping theorem permits to conclude that $\mathbb{G}_{n,\theta} \rightsquigarrow \mathbb{G}_{\theta}$ in $\ell^{\infty}([0,1]^p)$. Since the trajectories of $\mathbb{G}$ are ρ-continuous almost surely and since ϕ is continuous, it follows that the sample paths of $\mathbb{G}_{\theta}$ are continuous almost surely as well. This concludes the proof of Lemma 3.

We now proceed with the proof of Theorem 1. Define

$$B_n^{\mathrm{P}}(\boldsymbol{w}) = n^{1/2}\left(\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} - \frac{1}{A(\boldsymbol{w})}\right),$$

$$B_n^{\mathrm{CFG}}(\boldsymbol{w}) = n^{1/2}\left(\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - \log A(\boldsymbol{w})\right),$$

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

$$B_n(\boldsymbol{w}) = \int_0^{\infty} \mathbb{C}_n(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\, \mathrm{d}s, \tag{A.4}$$

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 (A.4) into three parts:

$$B_n(\boldsymbol{w}) = \int_0^{l_n} + \int_{l_n}^{k_n} + \int_{k_n}^{\infty} = I_{1,n}(\boldsymbol{w}) + I_{2,n}(\boldsymbol{w}) + I_{3,n}(\boldsymbol{w}). \tag{A.5}$$

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 $e^{-s} \geqslant 1 - l_n$ and thus $C_n(e^{-w_1 s}, \ldots, e^{-w_p s}) = 1$, which implies

$$0 \leqslant I_{1,n}(\boldsymbol{w}) = \int_0^{l_n} n^{1/2}\left(1 - e^{-s A(\boldsymbol{w})}\right) h(s)\, \mathrm{d}s$$

$$\leqslant n^{1/2} \int_0^{l_n} s\, h(s)\, \mathrm{d}s \leqslant n^{1/2} l_n \leqslant n^{-1/2}.$$

– 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

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

$$\leqslant \frac{n^{1/2}}{A(\boldsymbol{w})} e^{-k_n A(\boldsymbol{w})} \leqslant p\, n^{-1/2},$$

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 (A.5) is $I_{2,n}$. By Assumption 1 and by Proposition 4.2 in Segers (2011), Stute's expansion (2.8)–(2.9) 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

$$J_{0,n}(\boldsymbol{w}) = \int_{l_n}^{k_n} \alpha_n(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\, \mathrm{d}s,$$

$$J_{j,n}(\boldsymbol{w}) = \int_{l_n}^{k_n} \alpha_{n,j}(e^{-w_j s})\, \dot{C}_j(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\, \mathrm{d}s, \qquad j \in \{1, \ldots, p\},$$

$$J_{p+1,n}(\boldsymbol{w}) = \int_{l_n}^{k_n} R_n(e^{-w_1 s}, \ldots, e^{-w_p s})\, h(s)\, \mathrm{d}s.$$

In view of the bound (2.9) on R_n, the term $J_{p+1,n}$ is negligible: as $n \to \infty$,

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

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

$$J_{0,n}(\boldsymbol{w}) = \int_{l_n}^{k_n} \mathbb{G}_{n,\theta}(e^{-sw_1}, \ldots, e^{-sw_p})\, K_0(s, \boldsymbol{w})\, h(s)\, \mathrm{d}s,$$

$$J_{j,n}(\boldsymbol{w}) = \int_{l_n}^{k_n} \mathbb{G}_{n,\theta}(1, \ldots, 1, e^{-sw_j}, 1, \ldots, 1)\, K_j(s, \boldsymbol{w})\, h(s)\, \mathrm{d}s, \qquad j \in \{1, \ldots, p\},$$

with

$$K_0(s, \boldsymbol{w}) = q_\theta\big(\min(e^{-sw_1}, \ldots, e^{-sw_p})\big),$$

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

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

$$0 \leqslant K_j(s, \boldsymbol{w}) \leqslant K(s) = s^\theta\, \mathbf{1}_{(0,1]}(s) + e^{-(\theta/p)s}\, \mathbf{1}_{(1,\infty)}(s),$$

$$j \in \{0, \ldots, p\}, \qquad s \in (0, \infty), \qquad \boldsymbol{w} \in \Delta_{p-1}. \quad \text{(A.6)}$$

To prove these bounds, use equation (A.1), 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 $\boldsymbol{w} \in \Delta_{p-1}$. The function K in (A.6) satisfies $\int_0^\infty K(s)\, h(s)\, \mathrm{d}s < \infty$.

By Lemma 3 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

$$J_0(\boldsymbol{w}) = \int_0^\infty \mathbb{G}_\theta(e^{-sw_1}, \ldots, e^{-sw_p})\, K_0(s, \boldsymbol{w})\, h(s)\, \mathrm{d}s,$$

$$J_j(\boldsymbol{w}) = \int_0^\infty \mathbb{G}_\theta(1, \ldots, 1, e^{-sw_j}, 1, \ldots, 1)\, K_j(s, \boldsymbol{w})\, h(s)\, \mathrm{d}s, \qquad j \in \{1, \ldots, p\}.$$

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

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

An application of the functional delta method (van der Vaart and Wellner, 1996, Theorem 3.9.4) now yields the result.

A.3 Proof of Lemma 2

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 (4.4).

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 $j \in \{1,\ldots,p-1\}$; 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 $j \in \{1,\ldots,p-1\}$ 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 $j \in \{1,\ldots,p\}$ 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 $j \in \{1,\ldots,p-1\}$ 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

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

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

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

(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 $e_1,\ldots,e_{p-1}$, where e_j is the jth coordinate vector in $\mathbb{R}^p$. Specifically, we set

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

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

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

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

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

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

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 (4.4). 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

$$A_m(\boldsymbol{w}) = \int_{\Delta_{p-1}} f(\boldsymbol{w}, \boldsymbol{t}) \, H_m(\mathrm{d}\boldsymbol{t})$$

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

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,

$$A_m(\boldsymbol{w}) \leqslant B_m(\boldsymbol{w}) + a_1 w_1 + \cdots + a_{p-1} w_{p-1}$$

$$\leqslant B_m(\boldsymbol{w}) + \max(a_1, \ldots, a_{p-1}) < B_m(\boldsymbol{w}) + \frac{p}{m}$$

and, on the other hand,

$$A_m(\boldsymbol{w}) \geqslant (1 - a_0) B_m(\boldsymbol{w}) \geqslant B_m(\boldsymbol{w}) - \frac{a_0}{1 - a_0} = B_m(\boldsymbol{w}) - \frac{1}{m} \sum_{i=1}^{p-1} c_i$$

$$> B_m(\boldsymbol{w}) - \frac{p}{m}.$$

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,

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

$$= \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \left| \int_{\Delta_{p,\boldsymbol{v},m}} f(\boldsymbol{w}, \boldsymbol{t}) \, H(\mathrm{d}\boldsymbol{t}) - f(\boldsymbol{w}, \boldsymbol{v}) \, H(\Delta_{p,\boldsymbol{v},m}) \right|$$

$$= \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \left| \int_{\Delta_{p,\boldsymbol{v},m}} \big(f(\boldsymbol{w}, \boldsymbol{t}) - f(\boldsymbol{w}, \boldsymbol{v})\big) \, H(\mathrm{d}\boldsymbol{t}) \right|$$

$$\leqslant \sum_{\boldsymbol{v} \in \mathscr{V}_{p,m}} \int_{\Delta_{p,\boldsymbol{v},m}} \big| f(\boldsymbol{w}, \boldsymbol{t}) - f(\boldsymbol{w}, \boldsymbol{v}) \big| \, H(\mathrm{d}\boldsymbol{t}).$$

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 (4.4) follows.