STRONG CONVERGENCE OF MULTIVARIATE MAXIMA

MICHAEL FALK[(1)], SIMONE A. PADOAN[(2)], AND STEFANO RIZZELLI[(3)]

ABSTRACT. It is well known and readily seen that the maximum of n independent and uniformly on $[0,1]$ distributed random variables, suitably standardised, converges in total variation distance, as n increases, to the standard negative exponential distribution. We extend this result to higher dimensions by considering copulas. We show that the strong convergence result holds for copulas that are in a differential neighbourhood of a multivariate generalized Pareto copula. Sklar's theorem then implies convergence in variational distance of the maximum of n independent and identically distributed random vectors with arbitrary common distribution function and (under conditions on the marginals) of its appropriately normalised version. We illustrate how these convergence results can be exploited to establish the almost-sure consistency of some estimation procedures for max-stable models, using sample maxima.

(1) INSTITUTE OF MATHEMATICS, UNIVERSITY OF WÜRZBURG, WÜRZBURG, GERMANY

(2) DEPARTMENT OF DECISION SCIENCES, BOCCONI UNIVERSITY OF MILAN, MILANO, ITALY

(3) INSTITUTE OF MATHEMATICS, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, LAUSANNE, SWITZERLAND

E-mail addresses: michael.falk@uni-wuerzburg.de, simone.padoan@unibocconi.it, stefano.rizzelli@epfl.ch.

Date: March 12, 2020.

2010 *Mathematics Subject Classification.* Primary 60G70; secondary 62H10, 60F15.

Key words and phrases. Maxima; strong convergence; total variation; copula; generalized Pareto copula; D-norm; multivariate max-stable distribution; domain of attraction.

1. Introduction

Let U be a random variable (rv), which follows the uniform distribution on $[0,1]$, i.e.,

$$P(U \le u) = \begin{cases} 0, & u < 0 \\ u, & u \in [0,1] \\ 1, & u > 1 \end{cases} =: V(u). \tag{1}$$

Let $U^{(1)}, U^{(2)}, \dots$ be independent and identically distributed (iid) copies of U. Then, clearly, we have for $x \le 0$ and large $n \in \mathbb{N}$ (natural set),

$$\begin{aligned} P\left(n\left(\max_{1\le i\le n} U^{(i)} - 1\right) \le x\right) &= P\left(U_i \le 1 + \frac{x}{n},\ 1 \le i \le n\right) \\ &= V^n\left(1 + \frac{x}{n}\right) \\ &= \left(1 + \frac{x}{n}\right)^n \\ &\to_{n\to\infty} G(x), \end{aligned} \tag{2}$$

where

$$G(x) = \begin{cases} \exp(x), & x \le 0 \\ 1, & x > 0 \end{cases} \tag{3}$$

is the distribution function (df) of the standard negative exponential distribution. Thus, we have established convergence in distribution of the suitably normalised sample maixmum, i.e.

$$n\left(M^{(n)} - 1\right) \to_D \eta, \tag{4}$$

where $M^{(n)} := \max_{1\le i\le n} U^{(i)}$, $n \in \mathbb{N}$, the arrow "$\to_D$" denotes convergence in distribution, and the rv η has df G in (3).

Note that, with $v(x) := V'(x) = 1$, if $x \in [0,1]$, and zero elsewhere, we have

$$\begin{aligned} v_n(x) &:= \frac{\partial}{\partial x}\left(V^n\left(1+\frac{x}{n}\right)\right) = V^{n-1}\left(1+\frac{x}{n}\right) v\left(1+\frac{x}{n}\right) \\ &\to_{n\to\infty} g(x) := G'(x) = \begin{cases} \exp(x), & x \le 0 \\ 0, & x > 0 \end{cases}, \end{aligned}$$

i.e., we have pointwise convergence of the sequence of densities of normalised maximum $n\left(M^{(n)}-1\right)$, $n \in \mathbb{N}$, to that of η. Scheffé's lemma, see, e.g. Reiss (1989, Lemma 3.3.3) now implies convergence in total variation:

$$\sup_{A\in\mathbb{B}} \left| P\left(n\left(M^{(n)}-1\right) \in A\right) - P(\eta \in A)\right| \to_{n\to\infty} 0, \tag{5}$$

where $\mathbb{B}$ denotes the Borel-σ-field in $\mathbb{R}$.

Let now X be a rv with *arbitrary* df F and $F^{-1}(q) := \{t \in \mathbb{R} : F(t) \ge q\}$ with $q \in (0,1)$ be the usual *quantile function* or *generalized inverse* of F. Then, we can assume the representation

$$X = F^{-1}(U).$$

Let $X^{(1)}, X^{(2)}, \dots$ be independent copies of X. Again we can consider the representation

$$X^{(i)} = F^{-1}\left(U^{(i)}\right), \qquad i = 1, 2, \dots$$

The fact that each quantile function is a nondecreasing function yields

$$\begin{aligned} \max_{1\le i\le n} X^{(i)} &= \max_{1\le i\le n} F^{-1}\left(U^{(i)}\right) = F^{-1}\left(\max_{1\le i\le n} U^{(i)}\right) \\ &= F^{-1}\left(1+\frac{1}{n}\left(n\left(\max_{1\le i\le n} U^{(i)}-1\right)\right)\right). \end{aligned}$$

The strong convergence in equation (5) now implies the following convergence in total variation:

$$\sup_{A\in\mathbb{B}} \left| P\left(\max_{1\le i\le n} X^{(i)} \in A\right) - P\left(F^{-1}\left(1+\frac{1}{n}\eta\right) \in A\right)\right| \to_{n\to\infty} 0. \tag{6}$$

Finally, assume that F is a continuous df with density $f = F'$. We denote the right endpoint of F by $x_0 := \sup\{x \in \mathbb{R} : F(x) < 1\}$. Assume also that $F \in \mathcal{D}(G^*_\gamma)$, i.e. F belongs to the domain of attraction of a generalised extreme-value df G^*_γ, e.g. Falk et al. (2019, p. 21). This means, for $n \in \mathbb{N}$, there are norming constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) \to_{n\to\infty} \exp\left(-(1+\gamma x)_+^{-1/\gamma}\right) =: G^*_\gamma(x), \tag{7}$$

for all $x \in \mathbb{R}$, where $(x)_+ = \max(0, x)$ and $\gamma \in \mathbb{R}$ is the so-called *tail index*. Such a coefficient describes the heaviness of the upper tail of the probability density function corresponding to G^*_γ, see Falk et al. (2019, for details). Furthermore, in this general case, we also have the pointwise convergence at the density level, i.e.

$$f^{(n)}(x) := \frac{\partial}{\partial x} F^n(a_n x + b_n) \to_{n\to\infty} \frac{\partial}{\partial x} G^*_\gamma(x) =: g^*_\gamma(x) \tag{8}$$

for all $x \in \mathbb{R}$, if and only if

$$\lim_{x\to\infty} \frac{x f(x)}{1-F(x)} = 1/\gamma, \qquad \text{if } \gamma > 0 \tag{9}$$

$$\lim_{x\uparrow x_0} \frac{(x_0 - x) f(x)}{1-F(x)} = -1/\gamma, \qquad \text{if } \gamma < 0 \tag{10}$$

$$\lim_{x\uparrow x_0} \frac{f(x)}{(1-F(x))^2} \int_0^{x_0} 1 - F(t)\mathrm{d}t = 1, \qquad \text{if } \gamma = 0, \tag{11}$$

see e.g. Proposition 2.5 in Resnick (2008). In particular, if (8) holds true, Scheffé's lemma entails that

$$\sup_{A\in\mathbb{B}} \left| P\left(a_n^{-1}\left(\max_{1\le i\le n} X^{(i)} - b_n\right) \in A\right) - P\left(Y \in A\right)\right| \to_{n\to\infty} 0, \tag{12}$$

where Y is a rv with distribution G^*_γ and $X^{(i)}$, $i = 1, \ldots, n$ are independent copies of a rv X with distribution F, with $F \in \mathcal{D}(G^*_\gamma)$.

In this paper we extend the results in (5), (6) and (12) to higher dimensions. First, in Section 2, we consider copulas. In Theorem 2.4, we demonstrate that the strong convergence result holds for copulas that are in a differential neighbourhood of a multivariate generalized Pareto copula (Falk et al., 2019; Falk, 2019). As a result of this, we also establish strong convergence of the copula of the maximum of n iid random vectors with arbitrary common df to the limiting extreme-value copula (Corollary 2.7). Sklar's theorem is then used in Section 3 to derive convergence in variational distance of the maximum of n iid random vectors with arbitrary common df and, under restrictions (9)-(11) on the margins, of its normalised versions. These results address some still open problems in the literature on multivariate extremes.

Strong convergence for extremal order statistics of univariate iid rv has been well investigated; see, e.g. Section 5.1 in Reiss (1989) and the literature cited therein. Strong convergence holds in particular under suitable von Mises type conditions on the underlying df, see (9)-(12) for the univariate normalised maximum. Much less is known in the multivariate setup. In this case, a possible approach is to investigate a point process of exceedances over high thresholds and establish its convergence to a Poisson process. This is done under suitable assumptions on variational convergence for truncated point measures, see e.g. Theorem 7.1.4 in Falk et al. (2011). It is proven in Kaufmann and Reiss (1993) that strong convergence of such multivariate point processes holds if, and only if, strong convergence of multivariate maxima occurs. Differently from that, we provide simple conditions (namely (17) and (24)) under which strong convergence of multivariate maxima and its normalised version actually holds. Furthermore, our strong convergence results for sample maxima are valid for maxima with arbitrary dimensions, unlike those in de Haan and Peng (1997), which are tailored to the two-dimensional case. Section 4 concludes the paper, by illustrating how effective our variational convergence results are for statistical purposes. In particular, when the interest is on inferential procedures for sample maxima whose df is in a neighborhood of some multivariate max-stable model, we show that, e.g., our results can be used to establish almost-sure consistency for the empirical copula estimator of the extreme-value copula. Similar results can also be achieved within the Bayesian inferential approach.

2. Strong Results for Copulas

Suppose that the random vector (rv) $\boldsymbol{U} = (U_1, \dots, U_d)$ follows a *copula*, say C, on $\mathbb{R}^d$, i.e., each component U_j has df V_j given in formula (1). Let $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \dots$ be independent copies of $\boldsymbol{U}$ and put for $n \in \mathbb{N}$

$$\boldsymbol{M}^{(n)} := \left(M_1^{(n)}, \dots, M_d^{(n)}\right) := \left(\max_{1\le i\le n} U_1^{(i)}, \dots, \max_{1\le i\le n} U_d^{(i)}\right). \tag{13}$$

In the sequel the operations involving vectors are meant componentwise, furthermore, we set $\mathbf{0} = (0, \dots, 0)$, $\mathbf{1} = (1 \dots, 1)$ and $\boldsymbol{\infty} = (\infty, \dots, \infty)$. Finally, hereafter, we denote the copula of the random vector in (13) by $C^n(\boldsymbol{u})$, $\boldsymbol{u} \in [0,1]^d$.

Suppose that a convergence result analogous to (2) holds for the random *vector* $\boldsymbol{M}^{(n)}$ of componentwise maxima, i.e., suppose there exists a nondegenerate df G on $\mathbb{R}^d$ such that for $\boldsymbol{x} = (x_1, \dots, x_d) \le \mathbf{0} \in \mathbb{R}^d$

$$\begin{aligned} P\left(n\left(\boldsymbol{M}^{(n)} - \mathbf{1}\right) \le \boldsymbol{x}\right) &= P\left(n\left(M_1^{(n)} - 1\right) \le x_1, \dots, n\left(M_d^{(n)} - 1\right) \le x_d\right) \\ &\to_{n\to\infty} G(\boldsymbol{x}). \end{aligned} \tag{14}$$

Then, G is necessarily a *multivariate max-stable* or *multivariate extreme-value df*, with *extreme-value copula* C_G and standard negative exponential margins G_j, $j = 1, \dots, d$, see (3). In the sequel we refer to the df G in (14) as *standard* multivariate max-stable df. Precisely, the form of G is

$$G(\boldsymbol{x}) = C_G(G_1(x_1), \dots, G_d(x_d)),$$

where the copula C_G can be expressed in terms of $\|\cdot\|_D$, a *D-norm* on $\mathbb{R}^d$, via

$$C_G(\boldsymbol{u}) = \exp\left(-\|\log u_1, \dots, \log u_d\|_D\right), \quad \boldsymbol{u} \in [0,1]^d, \tag{15}$$

while the margins G_j, $j = 1, \dots, d$, are as in (3). Therefore, the distribution in (14) has the representation

$$G(\boldsymbol{x}) = \exp\left(-\|\boldsymbol{x}\|_D\right), \qquad \boldsymbol{x} \le \mathbf{0} \in \mathbb{R}^d. \tag{16}$$

The convergence result in (14) implies that $C^{(n)}(\boldsymbol{u}) := C^n(\boldsymbol{u}^{1/n}) \to_{n\to\infty} C_G(\boldsymbol{u})$, for all $\boldsymbol{u} \in [0,1]^d$, see e.g. Falk (2019, Corollary 3.1.12). For brevity, with a little abuse of notation we also denote this latter fact by $C \in \mathcal{D}(C_G)$. By Theorem 2.3.3 in Falk (2019), there exists a rv $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ with $Z_i \geq 0$, $E(Z_i) = 1$, $1 \leq i \leq d$, such that

$$\left\|\boldsymbol{x}\right\|_D = E\left(\max\left(|x_i|\, Z_i\right)\right), \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

Examples of D-norms are the sup-norm $\left\|\boldsymbol{x}\right\|_\infty = \max_{1\leq i\leq d} |x_i|$, or the complete family of logistic norms $\left\|\boldsymbol{x}\right\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{1/p}$, $p \geq 1$. For a recent account on multivariate extreme-value theory and D-norms we refer to Falk (2019). In particular, Proposition 3.1.5 in Falk (2019) implies that the convergence result in (14) is also equivalent to the expansion

$$C(\boldsymbol{u}) = 1 - \left\|\boldsymbol{1} - \boldsymbol{u}\right\|_D + o(\left\|\boldsymbol{1} - \boldsymbol{u}\right\|) \tag{17}$$

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

In a first step we drop the term $o(\left\|\boldsymbol{1} - \boldsymbol{u}\right\|)$ in expansion (17) and require that there exists $\boldsymbol{u}_0 \in (0,1)^d$, such that

$$C(\boldsymbol{u}) = 1 - \left\|\boldsymbol{1} - \boldsymbol{u}\right\|_D, \qquad \boldsymbol{u} \in [\boldsymbol{u}_0, \boldsymbol{1}] \subset \mathbb{R}^d. \tag{18}$$

A copula, which satisfies the above expansion is a *generalized Pareto copula* (GPC). The significance of GPCs for multivariate extreme-value theory is explained in Falk et al. (2019) and in Falk (2019, Section 3.1).

Note that

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

defines a multivariate df only in dimension $d = 2$, see, e.g., McNeil and Nešlehová (2009, Examples 2.1, 2.2). But one can find for arbitrary dimension $d \geq 2$ a rv, whose df satisfies equation (18), see e.g. Falk (2019, equation 2.15). For this reason, we require the condition in (18) only on some upper interval $[\boldsymbol{u}_0, \boldsymbol{1}] \subset \mathbb{R}^d$.

The df of $n\left(\boldsymbol{M}^{(n)}-\mathbf{1}\right)$ is, for $\boldsymbol{x}<\mathbf{0}\in\mathbb{R}^d$ and n large so that $\mathbf{1}+\boldsymbol{x}/n\geq\boldsymbol{u}_0$,

$$P\left(n\left(\boldsymbol{M}^{(n)}-\mathbf{1}\right)\leq\boldsymbol{x}\right)=\left(1-\frac{1}{n}\left\|\boldsymbol{x}\right\|_D\right)^n=:F^{(n)}(\boldsymbol{x}).$$

Suppose that the norm $\|\cdot\|_D$ has partial derivatives of order d. Then the df $F^{(n)}(\boldsymbol{x})$ has for $\mathbf{1}+\boldsymbol{x}/n\geq\boldsymbol{u}_0$ the density

$$f^{(n)}(\boldsymbol{x}):=\frac{\partial^d}{\partial x_1\dots\partial x_d}F^{(n)}(\boldsymbol{x})=\frac{\partial^d}{\partial x_1\dots\partial x_d}\left(1-\frac{1}{n}\left\|\boldsymbol{x}\right\|_D\right)^n. \tag{19}$$

As for the standard multivariate max-stable df G in (16), its density exists and is given by

$$g(\boldsymbol{x}):=\frac{\partial^d}{\partial x_1\dots\partial x_d}G(\boldsymbol{x})=\frac{\partial^d}{\partial x_1\dots\partial x_d}\exp\left(-\left\|\boldsymbol{x}\right\|_D\right),\qquad\boldsymbol{x}\leq\mathbf{0}\in\mathbb{R}^d. \tag{20}$$

We are now ready to state our first multivariate extension of the convergence in total variation in equation (5). For brevity, we occasionally denote with the same letter a Borel measure and its distribution function.

Theorem 2.1. *Suppose the rv $\boldsymbol{U}$ follows a GPC C with corresponding D-norm $\|\cdot\|_D$, which has partial derivatives of order $d\geq 2$. Then*

$$\sup_{A\in\mathbb{B}^d}\left|P\left(n\left(\boldsymbol{M}^{(n)}-\mathbf{1}\right)\in A\right)-G(A)\right|\to_{n\to\infty}0,$$

where $\mathbb{B}^d$ denotes the Borel-σ-field in $\mathbb{R}^d$.

Remark 2.2. Note that we can write a GPC

$$C(\boldsymbol{u})=1-\left\|\mathbf{1}-\boldsymbol{u}\right\|_p=1-\left(\sum_{i=1}^d(1-u_i)^p\right)^{1/p},\qquad\boldsymbol{u}\in[\boldsymbol{u}_0,\mathbf{1}]\subset\mathbb{R}^d,$$

where the D-norm $\|\cdot\|_D$ is a logistic norm $\|\cdot\|_p$, $p\geq 1$, as an *Archimedean* copula

$$C(\boldsymbol{u})=\varphi^{-1}\left(\sum_{i=1}^d\varphi(u_i)\right),\qquad\boldsymbol{u}\in[\boldsymbol{u}_0,\mathbf{1}]\subset\mathbb{R}^d.$$

The *generator function* $\varphi:(0,1]\to[0,\infty)$ is in general strictly decreasing and convex, with $\varphi(1)=0$ (see, e.g. McNeil and Nešlehová 2009). Just set here $\varphi(u):=$

$(1-u)^p$, $u \in [0,1]$. Note that we require the Archimedean structure of C only in its upper tail; this allows the incorporation of $\varphi(u) = (1-u)^p$ as a generator function in arbitrary dimension $d \geq 2$, not only for $d = 2$. The partial differentiability condition on the D-norm in Theorem 2.1 now reduces to the existence of the derivative of order d of $\varphi(u)$ in a left neighbourhood of 1.

For the proof of Theorem 2.1 we establish the following auxiliary result.

Lemma 2.3. *Choose $\varepsilon \in (0,1)$ and $\boldsymbol{x}_\varepsilon < \boldsymbol{0} \in \mathbb{R}^d$ with $G([\boldsymbol{x}_\varepsilon, \boldsymbol{0}]) \geq 1-\varepsilon$. Then we have for $\boldsymbol{x} \in [\boldsymbol{x}_\varepsilon, \boldsymbol{0}]$*

$$f^{(n)}(\boldsymbol{x}) \to_{n\to\infty} g(\boldsymbol{x}). \tag{21}$$

Proof. $G(\boldsymbol{x})$ can be seen as the function composition $(\ell \circ \phi)(\boldsymbol{x})$, where we set $\ell(y) = \exp(y)$ and $\phi(\boldsymbol{x}) = -\left\|\boldsymbol{x}\right\|_D$. Then, by *Faá di Bruno's formula*, the density in (20) is equal to

$$g(\boldsymbol{x}) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} \exp(\phi(\boldsymbol{x})) = G(\boldsymbol{x}) \sum_{\mathcal{P}\in\mathscr{P}} \prod_{B\in\mathcal{P}} \frac{\partial^{|B|}\phi(\boldsymbol{x})}{\partial^B \boldsymbol{x}}, \tag{22}$$

where $\mathscr{P}$ is the set of all partitions of $\{1,\dots,d\}$ and the product is over all blocks B of a partition $\mathcal{P} \in \mathscr{P}$. In particular, $B = (i_1,\dots,i_k)$ with each $i_j \in \{1,\dots,d\}$, and the cardinality of each block is denoted by $|B| = k$. Finally, for a function $h : \mathbb{R}^d \to \mathbb{R}$ we define $\partial^{|B|}h/\partial^B \boldsymbol{x} := \partial^k h/\partial x_{i_1},\dots,\partial x_{i_k}$.

Similarly, $F^{(n)}(\boldsymbol{x})$ can be seen as the function composition $(\ell \circ \phi_n)(\boldsymbol{x})$, where we set $\phi_n(\boldsymbol{x}) := -n\log(1/(1-n^{-1}\left\|\boldsymbol{x}\right\|_D))$. Then, $F^{(n)}(\boldsymbol{x}) = \exp(\phi_n(\boldsymbol{x}))$ and, once again by the Faá di Bruno's formula, the density in (19) is equal to

$$f^{(n)}(\boldsymbol{x}) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} \exp(\phi_n(\boldsymbol{x})) = F^{(n)}(\boldsymbol{x}) \sum_{\mathcal{P}\in\mathscr{P}} \prod_{B\in\mathcal{P}} \frac{\partial^{|B|}\phi_n(\boldsymbol{x})}{\partial^B \boldsymbol{x}}.$$

Clearly, $F^{(n)}(\boldsymbol{x}) \to_{n\to\infty} G(\boldsymbol{x})$ for all $\boldsymbol{x} \in [\boldsymbol{x}_\varepsilon, \boldsymbol{0}]$. Next, $\phi_n(\boldsymbol{x})$ can be seen as the function composition $(\sigma_n \circ \phi)(\boldsymbol{x})$, where we set $\sigma_n(y) = -n\log(1/(1+n^{-1}y))$. Thus, again by the Faá di Bruno's formula we have that for each block B

$$\frac{\partial^{|B|}\phi_n(\boldsymbol{x})}{\partial^B \boldsymbol{x}} = \sum_{\mathcal{P}_B \in \mathscr{P}_B} \left.\frac{\partial^{|\mathcal{P}_B|}\sigma_n(y)}{\partial y^{|\mathcal{P}_B|}}\right|_{y=\phi(\boldsymbol{x})} \prod_{b\in\mathcal{P}_B} \frac{\partial^{|b|}\phi(\boldsymbol{x})}{\partial^b \boldsymbol{x}},$$

where $\mathscr{P}_B$ is the set of all partitions of $B = (i_1, \dots, i_k)$ and the product is over all blocks b of partition $\mathcal{P}_B \in \mathscr{P}_B$. It is not difficult to check that

$$\frac{\partial^{|\mathcal{P}_B|}\sigma_n(y)}{\partial y^{|\mathcal{P}_B|}} = (-1)^{1+|\mathcal{P}_B|}\,(|\mathcal{P}_B|-1)!\,(1+y/n)^{-|\mathcal{P}_B|}\; n^{-|\mathcal{P}_B|+1}.$$

Then,

$$\frac{\partial^{|\mathcal{P}_B|}\sigma_n(y)}{\partial y^{|\mathcal{P}_B|}} \to_{n\to\infty} \begin{cases} 1, & \text{if } |\mathcal{P}_B| = 1, \\ 0, & \text{if } |\mathcal{P}_B| > 1. \end{cases}$$

Notice that $|\mathcal{P}_B| = 1$ when $\mathcal{P}_B = B$ and in this case $b = B$. Consequently, for all $\boldsymbol{x} \in [\boldsymbol{x}_\varepsilon, \boldsymbol{0}]$, we have

$$\frac{\partial^{|B|}\phi_n(\boldsymbol{x})}{\partial^B \boldsymbol{x}} \to_{n\to\infty} \frac{\partial^{|B|}\phi(\boldsymbol{x})}{\partial^B \boldsymbol{x}}.$$

Therefore, the pointwise convergence in (21) follows. □

Proof of Theorem 2.1. It is sufficient to consider $A \subset \mathbb{B}^d \cap (-\infty, 0]^d$, where $\mathbb{B}^d$ denotes the Borel-σ-field in $\mathbb{R}^d$. Moreover, choose $\varepsilon > 0$ and $\boldsymbol{x}_\varepsilon < \boldsymbol{0} \in \mathbb{R}^d$ with $G([\boldsymbol{x}_\varepsilon, \boldsymbol{0}]) \geq 1 - \varepsilon$.

We already know that

$$\sup_{\boldsymbol{x}\leq \boldsymbol{0}} \left| P\left(n\left(\boldsymbol{M}^{(n)} - \boldsymbol{1}\right) \leq \boldsymbol{x}\right) - G(\boldsymbol{x})\right| \to_{n\to\infty} 0,$$

which implies

$$\left| P\left(n\left(\boldsymbol{M}^{(n)} - \boldsymbol{1}\right) \in [\boldsymbol{x}_\varepsilon, \boldsymbol{0}]\right) - G([\boldsymbol{x}_\varepsilon, \boldsymbol{0}])\right| \to_{n\to\infty} 0 \tag{23}$$

and, thus,

$$\limsup_{n\to\infty} P\left(n\left(\boldsymbol{M}^{(n)} - \boldsymbol{1}\right) \in [\boldsymbol{x}_\varepsilon, \boldsymbol{0}]^{\complement}\right) \leq \varepsilon$$

or

$$\limsup_{n\to\infty} \sup_{A\in\mathbb{B}^d\cap[\boldsymbol{x}_\varepsilon,\boldsymbol{0}]^\complement} \left| P\left(n\left(\boldsymbol{M}^{(n)}-\boldsymbol{1}\right)\in A\right)-G(A)\right|$$
$$\leq \limsup_{n\to\infty} P\left(n\left(\boldsymbol{M}^{(n)}-\boldsymbol{1}\right)\in [\boldsymbol{x}_\varepsilon,\boldsymbol{0}]^\complement\right)+G\left([\boldsymbol{x}_\varepsilon,\boldsymbol{0}]^\complement\right)\leq 2\varepsilon.$$

As $\varepsilon > 0$ was arbitrary, it is therefore sufficient to establish

$$\sup_{A\in\mathbb{B}^d\cap[\boldsymbol{x}_\varepsilon,\boldsymbol{0}]} \left| P\left(n\left(\boldsymbol{M}^{(n)}-\boldsymbol{1}\right)\in A\right)-G(A)\right| \to_{n\to\infty} 0.$$

Now, from equation (23) we know that

$$\int_{[\boldsymbol{x}_\varepsilon,\boldsymbol{0}]} f^{(n)}(\boldsymbol{x})\,dx \to_{n\to\infty} \int_{[\boldsymbol{x}_\varepsilon,\boldsymbol{0}]} g(\boldsymbol{x})\,d\boldsymbol{x}.$$

Together with (21), we can apply Scheffé's lemma and obtain

$$\int_{[\boldsymbol{x}_\varepsilon,\boldsymbol{0}]} \left|f^{(n)}(\boldsymbol{x})-g(\boldsymbol{x})\right|\,d\boldsymbol{x} \to_{n\to\infty} 0.$$

The bound

$$\sup_{A\in\mathbb{B}^d\cap[\boldsymbol{x}_\varepsilon,\boldsymbol{0}]} \left| P\left(n\left(\boldsymbol{M}^{(n)}-\boldsymbol{1}\right)\in A\right)-G(A)\right| \leq \int_{[\boldsymbol{x}_\varepsilon,\boldsymbol{0}]} \left|f^{(n)}(\boldsymbol{x})-g(\boldsymbol{x})\right|\,d\boldsymbol{x}$$

now implies the assertion of Theorem 2.1. □

Next we extend Theorem 2.1 to a copula C, which is in a *differentiable neighborhood* of a GPC, defined next. Suppose that C satisfies expansion (17), where the D-norm $\|\cdot\|_D$ on $\mathbb{R}^d$ has partial derivatives of order d. Assume also that C is such that for each nonempty block of indices $B=(i_1,\dots,i_k)$ of $\{1,\dots,d\}$,

$$\frac{\partial^k}{\partial x_{i_1},\dots,\partial x_{i_k}} n\left(C\left(\boldsymbol{1}+\frac{\boldsymbol{x}}{n}\right)-1\right) \to_{n\to\infty} \frac{\partial^k}{\partial x_{i_1},\dots,\partial x_{i_k}}\phi(\boldsymbol{x}), \tag{24}$$

holds true for all $\boldsymbol{x}<\boldsymbol{0}\in\mathbb{R}^d$, where $\phi(\boldsymbol{x})=-\|\boldsymbol{x}\|_D$.

Theorem 2.4. *Suppose the copula C satisfies conditions (17) and (24). Then we obtain*

$$\sup_{A\in\mathbb{B}^d} \left| P\left(n\left(\boldsymbol{M}^{(n)}-\boldsymbol{1}\right)\in A\right)-G(A)\right| \to_{n\to\infty} 0,$$

where G is the standard max-stable distribution with corresponding D-norm $\|\cdot\|_D$, i.e., it has df $G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_D)$, $\boldsymbol{x} \leq \boldsymbol{0} \in \mathbb{R}^d$.

Proof. The proof of Theorem 2.4 is similar to that of Theorem 2.1, but this time we resort to a variant of Lemma 2.3 as follows. Note that for $n \in \mathbb{N}$,

$$P\left(n\left(\boldsymbol{M}^{(n)} - \boldsymbol{1}\right) \leq \boldsymbol{x}\right) = C^n\left(\boldsymbol{1} + \frac{\boldsymbol{x}}{n}\right), \qquad \boldsymbol{x} \leq \boldsymbol{0} \in \mathbb{R}^d.$$

Moreover, $C^n(\boldsymbol{1} + \boldsymbol{x}/n)$ is the function composition $(\ell \circ \phi_n)(\boldsymbol{x})$, where we now set $\phi_n(\boldsymbol{x}) := n\log(C(\boldsymbol{1} + \boldsymbol{x}/n))$. Furthermore, $\phi_n(\boldsymbol{x})$ is the composition function $(\sigma_n \circ v_n)(\boldsymbol{x})$, where we set $v_n(\boldsymbol{x}) := n(C(\boldsymbol{1} + \boldsymbol{x}/n) - 1)$ and σ_n is as in the proof of Lemma 2.3. Then, in the Faá di Bruno's formula we have that for each block B,

$$\frac{\partial^{|B|}\phi_n(\boldsymbol{x})}{\partial^B \boldsymbol{x}} = \sum_{\mathcal{P}_B \in \mathscr{P}_B} \frac{\partial^{|\mathcal{P}_B|}\sigma_n(y)}{\partial y^{|\mathcal{P}_B|}}\bigg|_{y = v_n(\boldsymbol{x})} \prod_{b \in \mathcal{P}_B} \frac{\partial^{|b|} v_n(\boldsymbol{x})}{\partial^b \boldsymbol{x}}.$$

By assumptions (17) and (24) we obtain that, for each block b of a partition $\mathcal{P}_B \in \mathscr{P}_B$,

$$\frac{\partial^{|b|} v_n(\boldsymbol{x})}{\partial^b \boldsymbol{x}} \to_{n\to\infty} \frac{\partial^{|b|}\phi(\boldsymbol{x})}{\partial^b \boldsymbol{x}}, \quad \boldsymbol{x} < \boldsymbol{0} \in \mathbb{R}^d.$$

Therefore, as in Lemma 2.3, we obtain

$$\frac{\partial^{|B|}\phi_n(\boldsymbol{x})}{\partial^B \boldsymbol{x}} \to_{n\to\infty} \frac{\partial^{|B|}\phi(\boldsymbol{x})}{\partial^B \boldsymbol{x}}, \quad \boldsymbol{x} < \boldsymbol{0} \in \mathbb{R}^d. \tag{25}$$

and the result follows. □

EXAMPLE 2.5. Consider, the *Gumbel-Hougaard family* $\{C_p : p \geq 1\}$ of Archimedean copulas, with generator function $\varphi_p(u) := (-\log(u))^p$, $p \geq 1$. This is an extreme-value family of copulas. In particular, we have

$$C_p(\boldsymbol{u}) = \exp\left(-\left(\sum_{i=1}^d (-\log(u_i))^p\right)^{1/p}\right) = 1 - \|\boldsymbol{1} - \boldsymbol{u}\|_p + o(\|1 - \boldsymbol{u}\|),$$

as $\boldsymbol{u} \in (0,1]^d$ converges to $\boldsymbol{1} \in \mathbb{R}^d$, i.e., condition (17) is satisfied, where the D-norm is the logistic norm $\|\cdot\|_p$ and the limiting distribution is $G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_p)$. The copula C_p also satisfies conditions (24). To prove it, we express $C_p(\boldsymbol{1} + \boldsymbol{x}/n)$

as the function composition $(\ell \circ \varphi_n)(\boldsymbol{x})$, with ℓ as in the proof of Lemma 2.3 and $\varphi_n(\boldsymbol{x}) := \log\left(C_p\left(\mathbf{1}+\boldsymbol{x}/n\right)\right)$. Observe that

$$n\varphi_n(\boldsymbol{x}) = -n\left\|\log\left(1+\frac{\boldsymbol{x}}{n}\right)\right\|_p =: -nt(s_n(\boldsymbol{x})),$$

where $t(\cdot) = \|\cdot\|_p$, $s_n(\boldsymbol{x}) = (s_n(x_1), \ldots, s_n(x_d))$, and $s_n(\cdot) = \log(1 + \cdot/n)$. Hence, applying the Faá di Bruno's formula to the partial derivatives of $n(\ell \circ \varphi_n(x) - 1)$ and noting that, on one hand, $C_p\left(\mathbf{1}+\boldsymbol{x}/n\right) \to_{n\to\infty} 1$, on the other hand,

$$\begin{aligned}
&\frac{\partial^k}{\partial x_{i_1}, \ldots, \partial x_{i_k}} n\varphi_n(\boldsymbol{x}) \\
&= -n\frac{\partial^k}{\partial y_{i_1}, \ldots, \partial y_{i_k}} t(\boldsymbol{y})\big|_{\boldsymbol{y}=s_n(\boldsymbol{x})} \frac{\partial s_n(x_{i_1})}{\partial x_{i_1}} \cdot \cdots \cdot \frac{\partial s_n(x_{i_k})}{\partial x_{i_k}} \\
&\simeq -n\prod_{j=1}^{k-1}(1-jp)\,\|\boldsymbol{x}\|_p^{1-kp}\, n^{kp-1}\prod_{j=1}^{k}\frac{|x_{i_j}|^p}{x_{i_j}} n^{-k(p-1)}\prod_{j=1}^{k}\left(1+\frac{x_{i_j}}{n}\right)^{-1} n^{-k} \\
&\to_{n\to\infty} -\frac{\partial^k}{\partial x_{i_1}, \ldots, \partial x_{i_k}}\|\boldsymbol{x}\|_p\,,
\end{aligned}$$

the desired result obtains. In particular, notice that we pass from the first to second line of the above display by computing partial derivatives, then from the second to the third one by exploiting the asymptotic equivalence $\log(1+y) \simeq y$, for $y \to 0$.

Example 2.6. Consider the copula

$$C(\boldsymbol{u}) = 1 - d + \sum_{i=1}^{d} u_i + \sum_{2\le i\le d}\left((-1)^i \sum_{\substack{B\subseteq\{1,\ldots,d\}\\|B|=i}}\left(\sum_{j\in B}\frac{1}{1-u_j} - d + 1\right)^{-1}\right). \tag{26}$$

This provides the d-dimensional version (with $d \geq 2$) of the 2-dimensional copula associated to the df discussed in (Resnick, 2008, Example 5.14). It can be checked that $C \in \mathcal{D}(C_G)$, where C_G is, for all $\boldsymbol{u} \in [0,1]^d$, the extreme-value copula

$$C_G(\boldsymbol{u}) = \exp\left(\sum_{i=1}^{d}\log u_i + \sum_{2\le i\le d}\left((-1)^{i+1}\sum_{\substack{B\subseteq\{1,\ldots,d\}\\|B|=i}}\left(\sum_{j\in B}\frac{1}{\log u_j} - d + 1\right)^{-1}\right)\right). \tag{27}$$

Then, by Falk (2019, Proposition 3.1.5 and Corollary 3.1.12) the copula in (26) satisfies condition (17), with D-norm

$$\|\boldsymbol{x}\|_D = \sum_{i=1}^{d} |x_i| + \sum_{2\le i\le d} \left((-1)^{i+1} \sum_{\substack{B\subseteq\{1,\dots,d\}\\|B|=i}} \left(\sum_{j\in B} \frac{1}{|x_j|} \right)^{-1} \right).$$

The copula in (26) also complies conditions in (24). Indeed, for $2 \le k \le d$,

$$\frac{\partial^k}{\partial x_{i_1},\dots,\partial x_{i_k}} \|\boldsymbol{x}\|_D = \sum_{k\le j\le d} \left((-1)^{j+1} k! \sum_{\substack{\mathcal{I}\subseteq B\subseteq\{1,\dots,d\}\\|B|=j}} \left(\sum_{l\in B} \frac{1}{|x_l|} \right)^{-(k+1)} \prod_{v=1}^{k} \frac{1}{x_{i_v}^2} \frac{|x_{i_v}|}{x_{i_v}} \right),$$

where $\mathcal{I} = \{i_1,\dots,i_k\}$. When $k = 1$, $(\partial/\partial x_{i_k}) \|\boldsymbol{x}\|_D$ is given by the right-hand side of the above expression plus the term $|x_{i_k}|/x_{i_k}$. Furthermore, for $2 \le k \le d$,

$$\begin{aligned}&\frac{\partial^k}{\partial x_{i_1},\dots,\partial x_{i_k}} C(\mathbf{1} + \boldsymbol{x}/n)\\ &= \frac{1}{n} \left(\sum_{k\le j\le d} \left((-1)^{j+1} k! \sum_{\substack{\mathcal{I}\subseteq B\subseteq\{1,\dots,d\}\\|B|=j}} \left(\sum_{l\in B} \frac{1}{x_l} + \frac{d-1}{n} \right)^{-(k+1)} \prod_{v=1}^{k} \frac{1}{x_{i_v}^2} \right) \right).\end{aligned}$$

When $k = 1$, $n(\partial/\partial x_{i_k}) C(\mathbf{1} + \boldsymbol{x}/n)$ is given by the right-hand side of the above expression plus 1. Therefore, for $k = 1,\dots,d$, we have that

$$n \frac{\partial^k}{\partial x_{i_1},\dots,\partial x_{i_k}} C(\mathbf{1} + \boldsymbol{x}/n) \to_{n\to\infty} -\frac{\partial^k}{\partial x_{i_1},\dots,\partial x_{i_k}} \|\boldsymbol{x}\|_D$$

and the desired result obtains.

Let C be a copula and C^n be the copula of the corresponding componentwise maxima, see (13). We recall that $C^{(n)}(\boldsymbol{u}) := C^n(\boldsymbol{u}^{1/n})$, $\boldsymbol{u} \in [0,1]^d$. Assume that $C \in \mathcal{D}(C_G)$, where C_G is an extreme-value copula. A readily demonstrable result implied by Theorem 2.4 is the convergence of $C^{(n)}$ to C_G in variational distance.

Corollary 2.7. *Assume C satisfies conditions (17) and (24), with continuous partial derivatives of order up to d on $(0,1)^d$, then*

$$\sup_{A\in\mathbb{B}^d\cap[0,1]^d} |C^{(n)}(A) - C_G(A)| \to_{n\to\infty} 0.$$

Proof. For any $\boldsymbol{u}\in[0,1]^d$, define

$$\widetilde{C}^{(n)}(\boldsymbol{u}) := P\left(n\left(\boldsymbol{M}^{(n)} - \boldsymbol{1}\right) \leq \log \boldsymbol{u}\right) = C^n(1+\log \boldsymbol{u}/n).$$

By Theorem 2.4, $\widetilde{C}^{(n)}$ converges to C_G in variational distance. Now, for some $\varepsilon\in(0,1)$, set

$$\mathcal{U}_\varepsilon := \cup_{j=1}^d \{\boldsymbol{u}\in[0,1] : u_j < \varepsilon \text{ or } u_j > 1-\varepsilon\}.$$

In particular, fix $\varepsilon>0$ such that $C_G(\mathcal{U}_\varepsilon^\complement) > 1-\varepsilon_0$, for some arbitrarily small $\varepsilon_0\in(0,1)$. Then, using the Taylor expansion $u^{1/n} = 1 + n^{-1}\log u + o(1/n)$, with uniform reminder over $\mathcal{U}_\varepsilon^\complement$, together with the Lipschitz continuity of C, we obtain

$$\sup_{\boldsymbol{u}\in\mathcal{U}_\varepsilon^\complement} \left|C^{(n)}(\boldsymbol{u}) - \widetilde{C}^{(n)}(\boldsymbol{u})\right| \to_{n\to\infty} 0,$$

and therefore $\limsup_{n\to\infty} C^{(n)}(\mathcal{U}_\varepsilon) < \varepsilon_0$. This implies that, as $n\to\infty$, we have

$$\sup_{A\in\mathbb{B}^d\cap[0,1]^d} \left|C^{(n)}(A) - C_G(A)\right| \leq \sup_{A\in\mathbb{B}^d\cap\mathcal{U}_\varepsilon^\complement} \left|C_\varepsilon^{(n)}(A) - \widetilde{C}_\varepsilon^{(n)}(A)\right| + O(\varepsilon_0), \tag{28}$$

where $C_\varepsilon^{(n)}$ and $\widetilde{C}_\varepsilon^{(n)}$ are the normalised versions $C_\varepsilon^{(n)} = C^{(n)}/C^{(n)}(\mathcal{U}_\varepsilon^\complement)$ and $\widetilde{C}_\varepsilon^{(n)} = \widetilde{C}^{(n)}/\widetilde{C}^{(n)}(\mathcal{U}_\varepsilon^\complement)$. Finally, denote their densities by $c_\epsilon^{(n)}$ and $\widetilde{c}_\varepsilon^{(n)}$, respectively. Then, the supremum on the right hand side in (28) is attained at the set

$$\widetilde{\mathcal{U}}_\varepsilon^{(n)} := \left\{\boldsymbol{u}\in\mathcal{U}_\varepsilon^\complement : c_\varepsilon^{(n)}(\boldsymbol{u}) > \widetilde{c}_\varepsilon^{(n)}(\boldsymbol{u})\right\}.$$

Notice that $c_\varepsilon^{(n)}$ and $\widetilde{c}_\varepsilon^{(n)}$ are both positive on $\mathcal{U}_\varepsilon^\complement$, for n sufficiently large. Following steps similar to those in the proof of Theorem 2.4 and exploiting the continuity of the partial derivatives of C, we obtain

$$c_\varepsilon^{(n)}(\boldsymbol{u})/\widetilde{c}_\varepsilon^{(n)}(\boldsymbol{u}) \to_{n\to\infty} 1,$$

for all $\boldsymbol{u} \in \mathcal{U}_\varepsilon^{\complement}$. Therefore, $\widetilde{\mathcal{U}}_\varepsilon^{(n)} \downarrow \emptyset$ as $n \to \infty$ and the result follows. $\square$

3. The General Case

Let $\boldsymbol{X} = (X_1, \dots, X_d)$ be a rv with arbitrary df F. By Sklar's theorem (Sklar 1959, 1996) we can assume the representation

$$\boldsymbol{X} = (X_1, \dots, X_d) = \left(F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)\right),$$

where F_i is the df of X_i, $i = 1, \dots, d$, and $\boldsymbol{U} = (U_1, \dots, U_d)$ follows a copula, say C, corresponding to F.

Let $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots$ be independent copies of $\boldsymbol{X}$ and let $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \dots$ be independent copies of $\boldsymbol{U}$. Again we can assume the representation

$$\boldsymbol{X}^{(i)} = \left(X_1^{(i)}, \dots, X_d^{(i)}\right) = \left(F_1^{-1}\left(U_1^{(i)}\right), \dots, F_d^{-1}\left(U_d^{(i)}\right)\right), \qquad i = 1, 2, \dots$$

From the fact that each quantile function F_i^{-1} is monotone increasing, we obtain

$$\begin{aligned}
&\boldsymbol{M}^{(n)} \\
&:= \left(\max_{1 \le i \le n} X_1^{(i)}, \dots, \max_{1 \le i \le n} X_d^{(i)}\right) \\
&= \left(\max_{1 \le i \le n} F_1^{-1}\left(U_1^{(i)}\right), \dots, \max_{1 \le i \le n} F_d^{-1}\left(U_d^{(i)}\right)\right) \\
&= \left(F_1^{-1}\left(\max_{1 \le i \le n} U_1^{(i)}\right), \dots, F_d^{-1}\left(\max_{1 \le i \le n} U_d^{(i)}\right)\right) \\
&= \left(F_1^{-1}\left(1 + \frac{1}{n}\left(n\left(\max_{1 \le i \le n} U_1^{(i)} - 1\right)\right)\right), \dots, F_d^{-1}\left(1 + \frac{1}{n}\left(n\left(\max_{1 \le i \le n} U_d^{(i)} - 1\right)\right)\right)\right).
\end{aligned}$$

Theorem 2.1 now implies the following result.

Proposition 3.1. *Let $\boldsymbol{\eta} = (\eta_1, \dots \eta_d)$ be a rv with standard multivariate max-stable df $G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_D)$, $\boldsymbol{x} \le \boldsymbol{0} \in \mathbb{R}^d$. Let $\boldsymbol{X}$ be a rv with some distribution F and a copula C. Suppose that either C is a GPC with corresponding D-norm $\|\cdot\|_D$, which has partial derivatives of order $d \ge 2$, or C satisfies conditions (17) and*

(24). *Then,*

$$\sup_{A\in\mathbb{B}^d}\left|P\left(\boldsymbol{M}^{(n)}\in A\right)-P\left(\left(F_1^{-1}\left(1+\frac{1}{n}\eta_1\right),\dots,F_d^{-1}\left(1+\frac{1}{n}\eta_d\right)\right)\in A\right)\right|$$

$$\to_{n\to\infty} 0.$$

Finally, we generalise the result in Proposition 3.1 to the case where the rv of componentwise maxima is suitably normalised. Precisely, we now consider the case that $F\in\mathcal{D}(G^*_{\boldsymbol{\gamma}})$, i.e. F belongs to the domain of attraction of a *generalised* multivariate max-stable df $G^*_{\boldsymbol{\gamma}}$, with tail index $\boldsymbol{\gamma}=(\gamma_1,\dots,\gamma_d)$, e.g. Falk et al. (2019, Ch. 4). This means that there exist sequences of norming vectors $\boldsymbol{a}_n=(a_n^{(1)},\dots,a_n^{(d)})>\boldsymbol{0}$ and $\boldsymbol{b}_n=(b_n^{(1)},\dots,b_n^{(d)})\in\mathbb{R}^d$, for $n\in\mathbb{N}$, such that $(\boldsymbol{M}^{(n)}-\boldsymbol{b}_n)/\boldsymbol{a}_n\to_D\boldsymbol{Y}$ as $n\to\infty$, where $\boldsymbol{Y}$ is a rv with distribution $G^*_{\boldsymbol{\gamma}}$. The copula of $G^*_{\boldsymbol{\gamma}}$ is the extreme-value copula in (15) and its margins $G^*_{\gamma_j}$, $j=1,\dots,d$, are members of the generalised extreme-value family of dfs in (7).

To attain the convergence in variational distance, we combine Proposition 3.1, obtained under conditions involving only dependence structures, with univariate von Mises conditions on the margins $F_1,\dots,F_d$, see (8)-(11). We denote by $\boldsymbol{x}_0:=(x_0^{(1)},\dots,x_0^{(d)})$, where $x_0^{(j)}:=\sup\{x\in\mathbb{R}:\ F_j(x)<1\}$, $j=1,\dots,d$, the vector of endpoints.

Corollary 3.2. *Let $\boldsymbol{Y}$ and $\boldsymbol{X}$ be rvs with a generalised multivariate max-stable df $G^*_{\boldsymbol{\gamma}}$ and a continuous df F, respectively. Assume that $F\in\mathcal{D}(G^*_{\boldsymbol{\gamma}})$ and that its copula C satisfies the assumptions of Proposition (3.1). Assume further that, for $1\le j\le d$, the density of the j-th margin F_j of F satisfies one of the conditions (9)-(11) with f', γ and x_0 replaced by f_j', γ_j and $x_0^{(j)}$. Then,*

$$\sup_{A\in\mathbb{B}^d}\left|P\left(\frac{\boldsymbol{M}^{(n)}-\boldsymbol{b}_n}{\boldsymbol{a}_n}\in A\right)-P\left(\boldsymbol{Y}\in A\right)\right|\to_{n\to\infty}0.$$

Proof. Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_d)$ be rv with standard multivariate max-stable distribution $G(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}\|_D)$. Define,

$$\boldsymbol{Y}_n := \left(\frac{1}{a_n^{(1)}}\left(F_1^{-1}\left(1+\frac{1}{n}\eta_1\right) - b_n^{(1)}\right), \ldots, \frac{1}{a_n^{(d)}}\left(F_d^{-1}\left(1+\frac{1}{n}\eta_d\right) - b_n^{(d)}\right)\right).$$

Observe that

$$\sup_{A\in\mathbb{B}^d}\left|P\left(\frac{\boldsymbol{M}^{(n)}-\boldsymbol{b}_n}{\boldsymbol{a}_n}\in A\right) - P\left(\boldsymbol{Y}\in A\right)\right| \le T_{1,n} + T_{2,n},$$

where

$$T_{1,n} := \sup_{A\in\mathbb{B}^d}\left|P\left(\boldsymbol{M}^{(n)}\in A\right) - P\left(\left(F_1^{-1}\left(1+\frac{1}{n}\eta_1\right), \ldots, F_d^{-1}\left(1+\frac{1}{n}\eta_d\right)\right)\in A\right)\right|$$

and

$$T_{2,n} := \sup_{A\in\mathbb{B}^d}\left|P\left(\boldsymbol{Y}_n\in A\right) - P\left(\boldsymbol{Y}\in A\right)\right|.$$

By Proposition 3.1, $T_{1,n} \to_{n\to\infty} 0$. To show that $T_{2,n} \to_{n\to\infty} 0$, it is sufficient to show pointwise convergence of the probability density function of $\boldsymbol{Y}_n$ to that of $\boldsymbol{Y}$ and then to appeal to the Scheffé's lemma. First, notice that $G^*_{\boldsymbol{\gamma}}$ and G have the same extreme-value copula. Thus, from (15) it follows that, for $\boldsymbol{x}\in\mathbb{R}^d$, $G^*_{\boldsymbol{\gamma}}(\boldsymbol{x}) = G(\boldsymbol{u}(\boldsymbol{x}))$, where $\boldsymbol{u}(\boldsymbol{x}) = \left(u^{(1)}(x_1), \ldots, u^{(d)}(x_d)\right)$ with $u^{(j)}(x_j) = \log G^*_{\gamma_j}(x_j)$ for $j = 1, \ldots, d$. Now, define $Q^{(n)}(\boldsymbol{x}) := P(\boldsymbol{Y}_n \le \boldsymbol{x}) = G(\boldsymbol{u}_n(\boldsymbol{x}))$, for $\boldsymbol{x}\in\mathbb{R}^d$, where $\boldsymbol{u}_n(\boldsymbol{x}) = \left(u_n^{(1)}(x_1), \ldots, u_n^{(d)}(x_d)\right)$ with

$$u_n^{(j)}(x_j) := -n\left(1 - F_j\left(a_n^{(j)}x_j + b_n^{(j)}\right)\right), \quad 1\le j\le d.$$

Consequently, as $n\to\infty$,

$$\begin{aligned}
\frac{\partial^d}{\partial x_1\ldots\partial x_d}Q^{(n)}(\boldsymbol{x}) &= g(\boldsymbol{u}_n(\boldsymbol{x}))\prod_{j=1}^d \frac{n a_n^{(j)} F_j\left(a_n^{(j)}x_j + b_n^{(j)}\right)^{n-1} f_j\left(a_n^{(j)}x_j + b_n^{(j)}\right)}{F_j\left(a_n^{(j)}x_j + b_n^{(j)}\right)^{n-1}}\\
&\simeq g(\boldsymbol{u}(\boldsymbol{x}))\prod_{j=1}^d \frac{g^*_{\gamma_j}(x_j)}{G^*_{\gamma_j}(x_j)}\\
&= \frac{\partial^d}{\partial x_1\ldots\partial x_d}G(\boldsymbol{u}(\boldsymbol{x})) = \frac{\partial^d}{\partial x_1\ldots\partial x_d}G^*_{\boldsymbol{\gamma}}(\boldsymbol{x}),
\end{aligned}$$

where g is as in (22) and $g^*_{\gamma_j}(x) = (\partial/\partial x)G^*_{\gamma_j}(x)$, $1 \leq j \leq d$. In particular, the second line follows from the continuity of g and Proposition 2.5 in Resnick (2008). The proof is now complete. $\square$

4. Applications

The strong convergence results established in Sections 2 and 3 can be used to refine asymptotic statistical theory for extremes. Max-stable distributions have been used for modelling extremes in several statistical analyses (e.g. Coles 2001, Ch. 8; Beirlant et al. 2004, Ch. 9; Marcon et al. 2017; Mhalla et al. 2017 to name a few). Parametric and nonparametric inferential procedures have been proposed for fitting max-stable models to the data (e.g., Gudendorf and Segers 2012; Berghaus et al. 2013; Marcon et al. 2017; Dombry et al. 2017). The asymptotic theory of the corresponding estimators is well established assuming that a sample of (componentwise) maxima follows a max-stable distribution. In practice, the latter provides only an approximate distribution for sample maxima. The recent results in Ferreira and de Haan (2015), Dombry (2015), Bücher and Segers (2018) and Berghaus and Bücher (2018) account for such model misspecification, in the univariate setting. In the multivariate case, in Bücher and Segers (2014), weak convergence and consistency in probability of empirical copulas, under suitable second order conditions (Bücher et al. e.g. 2019), have been studied. This is the only multivariate contribution focusing on the problem of convergences, under model misspecification, as far as we known. In the sequel, we illustrate how our variational convergence results, obtained under conditions (17) and (24), allow to establish a stronger form of consistency, for both frequentist and Bayesian procedures. To do that, we resort to the notion of remote contiguity.

Definition 4.1. *(Kleijn 2017) For $k \in \mathbb{N}$, let r_k, s_k be real valued sequences such that $0 < r_k, s_k \to_{k\to\infty} 0$. Let μ_k and ν_k be sequences of probability measures. Then, ν_k is said r_k-to-s_k-remotely contiguous with respect to μ_k if $\mu_k(E_k) = o(r_k)$, for a* *sequence of measurable events E_k, implies $\nu_k(E_k) = o(s_k)$. In this case, we write $s_k^{-1}\nu_k \lhd r_k^{-1}\mu_k$.*

4.1. **Frequentist approach.** Let Θ denote a parameter space (possibly infinite dimensional) and $\theta \in \Theta$ be a parameter of interest. Let $\boldsymbol{Y}$ be a d-dimensional rv with a df F, pertaining to a probability measure μ_0 on $\mathbb{B}^d$. Denote by μ_k the corresponding k-fold product measure. Let $\boldsymbol{Y}^{(1:k)} = (\boldsymbol{Y}^{(1,k)}, \dots, \boldsymbol{Y}^{(k,k)})$ be a sequence of k iid copies of $\boldsymbol{Y}$. Consider a measurable map $T_k : \times_{i=1}^k \mathbb{R}^d \to \Theta$ and let

$$\widehat{\theta}_k := T_k(\boldsymbol{Y}^{(1:k)})$$

be an estimator of θ. Let $\mathscr{D}$ denote a metric on Θ.

If for every $\varepsilon > 0$ there are constants $c_\varepsilon, c'_\varepsilon > 0$ such that $\mu_k(\mathscr{D}(\widehat{\theta}_k, \theta) > \varepsilon) = o(e^{-c_\varepsilon k})$ and $k^{1+c'_\varepsilon}\nu_k \lhd e^{c_\varepsilon k}\mu_k$, then, we can conclude by Borel-Cantelli lemma that

$$\mathscr{D}(T_k(\boldsymbol{Z}^{(1:k)}), \theta) \to_{k\to\infty} 0, \qquad \nu_k\text{-almost surely},$$

where $\boldsymbol{Z}^{(1:k)} = (\boldsymbol{Z}^{(1,k)}, \dots, \boldsymbol{Z}^{(k,k)})$ is a sequence of iid rv with common probability measure $\nu_{0,k}$ on $\mathbb{B}^d$, and ν_k is the corresponding k-fold product measure. The required form of remote contiguity easily obtains if $\sup_{A\in\mathbb{B}^d} |\nu_{0,k}(A) - \mu_0(A)| \to_{k\to\infty} 0$, μ_0 and $\nu_{0,k}$ have the same support and continuous Lebesgue densities, $p_{0,k}$ and m_0, satisfying

$$\sup_{k\geq k_0} \rho_\delta(\nu_{0,k}, \mu_0) := \sup_{k\geq k_0} \int_{\mathcal{X}_{\delta,k}} \left(p_{0,k}(\boldsymbol{x})/m_0(\boldsymbol{x})\right)^\delta p_{0,k}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} < \infty, \tag{29}$$

for some $\delta \in (0,1]$ and $k_0 \in \mathbb{N}$, where $\mathcal{X}_{\delta,k} = \{\boldsymbol{x} \in \mathbb{R}^d : p_{0,k}(\boldsymbol{x})/\mu_0(\boldsymbol{x}) > e^{1/\delta}\}$. Essentially, variational convergence and (29) guarantee that the fourth moments and the expectations of the triangular array of variables $\{\log p_{0,k}(\boldsymbol{Z}^{(i,k)}) - \log m_0(\boldsymbol{Z}^{(i,k)}), 1 \leq i \leq k; k \geq k_0 + k'_0\}$ are uniformly bounded and asymptotically null, respectively, for a sufficiently large $k'_0 \in \mathbb{N}$. The corresponding sequence of (rescaled) log-likelihood ratios, then, converges to 0 by the strong law of large numbers.

This novel asymptotic technique can be fruitfully applied to parameter estimation problems for multivariate max-stable models. In this context, the probability measure μ_0 can be associated to a multivariate max-stable df $G^*_{\boldsymbol{\gamma}}$ or to its extreme-value copula. Accordingly, we see the probability measure $\nu_{0,k}$ as associated to the df of a normalized rv of componentwise maxima, computed over a number of underlying rv indexed by k, say n_k.

Exploiting Corollary 2.7, herein we specialise the above procedure to the estimation of an extreme-value copula via the empirical copula of sample maxima.

First, we recall some basic notions. Let $\boldsymbol{Z}^{(1:k)}$ be a sequence of iid copies of a rv $\boldsymbol{Z}$ with some copula C. Then, the empirical copula function $\widehat{C}_k$ is a map $T_k : \times_{i=1}^k \mathbb{R}^d \mapsto \ell^\infty([0,1]^d)$ defined by

$$\begin{aligned}
&\widehat{C}_k(\boldsymbol{u};\boldsymbol{Z}^{(1:k)}) := (T_k(\boldsymbol{Z}^{(1:k)}))(\boldsymbol{u})\\
&= \frac{1}{k}\sum_{i=1}^k \mathbb{1}\left(\frac{\sum_{l=1}^k \mathbb{1}(Z_1^{(l,k)} \leq Z_1^{(i,k)})}{k} \leq u_1, \dots, \frac{\sum_{l=1}^k \mathbb{1}(Z_d^{(l,k)} \leq Z_d^{(i,k)})}{k} \leq u_d\right),
\end{aligned}$$

for $\boldsymbol{u} \in [0,1]^d$, with $\mathbb{1}(E)$ denoting the indicator function of the event E.

Proposition 4.2. *Let $\boldsymbol{M}^{(n)} = (M_1^{(n)}, \dots, M_d^{(n)})$, C and G be as in Proposition 3.1, with C satisfying the assumptions of Corollary 2.7. Let $\boldsymbol{M}^{(n,1:k)} = (\boldsymbol{M}^{(n,1)}, \dots, \boldsymbol{M}^{(n,k)})$ be k independent copies of $\boldsymbol{M}^{(n)}$, with $n \equiv n_k \to_{k\to\infty} \infty$. Assume that $C^{(n)}$ and C_G satisfy*

$$\sup_{k \geq k_0} \rho_\delta(C^{(n)}, C_G) < \infty, \tag{30}$$

for some $\delta \in (0,1]$, $k_0 \in \mathbb{N}$, with ρ_δ as in (29). Then, almost surely

$$\sup_{\boldsymbol{u}\in[0,1]^d} \left|\widehat{C}_k(\boldsymbol{u}) - C_G(\boldsymbol{u})\right| \to_{k\to\infty} 0,$$

where $\widehat{C}_k \equiv \widehat{C}_k(\cdot;\boldsymbol{M}^{(n,1:k)})$.

For the proof of Proposition of 4.2 we establish the following remote contiguity relation.

Lemma 4.3. *Let $C^{(n,k)}$ and C_G^k denote the k-fold product measures pertaining to $C^{(n)}$ and C_G, respectively. Then, $k^2 C^{(n,k)} \lhd e^{ck} C_G^k$, for any $c > 0$.*

Proof. Let E_k, $k = 1, 2, \dots$ be a sequence of measurable events satisfying $C_G^k(E_k) = o(e^{-ck})$, for some $c > 0$. It is not difficult to see that, for any $\varepsilon > 0$,

$$C^{(n,k)}(E_k) \leq e^{\varepsilon k} C_G^k(E_k) + C^{(n,k)}(S_k > \varepsilon k),$$

where $S_k = \sum_{i=1}^k \log\left\{c^{(n)}(\boldsymbol{U}^{(n,i)})/c_G(\boldsymbol{U}^{(n,i)})\right\}$, $\boldsymbol{U}^{(n,i)}$, $1 \leq i \leq k$, are iid according to $C^{(n)}$, $c^{(n)}$ and c_G are the Lebesgue densities of $C^{(n)}$ and C_G, respectively. Choosing $\varepsilon < c$, the first term on the right-hand side is of order $o(e^{-(c-\varepsilon)k})$. As for the second term, as $k \to +\infty$ we have that $n \equiv n_k \to \infty$ and, by Corollary 2.7, $\varepsilon_k := \sup_{A \in \mathbb{B}^d \cap [0,1]^d} |C^{(n)}(A) - C_G(A)| = o(1)$. Thus, defining

$$\eta_{\alpha,k} := E\left[\log^\alpha \left\{\frac{c^{(n)}(\boldsymbol{U}^{(n,1)})}{c_G(\boldsymbol{U}^{(n,1)})}\right\}\right], \quad \alpha \in \mathbb{N},$$

under assumption (30), Theorem 6 in Wong and Shen (1995) guarantees that, as $k \to +\infty$, $\max(\eta_{1,k}, \eta_{2,k}) = O(\varepsilon_k \log^2(1/\varepsilon_k)) \leq \varepsilon/2$. Furthermore, simple analytical derivations lead to show that

$$\sup_{k \geq k_0} (-\eta_{3,k}) \leq 1 + \sup_{k \geq k_0} \eta_{4,k} \leq 2 + \log^4(K) + \sup_{k \geq k_0} \rho_\delta(C^{(n)}, C_G) < +\infty,$$

for some large but fixed $K > e^{1/\delta}$. Together with triangular and Markov inequalities, these facts entail that as $k \to +\infty$

$$\begin{aligned}
C^{(n,k)}(S_k > \varepsilon k) &\leq C^{(n,k)}(|S_k - k\eta_{1,k}| > \varepsilon/2k) \\
&\leq \left(\frac{2}{\varepsilon k}\right)^4 E\left[(S_k - k\eta_{1,k})^4\right] \\
&\leq \left(\frac{2}{\varepsilon}\right)^4 \left[\frac{1}{k^3}(\eta_{4,k} - 4\eta_{1,k}\eta_{3,k} + 6\eta_{1,k}^2\eta_{2,k}) + \frac{3}{k^2}(\eta_{2,k} - \eta_{1,k})^2\right] \\
&= o(k^{-2}),
\end{aligned}$$

where, in the third line, we exploit nonnegativity of $\eta_{1,k}$. The result now follows. $\square$

Proof of Proposition 4.2. Let $\boldsymbol{V}$ be a rv distributed according to the extreme-value copula C_G. Let $\boldsymbol{V}^{(1:k)} = (\boldsymbol{V}^{(1)}, \dots, \boldsymbol{V}^{(k)})$ be a sequence of iid copies of $\boldsymbol{V}$ with joint distribution $C_G^{(k)}$. Then, standard empirical process arguments (Gudendorf and Segers 2012, Deheuvels 1980, Wellner 1992) yield that, for any $\varepsilon > 0$,

$$C_G^{(k)}\left(\sup_{\boldsymbol{u}\in[0,1]^d}\left|\widehat{C}_k(\boldsymbol{u};\boldsymbol{V}^{(1:k)}) - C_G(\boldsymbol{u})\right| > \varepsilon\right) \\ \leq 2d\exp\left(-\frac{b_\varepsilon^2 k}{(d+1)^2}\right) + 16\frac{kb_\varepsilon^2}{(d+1)^2}\exp\left(-\frac{2b_\varepsilon^2 k}{(d+1)^2}\right)$$

for some $b_\varepsilon \in (0,\varepsilon)$. The term on the right hand side is of order $O(e^{-c_\varepsilon k})$, for some $c_\varepsilon > 0$. By Lemma 4.3, we have that $k^2 C^{(n,k)} \lhd e^{ck} C_G^{(k)}$ for all $c > 0$, where $C^{(n,k)}$ is the k-fold product measure corresponding to $C^{(n)}$. Let $\boldsymbol{U}^{(n,1:k)} = (\boldsymbol{U}^{(n,1)}, \dots, \boldsymbol{U}^{(n,k)})$, where

$$\boldsymbol{U}^{(n,i)} = \left(F_1\left(M_1^{(n,i)}\right)^n, \dots, F_d\left(M_d^{(n,i)}\right)^n\right), \quad i = 1, \dots, k.$$

The result now follows observing that $\boldsymbol{U}^{(n,1:k)}$ is distributed according to $C^{(n,k)}$ and that $\widehat{C}_k(\boldsymbol{u}) \equiv \widehat{C}_k(\boldsymbol{u};\boldsymbol{M}^{(n,1:k)}) = \widehat{C}_k(\boldsymbol{u};\boldsymbol{U}^{(n,1:k)})$. $\square$

Remark 4.4. Notice that the assumption in (30) of Proposition 4.2 is not overambitious. Indeed, when $C^{(n)}$ is obtained from copulas that are extreme-value copulas, the required condition is always satisfied. While, when $C^{(n)}$ is obtained from copulas that are in the domain of attraction of extreme-value copulas, analytically verifying (30) seems troublesome. Still, numerically checking whether some copula models meet this asumption can be fairly simple. For instance, consider the copula of Example 2.6, given in equation (26), and let c denote its density. Denote by $c^{(n)}$ the density of the copula $C^{(n)}$ pertaining to C and by c_G the density of the extreme-value copula model in (27). In this case, Corollary 2.7 applies and $C^{(n)}$ converges to C_G in variational distance. Figure 1 displays the plots of the densities c, c_G and $c^{(n)}$, with $n = 100$. Outside a neighborhood of the origin, pointwise convergence of $c^{(n)}$ to c_G turns out to be quite fast. In addition, the middle-right to bottom-right panels show that the density ratio $c^{(n)}/c_G$ is uniformly bounded by a finite constant, as the sample size n increases. Consequently, the condition in (30) is satisfied.

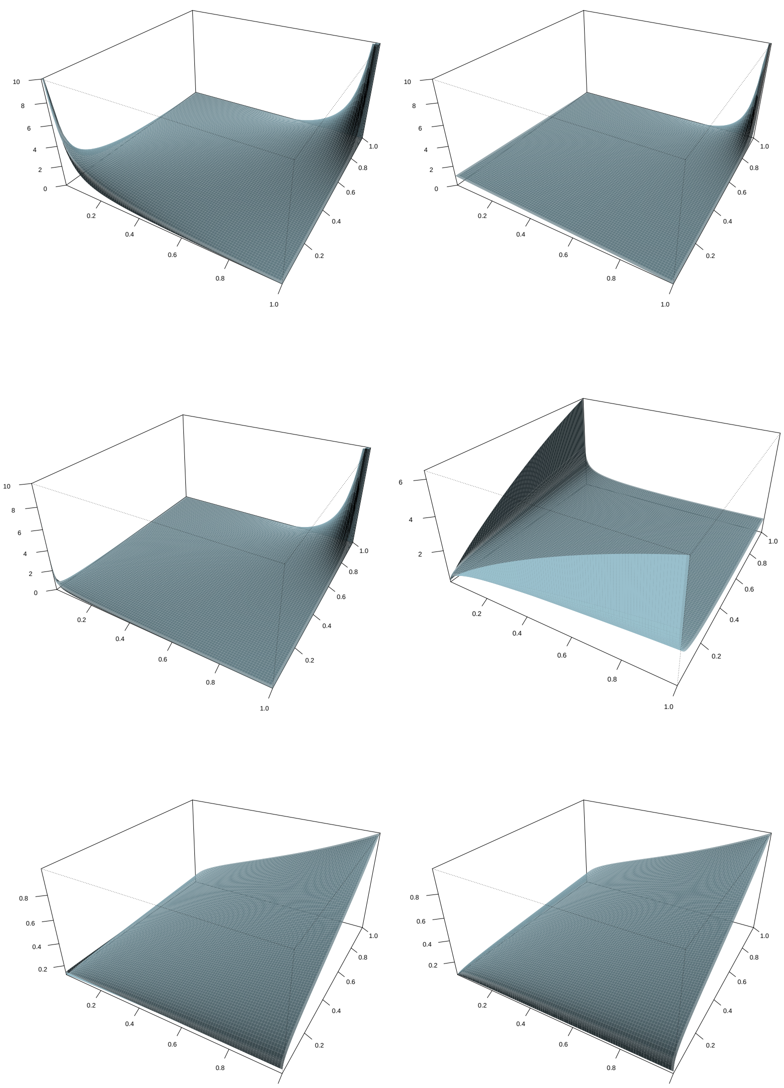

FIGURE 1. Top-left and -right panels display the densities c_G and c of the copula models in (27) and in (26), respectively. Middle-left panel shows the density $c^{(n)}$ of the copula $C^{(n)}$ pertaining to the copula model in (26), with sample size $n = 100$. Middle-right to bottom right panels depict the density ratio $c^{(n)}/c_G$, for $n = 2, 50, 100$, respectively.

4.2. **Bayesian approach.** A similar scheme is exploited by Padoan and Rizzelli (2019) in a Bayesian context, where extended Schwartz' theorem, e.g. Ghosal and van der Vaart (2017, Theorem 6.23), provides with exponential bounds for posterior concentration in a neighborhood of the true parameter. In particular, Padoan and Rizzelli (2019) consider a nonparametric Bayesian approach for estimating the D-norm $\|\cdot\|_D$ and the densities of the associated angular measure, see Falk (2019, pp. 25–29). Therein, Corollary 3.2 is leveraged to obtain a suitable remote contiguity result, allowing to extend almost-sure consistency of the proposed estimators from the case of data following a max-stable model, to the case of suitably normalised sample maxima, whose distribution lies in a variational neighbourhood of the latter.

Acknowledgements

The authors are indebted to the Associate Editor and two anonymous reviewers for their careful reading of the manuscript and their constructive remarks. Simone A. Padoan is supported by the Bocconi Institute for Data Science and Analytics (BIDSA).

References

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics. Chichester, UK: Wiley.

Berghaus, B. and A. Bücher (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. Ann. Statist. 46(5), 2307–2335.

Berghaus, B., A. Bücher, and H. Dette (2013). Minimum distance estimators of the Pickands dependence function and related tests of multivariate extreme-value dependence. J. SFdS 154(1), 116–137.

Bücher, A. and J. Segers (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. Extremes 17(3), 495–528.

Bücher, A. and J. Segers (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. Bernoulli 24(2), 1427–1462.

Bücher, A., S. Volgushev, and N. Zou (2019). On second order conditions in the multivariate block maxima and peak over threshold method. Journal of Multivariate Analysis 173, 604 – 619.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values, Volume 208. Springer.

de Haan, L. and L. Peng (1997). Rates of convergence for bivariate extremes. Journal of Multivariate Analysis 61(2), 195 – 230.

Deheuvels, P. (1980). Non parametric tests of independence. In J.-P. Raoult (Ed.), Statistique non Paramétrique Asymptotique, Berlin, Heidelberg, pp. 95–107. Springer Berlin Heidelberg.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. Bernoulli 21(1), 420–436.

Dombry, C., S. Engelke, and M. Oesting (2017). Bayesian inference for multivariate extreme value distributions. Electronic Journal of Statistics 11, 4813–4844.

Falk, M. (2019). Multivariate Extreme Value Theory and D-Norms. New York: Springer International Publishing.

Falk, M., J. Hüsler, and R.-D. Reiss (2011). Laws of Small Numbers: Extremes and Rare Events (3 ed.). Basel: Birkhäuser.

Falk, M., S. A. Padoan, and F. Wisheckel (2019). Generalized Pareto copulas: a key to multivariate extremes. Journal of Multivariate Analysis 174, 104538.

Ferreira, A. and L. de Haan (2015, 02). On the block maxima method in extreme value theory: Pwm estimators. Ann. Statist. 43(1), 276–298.

Ghosal, S. and A. van der Vaart (2017). Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Gudendorf, G. and J. Segers (2012). Nonparametric estimation of multivariate extreme-value copulas. Journal of Statistical Planning and Inference 142(12), 3073–3085.

Kaufmann, E. and R.-D. Reiss (1993). Strong convergence of multivariate point processes of exceedances. Ann. Inst. Statist. Math. 45(3), 433–444.

Kleijn, B. (2017). On the frequentist validity of Bayesian limits. arXiv e-prints, arXiv:1611.08444v3.

Marcon, G., S. A. Padoan, P. Naveau, P. Muliere, and J. Segers (2017). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. Journal of Statistical Planning and Inference 183, 1–17.

McNeil, A. J. and J. Nešlehová (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. Ann. Statist. 37(5B), 3059–3097.

Mhalla, L., V. Chavez-Demoulin, and P. Naveau (2017). Non-linear models for extremal dependence. Journal of Multivariate Analysis 159, 49–66.

Padoan, S. A. and S. Rizzelli (2019). Strong consistency of nonparametric Bayesian inferential methods for multivariate max-stable distributions. arXiv e-prints, arXiv:1904.00245v2.

Reiss, R.-D. (1989). Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics. Springer Series in Statistics. New York: Springer.

Resnick, S. I. (2008). Extreme Values, Regular Variation, and Point Processes. Springer Series in Operations Research and Financial Engineering. New York: Springer.

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

Sklar, A. (1996). Random variables, distribution functions, and copulas – a personal look backward and forward. In L. Rüschendorf, B. Schweizer, and M. D. Taylor (Eds.), Distributions with fixed marginals and related topics, Volume 28 of Lecture Notes – Monograph Series, Hayward, CA, pp. 1–14. Institute of Mathematical Statistics.

Wellner, J. A. (1992). Empirical processes in action: A review. International Statistical Review / Revue Internationale de Statistique 60(3), 247–269.

Wong, W. H. and X. Shen (1995). Probability inequalities for likelihood ratios and convergence rates of sieve mles. Ann. Statist. 23(2), 339–362.