J. Appl. Prob. **57,** 314–331 (2020)
*doi:*10.1017/jpr.2019.100

STRONG CONVERGENCE OF MULTIVARIATE MAXIMA

MICHAEL FALK,* *University of Würzburg*
SIMONE A. PADOAN,** *Bocconi University of Milan*
STEFANO RIZZELLI,*** *École Polytechnique Fédérale de Lausanne*

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 generalised 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.

Keywords: maxima; strong convergence; total variation; copula; generalised Pareto copula; D-norm; multivariate max-stable distribution; domain of attraction

2010 Mathematics Subject Classification: Primary 60G70
Secondary 62H10; 60F15

1. Introduction

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

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

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

$$\begin{aligned} \mathrm{P}\Big(n\Big(\max_{1\le i\le n} U^{(i)} - 1\Big) \le x\Big) &= \mathrm{P}\left(U_i \le 1 + \frac{x}{n},\ 1 \le i \le n\right) \\ &= V^n\left(1 + \frac{x}{n}\right) \end{aligned}$$

Received 20 July 2019; revision received 3 December 2019.
* Postal address: Chair of Mathematics VIII, Emil-Fischer-Str. 30, 97074 Würzburg, Germany. Email address: michael.falk@uni-wuerzburg.de
** Postal address: Department of Decision Sciences, via Roentgen, 1 20136 Milan, Italy. Email address: simone.padoan@unibocconi.it
*** Postal address: EPFL-SB-MATH-STAT, MA B1 507, Station 8, 1015 Lausanne, Switzerland. Email address: stefano.rizzelli@epfl.ch

$$= \left(1 + \frac{x}{n}\right)^n$$

$$\to_{n\to\infty} G(x), \tag{2}$$

where

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

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

$$n\big(M^{(n)} - 1\big) \to_D \eta,$$

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 random variable η has distribution function G in (3).

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

$$v_n(x) := \frac{\partial}{\partial x}\left(V^n\left(1 + \frac{x}{n}\right)\right) = V^{n-1}\left(1 + \frac{x}{n}\right) v\left(1 + \frac{x}{n}\right)$$

$$\to_{n\to\infty} g(x) := G'(x) = \begin{cases} \exp(x), & x \le 0, \\ 0, & x > 0, \end{cases}$$

i.e. we have pointwise convergence of the sequence of densities of the normalised maximum $n\left(M^{(n)} - 1\right)$, $n \in \mathbb{N}$, to that of η. Scheffé's lemma (see, e.g., [24, Lemma 3.3.3]) now implies convergence in total variation:

$$\sup_{A\in\mathbb{B}} \left|\mathrm{P}\big(n\big(M^{(n)} - 1\big) \in A\big) - \mathrm{P}(\eta \in A)\right| \to_{n\to\infty} 0, \tag{4}$$

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

Now let X be a random variable with *arbitrary* distribution function F and $F^{-1}(q) := \{t \in \mathbb{R} : F(t) \ge q\}$, with $q \in (0, 1)$, be the usual *quantile function* or *generalised inverse* of F. Then, we can assume the representation

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

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

$$X^{(i)} = F^{-1}\big(U^{(i)}\big), \qquad i = 1, 2, \ldots$$

The fact that each quantile function is a nondecreasing function yields

$$\max_{1\le i\le n} X^{(i)} = \max_{1\le i\le n} F^{-1}\big(U^{(i)}\big) = F^{-1}\left(\max_{1\le i\le n} U^{(i)}\right)$$

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

The strong convergence in equation (4) now implies the following convergence in total variation:

$$\sup_{A\in\mathbb{B}} \left|\mathrm{P}\left(\max_{1\le i\le n} X^{(i)} \in A\right) - \mathrm{P}\left(F^{-1}\left(1 + \frac{1}{n}\eta\right) \in A\right)\right| \to_{n\to\infty} 0. \tag{5}$$

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

Finally, assume that F is a continuous distribution function 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 distribution function G^*_γ, e.g. [14, p. 21]. This means, for $n \in \mathbb{N}$, there are norming constants $a_n > 0$, $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) \to_{n\to\infty} \exp\Big(-(1+\gamma x)_+^{-1/\gamma}\Big) =: G^*_\gamma(x), \tag{6}$$

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 [14] for details. In this general case we also have locally uniform convergence at the density level, i.e.

$$f^{(n)}(x) := \frac{\partial}{\partial x} F^n(a_n x + b_n) \to_{n\to\infty} \frac{\partial}{\partial x} G^*_\gamma(x) =: g^*_\gamma(x), \tag{7}$$

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

$$\lim_{x\to\infty} \frac{xf(x)}{1-F(x)} = 1/\gamma, \qquad \text{if } \gamma > 0, \tag{8}$$

$$\lim_{x\uparrow x_0} \frac{(x_0 - x)f(x)}{1-F(x)} = -1/\gamma, \qquad \text{if } \gamma < 0, \tag{9}$$

$$\lim_{x\uparrow x_0} \frac{f(x)}{(1-F(x))^2} \int_x^{x_0} (1-F(t))\,\mathrm{d}t = 1, \qquad \text{if } \gamma = 0, \tag{10}$$

see, e.g., Proposition 2.5 in [25]. In particular, if (7) holds true, Scheffé's lemma entails that

$$\sup_{A\in\mathbb{B}} \left| \mathrm{P}\left(a_n^{-1}\left(\max_{1\le i\le n} X^{(i)} - b_n \right) \in A \right) - \mathrm{P}\,(Y \in A) \right| \to_{n\to\infty} 0, \tag{11}$$

where Y is a random variable with distribution G^*_γ and $X^{(i)}$, $i = 1, \ldots, n$, are independent copies of a random variable X with distribution F.

In this paper we extend the results in (4), (5), and (11) to higher dimensions. First, in Section 2 we consider copulas. In Theorem 2.2, we demonstrate that the strong convergence result holds for copulas that are in a differential neighbourhood of a multivariate generalised Pareto copula [12, 14]. As a result of this, we also establish strong convergence of the copula of the maximum of n i.i.d. random vectors with arbitrary common distribution function to the limiting extreme-value copula (Corollary 2.1). Sklar's theorem is then used in Section 3 to derive convergence in variational distance of the maximum of n i.i.d. random vectors with arbitrary common distribution function and, under restrictions (8)–(10) 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 i.i.d. random variables has been well investigated; see, e.g., Section 5.1 in [24] and the literature cited therein. Strong convergence holds in particular under suitable von Mises type conditions on the underlying distribution function; see (8)–(11) 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 [13]. It is proven in [18] 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 (16) and (23)) under which strong convergence of multivariate maxima and their normalised versions actually holds. Furthermore, our strong convergence results for sample maxima are valid for maxima with arbitrary dimensions, unlike those in [8], 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 distribution function 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 $\boldsymbol{U} = (U_1, \ldots, U_d)$ follows a *copula*, say C, on $\mathbb{R}^d$, i.e. each component U_j has the distribution function V_j given in (1). Let $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \ldots$ be independent copies of $\boldsymbol{U}$ and put, for $n \in \mathbb{N}$,

$$\boldsymbol{M}^{(n)} := \left(M_1^{(n)}, \ldots, M_d^{(n)}\right) := \left(\max_{1 \le i \le n} U_1^{(i)}, \ldots, \max_{1 \le i \le n} U_d^{(i)}\right). \tag{12}$$

In the following, the operations involving vectors are meant componentwise; furthermore, we set $\boldsymbol{0} = (0, \ldots, 0)$, $\boldsymbol{1} = (1 \ldots, 1)$, and $\boldsymbol{\infty} = (\infty, \ldots, \infty)$. Finally, we denote the copula of the random vector in (12) by $C^{(n)}(\boldsymbol{u}) := C^n(\boldsymbol{u}^{1/n})$, $\boldsymbol{u} \in [0, 1]^d$.

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 distribution function G on $\mathbb{R}^d$ such that, for $\boldsymbol{x} = (x_1, \ldots, x_d) \le \boldsymbol{0} \in \mathbb{R}^d$,

$$\begin{aligned} \mathrm{P}\left(n\left(\boldsymbol{M}^{(n)} - \boldsymbol{1}\right) \le \boldsymbol{x}\right) &= \mathrm{P}\left(n\left(M_1^{(n)} - 1\right) \le x_1, \ldots, n\left(M_d^{(n)} - 1\right) \le x_d\right) \\ &\to_{n\to\infty} G(\boldsymbol{x}). \end{aligned} \tag{13}$$

Then, G is necessarily a *multivariate max-stable* or *multivariate extreme-value* distribution function, with *extreme-value copula* C_G and standard negative exponential margins G_j, $j = 1, \ldots, d$; see (3). In the following we refer to the distribution function G in (13) as the *standard* multivariate max-stable distribution function. Precisely, the form of G is

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

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

$$C_G(\boldsymbol{u}) = \exp\left(-\|\log u_1, \ldots, \log u_d\|_D\right), \qquad \boldsymbol{u} \in [0, 1]^d, \tag{14}$$

while the margins G_j, $j = 1, \ldots, d$, are as in (3). Therefore, the distribution in (13) has the representation

$$G(\boldsymbol{x}) = \exp\left(-\|\boldsymbol{x}\|_D\right), \qquad \boldsymbol{x} \le \boldsymbol{0} \in \mathbb{R}^d. \tag{15}$$

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

The convergence result in (13) implies that $C^{(n)}(\boldsymbol{u}) \to_{n\to\infty} C_G(\boldsymbol{u})$, for all $\boldsymbol{u} \in [0, 1]^d$; see, e.g., [12, 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 [12], there exists a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ with $Z_j \geq 0$, $\mathrm{E}(Z_j) = 1$, $1 \leq j \leq d$, such that

$$\|\boldsymbol{x}\|_D = \mathrm{E}\left(\max_{1\leq j\leq d}\left(|x_j| Z_j\right)\right), \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

Examples of D-norms are the sup-norm $\|\boldsymbol{x}\|_\infty = \max_{1\leq j\leq d}|x_j|$, or the complete family of logistic norms $\|\boldsymbol{x}\|_p = \left(\sum_{j=1}^d |x_j|^p\right)^{1/p}$, $p \geq 1$. For a recent account on multivariate extreme-value theory and D-norms we refer to [12]. In particular, Proposition 3.1.5 in [12] implies that the convergence result in (13) is also equivalent to the expansion

$$C(\boldsymbol{u}) = 1 - \|\boldsymbol{1} - \boldsymbol{u}\|_D + o(\|\boldsymbol{1} - \boldsymbol{u}\|) \tag{16}$$

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 $o(\|\boldsymbol{1} - \boldsymbol{u}\|)$ in expansion (16) and require that there exists $\boldsymbol{u}_0 \in (0, 1)^d$ such that

$$C(\boldsymbol{u}) = 1 - \|\boldsymbol{1} - \boldsymbol{u}\|_D, \qquad \boldsymbol{u} \in [\boldsymbol{u}_0, \boldsymbol{1}] \subset \mathbb{R}^d. \tag{17}$$

A copula that satisfies the above expansion is a *generalised Pareto copula* (GPC). The significance of GPCs for multivariate extreme-value theory is explained in [14] and in [12, Section 3.1].

Note that

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

defines a multivariate distribution function only in dimension $d = 2$; see, e.g., [20, Examples 2.1, 2.2]. But one can find, for arbitrary dimension $d \geq 2$, a random vector whose distribution function satisfies (17); see, e.g., [12, Equation (2.15)]. For this reason, we require the condition in (17) only on some upper interval $[\boldsymbol{u}_0, \boldsymbol{1}] \subset \mathbb{R}^d$.

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

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

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

$$f^{(n)}(\boldsymbol{x}) := \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F^{(n)}(\boldsymbol{x}) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d}\left(1 - \frac{1}{n}\|\boldsymbol{x}\|_D\right)^n. \tag{18}$$

As for the standard multivariate max-stable distribution function G in (15), its density exists and is given by

$$g(\boldsymbol{x}) := \frac{\partial^d}{\partial x_1 \cdots \partial x_d} G(\boldsymbol{x}) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} \exp\left(-\|\boldsymbol{x}\|_D\right), \qquad \boldsymbol{x} \leq \boldsymbol{0} \in \mathbb{R}^d. \tag{19}$$

We are now ready to state our first multivariate extension of the convergence in total variation in (4). For brevity, we occasionally denote with the same letter a Borel measure and its distribution function.

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

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

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

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

Remark 2.1. Note that we can write a GPC

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

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

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

The *generator function* $\varphi : (0, 1] \to [0, \infty)$ is in general strictly decreasing and convex, with $\varphi(1) = 0$; see, e.g., [20]. 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 2.1. *Choose $\varepsilon \in (0, 1)$ and $\boldsymbol{x}_\varepsilon < \mathbf{0} \in \mathbb{R}^d$ with $G([\boldsymbol{x}_\varepsilon, \mathbf{0}]) \geq 1 - \varepsilon$. Then we have, for $\boldsymbol{x} \in [\boldsymbol{x}_\varepsilon, \mathbf{0}]$,*

$$f^{(n)}(\boldsymbol{x}) \to_{n\to\infty} g(\boldsymbol{x}). \tag{20}$$

Proof. $G(\boldsymbol{x})$ can be seen as the function composition $(\ell \circ \phi)(\boldsymbol{x})$, where we set $\ell(y) = \exp(y)$ and $\phi(\boldsymbol{x}) = -\|\boldsymbol{x}\|_D$. Then, by *Faá di Bruno's formula*, the density in (19) is equal to

$$g(\boldsymbol{x}) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} \exp(\phi(\boldsymbol{x})) = G(\boldsymbol{x}) \sum_{\mathcal{P} \in \mathscr{P}} \prod_{B \in \mathcal{P}} \frac{\partial^{|B|}\phi(\boldsymbol{x})}{\partial^B \boldsymbol{x}}, \tag{21}$$

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

Similarly, $F^{(n)}(\boldsymbol{x})$ can be seen as the function composition $(\ell \circ \phi_n)(\boldsymbol{x})$, where we set $\phi_n(\boldsymbol{x}) := -n \log(1/(1 - n^{-1}\|\boldsymbol{x}\|_D))$. Then, $F^{(n)}(\boldsymbol{x}) = \exp(\phi_n(\boldsymbol{x}))$ and, once again by Faá di Bruno's formula, the density in (18) is equal to

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

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

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 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 $B = (i_1, \ldots, i_k)$ and the product is over all blocks b of a 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 (20) 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

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

which implies

$$\left| \mathrm{P}\big(n\big(\boldsymbol{M}^{(n)} - \boldsymbol{1}\big) \in [\boldsymbol{x}_\varepsilon, \boldsymbol{0}]\big) - G([\boldsymbol{x}_\varepsilon, \boldsymbol{0}]) \right| \to_{n\to\infty} 0 \tag{22}$$

and, thus,

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

or

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

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

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

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

Now, from equation (22) 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 (20), 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

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

now implies the assertion of Theorem 2.1. □

Next, we extend Theorem 2.1 to a copula C which is in a *differentiable neighbourhood* of a GPC, defined next. Suppose that C satisfies the expansion (16), 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 $B=(i_1, \ldots, i_k)$ of $\{1, \ldots, d\}$,

$$\frac{\partial^k}{\partial x_{i_1}, \ldots, \partial x_{i_k}} n\left(C\left(\boldsymbol{1}+\frac{\boldsymbol{x}}{n}\right)-1\right) \to_{n\to\infty} \frac{\partial^k}{\partial x_{i_1}, \ldots, \partial x_{i_k}} \phi(\boldsymbol{x}), \tag{23}$$

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

Theorem 2.2. *Suppose the copula C satisfies the conditions in (16) and (23). Then we obtain*

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

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

Proof. The proof of Theorem 2.2 is similar to that of Theorem 2.1, but this time we resort to a variant of Lemma 2.1 as follows. Note that for $n \in \mathbb{N}$,

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

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 2.1. Then, in 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 (16) and (23) we obtain that, for each block b of a partition $\mathcal{P}_B \in \mathscr{P}_B$,

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

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

Therefore, as in Lemma 2.1, we obtain

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

and the result follows. □

Example 2.1. 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

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

as $\boldsymbol{u} \in (0, 1]^d$ converges to $\boldsymbol{1} \in \mathbb{R}^d$, i.e. condition (16) 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 (23). 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 2.1 and $\varphi_n(\boldsymbol{x}) := \log\left(C_p(\boldsymbol{1}+\boldsymbol{x}/n)\right)$. 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 Faá di Bruno's formula to the partial derivatives of $n(\ell \circ \varphi_n(x) - 1)$ and noting that, on one hand, $C_p(\boldsymbol{1}+\boldsymbol{x}/n) \to_{n\to\infty} 1$, and on the other hand,

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

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

Example 2.2. Consider the copula

$$C(\boldsymbol{u}) = 1 - d + \sum_{j=1}^d u_j + \sum_{2\leq i\leq d}\left((-1)^i \sum_{\substack{B\subseteq\{1,\ldots,d\}\\|B|=i}} \left(\sum_{j\in B} \frac{1}{1-u_j} - d + 1\right)^{-1}\right). \tag{24}$$

This provides the d-dimensional version (with $d \geq 2$) of the 2-dimensional copula associated to the distribution function discussed in [25, 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

$$C_G(\boldsymbol{u}) = \exp\left(\sum_{j=1}^d \log u_j + \sum_{2\leq i\leq d}\left((-1)^{i+1} \sum_{\substack{B\subseteq\{1,\ldots,d\}\\|B|=i}} \left(\sum_{j\in B} \frac{1}{\log u_j} - d + 1\right)^{-1}\right)\right). \tag{25}$$

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

Then, by Proposition 3.1.5 and Corollary 3.1.12 in [12], the copula in (24) satisfies condition (16) with D-norm

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

The copula in (24) also complies with the conditions in (23). 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 expression on the right-hand side of the above formula 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 expression on the right-hand side of the above formula 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 (12). 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 2.2 is the convergence of $C^{(n)}$ to C_G in variational distance.

Corollary 2.1. *Assume C satisfies conditions (16) and (23), 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

$$\widetilde{C}^{(n)}(\boldsymbol{u}) := \mathrm{P}\big(n\big(\boldsymbol{M}^{(n)} - \mathbf{1}\big) \le \log \boldsymbol{u}\big) = C^n(1 + \log \boldsymbol{u}/n).$$

By Theorem 2.2, $\widetilde{C}^{(n)}$ converges to C_G in variational distance. Now, for some $\varepsilon \in (0, 1)$, set

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

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

In particular, fix $\varepsilon > 0$ such that $C_G\big(\mathcal{U}_\varepsilon^{\complement}\big) > 1 - \varepsilon_0$, 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 remainder 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

$$\sup_{A\in\mathbb{B}^d\cap[0,1]^d} \left|C^{(n)}(A) - C_G(A)\right| \le \sup_{A\in\mathbb{B}^d\cap\mathcal{U}_\varepsilon^{\complement}} \left|C_\varepsilon^{(n)}(A) - \widetilde{C}_\varepsilon^{(n)}(A)\right| + O(\varepsilon_0), \tag{26}$$

where $C_\varepsilon^{(n)}$ and $\widetilde{C}_\varepsilon^{(n)}$ are the normalised versions $C_\varepsilon^{(n)} = C^{(n)}/C^{(n)}\big(\mathcal{U}_\varepsilon^{\complement}\big)$ and $\widetilde{C}_\varepsilon^{(n)} = \widetilde{C}^{(n)}/\widetilde{C}^{(n)}\big(\mathcal{U}_\varepsilon^{\complement}\big)$. Finally, denote their densities by $c_\epsilon^{(n)}$ and $\widetilde{c}_\varepsilon^{(n)}$, respectively. Then the supremum on the right-hand side in (26) 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 2.2 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. □

3. The general case

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a random vector with arbitrary distribution function F. By Sklar's theorem [26, 27] we can assume the representation

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

where F_j is the distribution function of X_j, $j = 1, \ldots, d$, and $\boldsymbol{U} = (U_1, \ldots, U_d)$ follows a copula, say C, corresponding to F.

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

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

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

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

Theorem 2.1 now implies the following result.

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

Proposition 3.1. *Let $\boldsymbol{\eta}=(\eta_1, \ldots, \eta_d)$ be a random vector with standard multivariate max-stable distribution function $G(\boldsymbol{x})=\exp(-\|\boldsymbol{x}\|_D)$, $\boldsymbol{x}\le\boldsymbol{0}\in\mathbb{R}^d$. Let $\boldsymbol{X}$ be a random vector 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 (16) and (23). Then,*

$$\sup_{A\in\mathbb{B}^d}\left|\mathrm{P}\big(\boldsymbol{M}^{(n)}\in A\big)-\mathrm{P}\left(\left(F_1^{-1}\left(1+\frac{1}{n}\eta_1\right), \ldots, F_d^{-1}\left(1+\frac{1}{n}\eta_d\right)\right)\in A\right)\right|\to_{n\to\infty}0.$$

Finally, we generalise the result in Proposition 3.1 to the case where the random vector 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 distribution function $G^*_{\boldsymbol{\gamma}}$, with tail indices $\boldsymbol{\gamma}=(\gamma_1, \ldots, \gamma_d)$, e.g. [14, Chapter 4]. This means that there exist sequences of norming vectors $\boldsymbol{a}_n=\big(a_n^{(1)}, \ldots, a_n^{(d)}\big)>\boldsymbol{0}$ and $\boldsymbol{b}_n=\big(b_n^{(1)}, \ldots, b_n^{(d)}\big)\in\mathbb{R}^d$, 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 random vector with distribution $G^*_{\boldsymbol{\gamma}}$. The copula of $G^*_{\boldsymbol{\gamma}}$ is the extreme-value copula in (14), and its margins $G^*_{\gamma_j}$, $j=1, \ldots, d$, are members of the generalised extreme-value family of distribution functions in (6).

To attain convergence in variational distance, we combine Proposition 3.1, obtained under conditions involving only dependence structures, with univariate von Mises conditions on the margins $F_1, \ldots, F_d$, see (7)–(10). We denote the vector of endpoints by $\boldsymbol{x}_0:=(x_0^{(1)}, \ldots, x_0^{(d)})$, where $x_0^{(j)}:=\sup\{x\in\mathbb{R}: F_j(x)<1\}, j=1, \ldots, d$.

Corollary 3.1. *Let $\boldsymbol{Y}$ and $\boldsymbol{X}$ be random vectors with a generalised multivariate max-stable distribution function $G^*_{\boldsymbol{\gamma}}$ and a continuous distribution function 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 jth margin F_j of F satisfies one of the conditions (8)–(10) with f', γ, and x_0 replaced by f_j', γ_j, and $x_0^{(j)}$. Then,*

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

Proof. Let $\boldsymbol{\eta}=(\eta_1, \ldots, \eta_d)$ be a random vector 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

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

where

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

and

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

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

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 Scheffé's lemma. First, notice that $G^*_{\boldsymbol{\gamma}}$ and G have the same extreme-value copula. Thus, from (14) it follows that, for $\boldsymbol{x} \in \mathbb{R}^d$, $G^*_{\boldsymbol{\gamma}}(\boldsymbol{x}) = G(\boldsymbol{u}(\boldsymbol{x}))$, where $\boldsymbol{u}(\boldsymbol{x}) = \big(u^{(1)}(x_1), \dots, u^{(d)}(x_d)\big)$ with $u^{(j)}(x_j) = \log G^*_{\gamma_j}(x_j)$ for $j = 1, \dots, d$. Now, define $Q^{(n)}(\boldsymbol{x}) := \mathrm{P}(\boldsymbol{Y}_n \le \boldsymbol{x}) = G(\boldsymbol{u}_n(\boldsymbol{x}))$, for $\boldsymbol{x} \in \mathbb{R}^d$, where $\boldsymbol{u}_n(\boldsymbol{x}) = \big(u_n^{(1)}(x_1), \dots, u_n^{(d)}(x_d)\big)$ with

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

Consequently, as $n \to \infty$,

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

where g is as in (21) and $g^*_{\gamma_j}(x) = (\partial/\partial x) G^*_{\gamma_j}(x)$, $1 \le j \le d$. In particular, the second line follows from the continuity of g and Proposition 2.5 in [25]. The proof is now complete. □

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. [7, Chapter 8], [1, Chapter 9], [21, 22], to name a few. Parametric and nonparametric inferential procedures have been proposed for fitting max-stable models to the data; see, e.g., [3, 11, 17, 21]. 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 [2, 5, 10, 15] account for such model misspecification in the univariate setting. In the multivariate case, in [4] weak convergence and consistency in probability of empirical copulas have been studied under suitable second-order conditions; see also [6]. This is the only multivariate contribution focusing on the problem of convergences under model misspecification, as far as we know. In the following we illustrate how our variational convergence results, obtained under conditions (16) and (23), allow us to establish a stronger form of consistency, for both frequentist and Bayesian procedures. To do that, we resort to the notion of remote contiguity introduced in [19].

Definition 4.1. Let r_k, s_k, $k \in \mathbb{N}$, 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 to be 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 random vector with a distribution function F pertaining to

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

a probability measure μ_0 on $\mathbb{B}^d$. Denote by μ_k the corresponding k-fold product measure. Let $\boldsymbol{Y}^{(1:k)} = \big(\boldsymbol{Y}^{(1,k)}, \ldots, \boldsymbol{Y}^{(k,k)}\big)$ be a sequence of k i.i.d. copies of $\boldsymbol{Y}$. Consider a measurable map $T_k : \times_{i=1}^{k} \mathbb{R}^d \to \Theta$ and let

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

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 $\mu_k\big(\mathscr{D}(\widehat{\theta}_k, \theta) > \varepsilon\big) = o(\mathrm{e}^{-c_\varepsilon k})$ and $k^{1+c'_\varepsilon} \nu_k \lhd \mathrm{e}^{c_\varepsilon k} \mu_k$, then we can conclude by the Borel–Cantelli lemma that

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

where $\boldsymbol{Z}^{(1:k)} = \big(\boldsymbol{Z}^{(1,k)}, \ldots, \boldsymbol{Z}^{(k,k)}\big)$ is a sequence of i.i.d. random vectors 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 $\sup_{A\in\mathbb{B}^d} |\nu_{0,k}(A) - \mu_0(A)| \to_{k\to\infty} 0$, and μ_0 and $\nu_{0,k}$ have the same support and continuous Lebesgue densities, $p_{0,k}$ and m_0, satisfying

$$\sup_{k\geq k_0} \rho_\delta(\nu_{0,k}, \mu_0) := \sup_{k\geq k_0} \int_{\mathcal{X}_{\delta,k}} \big(p_{0,k}(\boldsymbol{x})/m_0(\boldsymbol{x})\big)^\delta p_{0,k}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} < \infty \tag{27}$$

for some $\delta \in (0, 1]$ and $k_0 \in \mathbb{N}$, where $\mathcal{X}_{\delta,k} = \{\boldsymbol{x} \in \mathbb{R}^d : p_{0,k}(\boldsymbol{x})/m_0(\boldsymbol{x}) > \mathrm{e}^{1/\delta}\}$. Essentially, variational convergence and (27) 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 distribution function $G^*_{\boldsymbol{\gamma}}$ or to its extreme-value copula. Accordingly, we see the probability measure $\nu_{0,k}$ as associated to the dostribution function of a normalized random vector of componentwise maxima, computed over a number of underlying random variables indexed by k, say n_k.

Exploiting Corollary 2.1, 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 i.i.d. copies of a random vector $\boldsymbol{Z}$ 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

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

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

Proposition 4.1. *Let $\boldsymbol{M}^{(n)} = \big(M_1^{(n)}, \ldots, M_d^{(n)}\big)$, and C and G be as in Proposition 3.1, with C satisfying the assumptions of Corollary 2.1. Let $\boldsymbol{M}^{(n,1:k)} = \big(\boldsymbol{M}^{(n,1)}, \ldots, \boldsymbol{M}^{(n,k)}\big)$ 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*

$$\sup_{k\geq k_0} \rho_\delta\left(C^{(n)}, C_G\right) < \infty \tag{28}$$

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

for some $\delta \in (0, 1]$ *and* $k_0 \in \mathbb{N}$, *with* ρ_δ *as in (27). 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 $\widehat{C}_k \equiv \widehat{C}_k\big(\,\cdot\,; \boldsymbol{M}^{(n,1:k)}\big)$.

For the proof of Proposition 4.1 we establish the following remote contiguity relation.

Lemma 4.1. *Let* $C^{(n,k)}$ *and* C_G^k *denote the k-fold product measures pertaining to* $C^{(n)}$ *and* C_G, *respectively. Then* $k^2 C^{(n,k)} \lhd \mathrm{e}^{ck} C_G^k$, *for any* $c > 0$.

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

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

where $S_k = \sum_{i=1}^k \log\{c^{(n)}(\boldsymbol{U}^{(n,i)})/c_G(\boldsymbol{U}^{(n,i)})\}$, $\boldsymbol{U}^{(n,i)},\ 1 \le i \le k$, are i.i.d. according to $C^{(n)}$, and $c^{(n)}$ 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 $o\big(\mathrm{e}^{-(c-\varepsilon)k}\big)$. As for the second term, as $k \to +\infty$ we have that $n \equiv n_k \to \infty$ and, by Corollary 2.1, $\varepsilon_k := \sup_{A\in\mathbb{B}^d\cap[0,1]^d} |C^{(n)}(A) - C_G(A)| = o(1)$. Thus, defining

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

under assumption (28), Theorem 6 in [28] guarantees that, as $k \to +\infty$, $\max_{i=1,2} \eta_{i,k} = O(\varepsilon_k \log^2(1/\varepsilon_k)) \le \varepsilon/2$. Furthermore, simple analytical derivations lead to

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

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

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

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

Proof of Proposition 4.1. Let $\boldsymbol{V}$ be a random vector distributed according to the extreme-value copula C_G. Let $\boldsymbol{V}^{(1:k)} = \big(\boldsymbol{V}^{(1)}, \ldots, \boldsymbol{V}^{(k)}\big)$ be a sequence of i.i.d. copies of $\boldsymbol{V}$ with joint

https://doi.org/10.1017/jpr.2019.100 Published online by Cambridge University Press

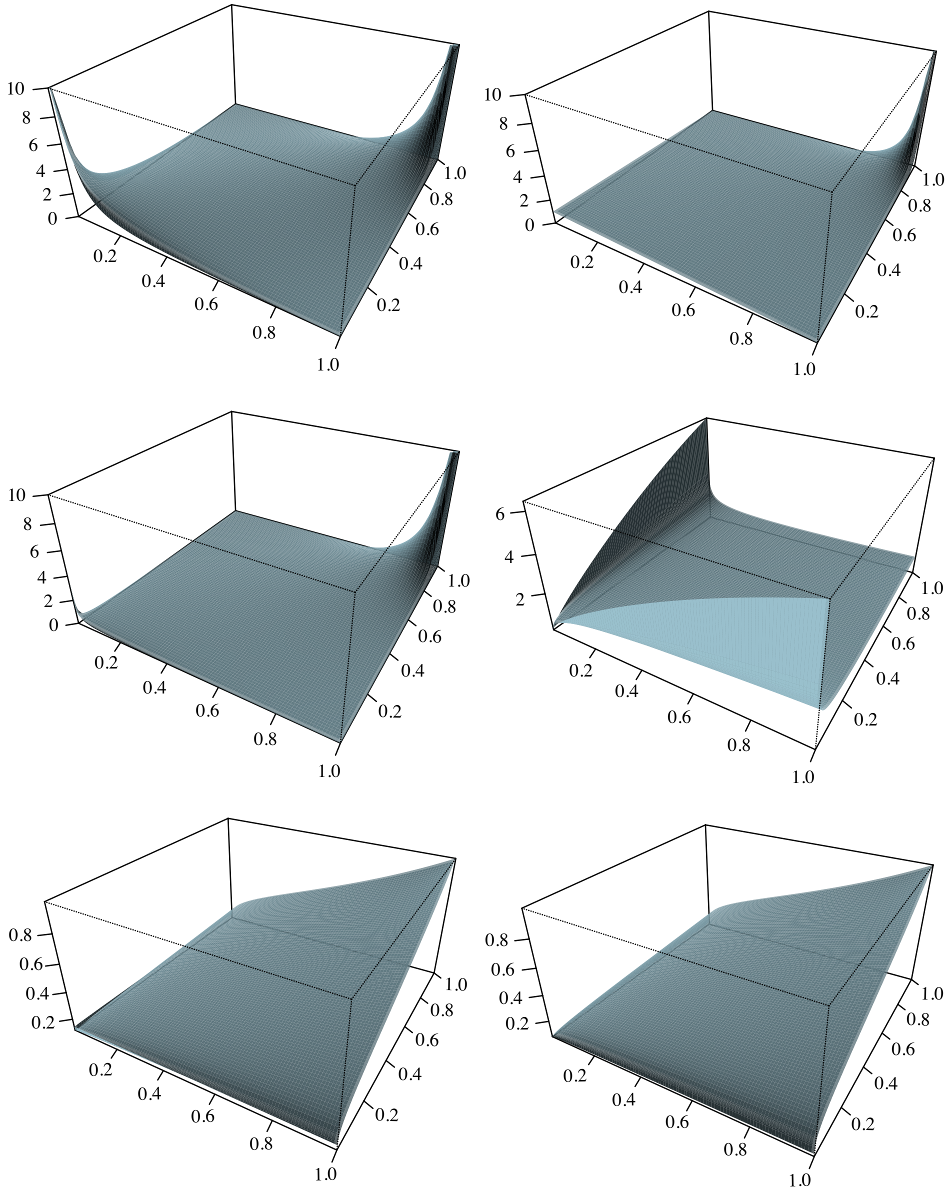

FIGURE 1. The top left and right panels display the densities c_G and c of the copula models in (25) and (24), respectively. The middle left panel shows the density $c^{(n)}$ of the copula $C^{(n)}$ pertaining to the copula model in (24), with sample size $n = 100$. The middle right to bottom right panels depict the density ratio $c^{(n)}/c_G$ for $n = 2, 50, 100$, respectively.

distribution $C_G^{(k)}$. Then, standard empirical process arguments (see, e.g., [9, 17, 29]) yield that, for any $\varepsilon > 0$,

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

for some $b_\varepsilon \in (0, \varepsilon)$. The term on the right-hand side is of order $O\big(\mathrm{e}^{-c_\varepsilon k}\big)$ for some $c_\varepsilon > 0$. By Lemma 4.1, we have that $k^2C^{(n,k)} \lhd \mathrm{e}^{ck}C_G^{(k)}$ for all $c > 0$, where $C^{(n,k)}$ is the k-fold product measure corresponding to $C^{(n)}$. Let $\boldsymbol{U}^{(n,1:k)} = \big(\boldsymbol{U}^{(n,1)}, \ldots, \boldsymbol{U}^{(n,k)}\big)$, where

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

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

Remark 4.1. Notice that the assumption in (28) of Proposition 4.1 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 (28) seems troublesome. Still, numerically checking whether some copula models meet this assumption can be fairly simple. For instance, consider the copula of Example 2.2, given in (24), 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 in (25). In this case, Corollary 2.1 applies and $C^{(n)}$ converges to C_G in variational distance. Figure 1 displays plots of the densities c, c_G, and $c^{(n)}$ with $n = 100$. Outside a neighbourhood 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 (28) is satisfied.

4.2. Bayesian approach

A similar scheme is exploited by [23] in a Bayesian context, where the extended Schwartz theorem, e.g. [16, Theorem 6.23], provides exponential bounds for posterior concentration in a neighbourhood of the true parameter. In particular, [23] considers a nonparametric Bayesian approach for estimating the D-norm $\|\cdot\|_D$ and the densities of the associated angular measure; see [12, pp. 25–29]. Therein, Corollary 3.1 is leveraged to obtain a suitable remote contiguity result, allowing us 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

[1] BEIRLANT, J., GOEGEBEUR Y., SEGERS, J. AND TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley, Chichester.

[2] BERGHAUS, B. AND BÜCHER, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.* **46**, 2307–2335.

[3] BERGHAUS, B., BÜCHER, A. AND DETTE, H. (2013). Minimum distance estimators of the Pickands dependence function and related tests of multivariate extreme-value dependence. *J. Soc. Française Statist.* **154**, 116–137.

[4] BÜCHER, A. AND SEGERS, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17**, 495–528.

[5] BÜCHER, A. AND SEGERS, J. (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24**, 1427–1462.

[6] BÜCHER, A., VOLGUSHEV, S. AND ZOU, N. (2019). On second order conditions in the multivariate block maxima and peak over threshold method. *J. Multivar. Anal.* **173**, 604–619.

[7] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer, London.

[8] DE HAAN, L. AND PENG, L. (1997). Rates of convergence for bivariate extremes. *J. Multivar. Anal.* **61**, 195–230.

[9] DEHEUVELS, P. (1980). Non parametric tests of independence. In *Statistique non Paramétrique Asymptotique,* ed. J. P. Raoult, Springer, Berlin, pp. 95–107.

[10] DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21**, 420–436.

[11] DOMBRY, C., ENGELKE, S. AND OESTING, M. (2017). Bayesian inference for multivariate extreme value distributions. *Electron. J. Statist.*, **11**, 4813–4844.

[12] FALK, M. (2019). *Multivariate Extreme Value Theory and D-Norms*. Springer, New York.

[13] FALK, M., HÜSLER, J. AND REISS, R.-D. (2011). *Laws of Small Numbers: Extremes and Rare Events*, 3rd edn. Birkhäuser, Basel.

[14] FALK, M., PADOAN, S. A. AND WISHECKEL, F. (2019). Generalized Pareto copulas: a key to multivariate extremes. *J. Multivar. Anal.* **174**, 104538.

[15] FERREIRA, A. AND DE HAAN, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43**, 276–298.

[16] GHOSAL, S. AND VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.

[17] GUDENDORF, G. AND SEGERS, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *J. Statist. Planning Infer.* **142**, 3073–3085.

[18] KAUFMANN, E. AND REISS, R.-D. (1993). Strong convergence of multivariate point processes of exceedances. *Ann. Inst. Statist. Math.* **45**, 433–444.

[19] KLEIJN, B. J. K. (2017). On the frequentist validity of Bayesian limits. Preprint, available at arXiv:1611.08444v3.

[20] MCNEIL, A. J. AND NEŠLEHOVÁ, J. (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Ann. Statist.* **37**, 3059–3097.

[21] MARCON, G., PADOAN, S. A., NAVEAU, P., MULIERE, P. AND SEGERS, J. (2017). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *J. Statist. Planning Infer.* **183**, 1–17.

[22] MHALLA, L., CHAVEZ-DEMOULIN, V. AND NAVEAU, P. (2017). Non-linear models for extremal dependence. *J. Multivar. Anal.* **159**, 49–66.

[23] PADOAN, S. A. AND RIZZELLI, S. (2019). Strong consistency of nonparametric Bayesian inferential methods for multivariate max-stable distributions. Preprint, available at arXiv:1904.00245v2.

[24] REISS, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer, New York.

[25] RESNICK, S. I. (2008). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.

[26] SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.

[27] SKLAR, A. (1996) Random variables, distribution functions, and copulas – a personal look backward and forward. In *Distributions with Fixed Marginals and Related Topics*, eds L. Rüschendorf, B. Schweizer and M. D. Taylor. Lecture Notes – Monograph Series, Vol. **28**. Institute of Mathematical Statistics, Hayward, CA, pp. 1–14.

[28] WONG, W. H. AND SHEN, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLES. *Ann. Statist.* **23**, 339–362.

[29] WELLNER, J. A. (1992). Empirical processes in action: a review. *Internat. Statist. Rev.* **60**, 247–269.