J. Appl. Prob. **53,** 733–746 (2016)
*doi:*10.1017/jpr.2016.37
© *Applied Probability Trust* 2016

ONE-COMPONENT REGULAR VARIATION AND GRAPHICAL MODELING OF EXTREMES

ADRIEN HITZ * ** AND
ROBIN EVANS,* *University of Oxford*

Abstract

The problem of inferring the distribution of a random vector given that its norm is large requires modeling a homogeneous limiting density. We suggest an approach based on graphical models which is suitable for high-dimensional vectors. We introduce the notion of one-component regular variation to describe a function that is regularly varying in its first component. We extend the representation and Karamata's theorem to one-component regularly varying functions, probability distributions and densities, and explain why these results are fundamental in multivariate extreme-value theory. We then generalize the Hammersley–Clifford theorem to relate asymptotic conditional independence to a factorization of the limiting density, and use it to model multivariate tails.

Keywords: Regular variation; Karamata's theorem; homogeneous distribution; multivariate exceedances over threshold; graphical model

2010 Mathematics Subject Classification: Primary 60F99
Secondary 60E05; 62H99; 60G70

1. Introduction

Regular variation describes the behavior of some functions when evaluated close to ∞ [27]. In the multivariate case, regular variation of a measurable function u on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ with limit v such that $u(\lambda\mathbf{1}), v(\lambda\mathbf{1}) > 0$ for all $\lambda > 0$, is defined as

$$\frac{u(t\boldsymbol{x})}{u(t\mathbf{1})} \to v(\boldsymbol{x}), \quad \text{for all } \boldsymbol{x} \text{ as } t \to \infty; \tag{1.1}$$

see [24]. A striking result is that, in this case, the limit is homogeneous, i.e. there exists $\alpha \in \mathbb{R}$ such that $v(\lambda\boldsymbol{x}) = \lambda^\alpha v(\boldsymbol{x})$ for all $\lambda > 0$, for all $\boldsymbol{x}$.

We suggest an alternative definition: a function u is called *regularly varying with respect to its first component* if

$$\frac{u(tx, \boldsymbol{y})}{u(t, \mathbf{1})} \to v(x, \boldsymbol{y}),$$

where u and v are measurable nonnegative functions on $(0, \infty) \times \mathbb{R}^{d-1}$ such that $u(\cdot, \mathbf{1}) > 0$ and $v(\cdot, \mathbf{1}) > 0$. One-component regular variation includes regular variation as a special case. For instance, u is regularly varying on $\mathbb{R}^d_+ \setminus \{\mathbf{0}\}$ if and only if $u \circ \varphi^{-1}$ is regularly varying in its first component, where $\varphi(\boldsymbol{x}) = (||\boldsymbol{x}||, \boldsymbol{x}/||\boldsymbol{x}||)$ and $\varphi^{-1}(r, \boldsymbol{\theta}) = r\boldsymbol{\theta}$. In Section 2 we generalize two important results in Karamata's theory—i.e. the study of regular variation—from the univariate to the multivariate case. The first is the representation theorem [7], which states

Received 2 July 2015; revision received 30 October 2015.
* Postal address: Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK.
** Email address: adrien.hitz@lincoln.ox.ac.uk

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

that any regularly varying function can be written in a specific form. The second, Karamata's theorem, reveals relations between a regularly varying function and its integral.

Regular variation is, above all, the limit of a fraction. The specific form of the limit $v(x) = x^\alpha$ in the univariate case comes from the fact that the operation $\alpha \cdot x \mapsto x^\alpha$ is distributive over the division; when regular variation is applied in probability, the fraction corresponds to conditioning. The multiplication tx, however, is an arbitrary choice of operation that can be replaced by a more general scaling [6]. Choosing $t \star x \mapsto T^{-1}\{T(t)T(x)\}$ for a diffeomorphism T gives a limit of the form $T(x)^\alpha$, thus extending regular variation to other decays than the power law. This is not just a trivial transformation, but enriches the representation and Karamata's theorems.

Extreme-value theory studies the distribution of a random vector $\boldsymbol{X}$ in regions far from the origin. In [2], $\boldsymbol{X}$ is called regularly varying if there exists a probability distribution ν on $C_{||\cdot||} = \{\boldsymbol{x} \in \mathbb{R}^d : ||\boldsymbol{x}|| \geq 1\}$ such that (s.t.) $\nu(\lambda C_{||\cdot||}) > 0$ for some $\lambda > 1$, and

$$t^{-1}\boldsymbol{X} \mid ||\boldsymbol{X}|| \geq t \xrightarrow{\text{w}} \nu, \tag{1.2}$$

where '$\xrightarrow{\text{w}}$' denotes weak convergence. In this case, ν is homogeneous, i.e. there exists $\alpha > 0$ such that $\nu(\lambda A) = \lambda^{-\alpha}\nu(A)$ for all Borel sets A and $\lambda \geq 1$. Basrak's exact definition involves vague convergence to a nonnull Radon measure $\tilde{\nu}$ on $(\{-\infty\} \cup \mathbb{R} \cup \{\infty\})^d \setminus \{\boldsymbol{0}\}$ called the exponent measure. For simplicity and without losing generality, we stick to weak convergence between probability distributions throughout the paper.

We say that a random vector $(X, \boldsymbol{Y})$ is *regularly varying with respect to its first component,* with a limiting probability distribution μ on $[1, \infty) \times \mathbb{R}^{d-1}$ if

$$(t^{-1}X, \boldsymbol{Y}) \mid X \geq t \xrightarrow{\text{w}} \mu,$$

and $\mu_X \neq \delta_1$, the Dirac mass on 1. Limits of random vectors with an extreme component have already been studied [16], [25] and it is known that $\mu = P_\alpha \times H$, where P_α is the Pareto distribution of index $\alpha > 0$ and H is a probability distribution on $\mathbb{R}^{d-1}$. In particular, (1.2) is equivalent to one-component regular variation of $\varphi(\boldsymbol{X})$, in which case H corresponds to the angular distribution also called the spectral measure [4].

In Section 3 we adapt the representation and Karamata's theorems for random variables and, as a result, characterize homogeneous probability measures. In addition, we show that if $\boldsymbol{X}$ admits a regularly varying probability density with nonnull limit and if the sequence is, for instance, dominated by an integrable function, then

$$f_{t^{-1}\boldsymbol{X} \mid ||\boldsymbol{X}|| \geq t} \to f_{\boldsymbol{Y}}, \tag{1.3}$$

for a probability density $f_{\boldsymbol{Y}}$ on $C_{||\cdot||}$, and, thus, $\boldsymbol{X}$ is regularly varying. In this case, $f_{\boldsymbol{Y}}$ is homogeneous of order $-\alpha - d$. This extends [11, Theorem 2.1], which implies regular variation of $\boldsymbol{X}$ when, in addition to regular variation of $f_{\boldsymbol{X}}$, the convergence is uniform, and the limit is bounded on $C_{||\cdot||}$.

Besides describing exceedances over a threshold, regular variation of $\boldsymbol{X}$ is directly related to the behavior of maxima of independent copies $\boldsymbol{X}^{(i)}$. As explained in [24], if $\boldsymbol{X}$ is nonnegative, then it is regularly varying if and only if there exists a sequence $a_n \to \infty$ such that $\max_{i=1,\ldots,n} a_n^{-1}\boldsymbol{X}^{(i)} \xrightarrow{\text{w}} G$ for a nondegenerate probability distribution G on $\mathbb{R}^d_+$; if the marginals of G are not degenerated, then they are Fréchet distributed and G is called a type-II multivariate extreme-value distribution. The cumulative distribution function

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

of G is $\exp\{-\tilde{\nu}([\mathbf{0}, \boldsymbol{x}]^c\}$, where $\tilde{\nu}$ is the exponent measure. The problem of modeling $\tilde{\nu}$ or the corresponding copula has received much attention in multivariate analysis. Unless d is particularly small, the density of G is approximated by the one of $\tilde{\nu}$ using a Taylor expansion, i.e. up to a constant, by $f_{\boldsymbol{Y}}$. Hence, modeling both maxima and exceedances of $\boldsymbol{X}$ requires modeling a homogeneous density.

The usual strategy is to write $f_{\boldsymbol{Y}}(\boldsymbol{y}) = ||\boldsymbol{y}||^{-\alpha-d} h(\boldsymbol{y}/||\boldsymbol{y}||)$ and model the angular density $h = dH$ satisfying some constraints to ensure that $\boldsymbol{Y}$ has Pareto marginals. Suggested models for h include the asymmetric logistic [28], the pairwise beta [9], its generalization [1], mixtures of Dirichlet [8] and the angular density of the Hüsler–Reiss exponent measure [13]. However, they suffer limitations in high-dimensions as the number of parameters explodes and may lack flexibility to describe multivariate tails.

As a consequence of (1.3), any parametric probability density $f_{\boldsymbol{X}}(\cdot; \boldsymbol{\theta})$ that is regularly varying with decay T and satisfies the additional condition induces a homogeneous parametric density $f_{\boldsymbol{Y}}(\cdot\,; \boldsymbol{\theta})$ in the limit. If the multivariate marginals and the censored densities of $f_{\boldsymbol{X}}(\cdot; \boldsymbol{\theta})$ are available, we directly find the ones of $f_{\boldsymbol{Y}}(\cdot; \boldsymbol{\theta})$ by passing to the limit. A complementary approach pursued from Section 4 is to simplify $f_{\boldsymbol{Y}}$ using graphical models, which have been successful in reducing dimensionality [29]. In Section 5 we define a new notion called *asymptotic conditional independence* and we generalize Hammersley–Clifford for a sequence of densities that factorize in the limit with respect to a graph.

In Section 6 we explain how to use asymptotic graphical models for multivariate exceedances. We consider a sequence $\boldsymbol{X}_t^{\mathrm{C}}$ based on censoring the marginals of $\boldsymbol{X}$ whose absolute values fall below t. Under (1.3) and $f_{\boldsymbol{Y}} > 0$, the generalized Hammersley–Clifford theorem states that some marginals of $\boldsymbol{X}_t^{\mathrm{C}}$ are asymptotically conditionally independent given the rest of the vector if and only if the censored density of $\boldsymbol{Y}$ is determined by lower dimensional marginals $\boldsymbol{Y}_C$, where C are the cliques of a graph. This enables us to perform inference on each clique separately once the graph has been selected.

2. One-component regular variation for functions

Consider the commutative and associative operation

$$x \star y := T^{-1}\{T(x)T(y)\}, \qquad x, y \in E = [e_0, e_1), \tag{2.1}$$

where $T\colon E \to [1, \infty)$ is a diffeomorphism, for $e_0 \in \mathbb{R}$, $e_1 \in \mathbb{R} \cup \{\infty\}$. When $E = [0, \infty)$, possible operations include multiplication ($T = \mathrm{id}$), addition ($T = \exp$), and $x \star y = ||x, y||_p$ for $p > 0$, $T(x) = \exp(x^p)$. When $E = [0, e_1)$ and $e_1 < \infty$, $T(x) = (1 - e_1^{-1}x)^{-1}$ gives $x \star y = x + y - e_1^{-1}xy$. We call a positive and measurable function u on E *regularly varying with decay* T and limit v if

$$\frac{u(t \star x)}{u(t \star e_0)} \to v(x) > 0, \tag{2.2}$$

where the arrow stands for pointwise convergence on the entire domain of definition as $t \uparrow e_1$. It follows that $u \circ T^{-1}$ is regularly varying (RV), and, thus, $v(x) = T(x)^\alpha$ for some $\alpha \in \mathbb{R}$. The convergence moreover is uniform [24]. We denote (2.2) by $u \in T - \mathrm{RV}_\alpha$.

Let us extend this notion to the multivariate setting. Consider two measurable and nonnegative functions u and v on $E \times \mathbb{R}^{d-1}$ such that $u(\cdot, \mathbf{1}) > 0$, $v(\cdot, \mathbf{1}) > 0$, and a nonnegative function h on $\mathbb{R}^{d-1}$ satisfying $h(\mathbf{1}) = 1$. We call $\mathbf{1}$ the pivot and its choice is arbitrary.

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

Lemma 2.1. (Characterization.) *The following are equivalent:*

(i) $u(t \star x, \boldsymbol{y})/u(t \star e_0, \mathbf{1}) \to v(x, \boldsymbol{y})$;

(ii) $u(\cdot, \mathbf{1}) \in T - \mathrm{RV}_\alpha$ *and* $u(t, \boldsymbol{y})/u(t, \mathbf{1}) \to h(\boldsymbol{y})$;

(iii) $u(\cdot, \boldsymbol{y}) \in T - \mathrm{RV}_\alpha$ *for all* $\boldsymbol{y}$ *s.t.* $h(\boldsymbol{y}) > 0$, *and* $u(t, \boldsymbol{y})/u(t, \mathbf{1}) \to h(\boldsymbol{y})$;

(iv) $u(t \star x, \boldsymbol{y})/u(t \star e_0, \mathbf{1}) \to T(x)^\alpha\ h(\boldsymbol{y})$ *uniformly in* x.

We say that u is *regularly varying in its first component* if Lemmas 2.1(i)–(iv) hold, written $u \in T - \mathrm{RV}_\alpha^x(h)$, or simply $\mathrm{RV}_\alpha^x(h)$ when $T = \mathrm{id}$. Condition $u(\cdot, \boldsymbol{y}) \in \mathrm{RV}_\alpha$ for all $\boldsymbol{y}$ corresponds to the uniform regular variation of [22] if it is assumed further that the convergence is uniform in $\boldsymbol{y}$.

We now generalize the representation theorem for univariate regular variation (see [7] in the case $T = \mathrm{id}$, [18] otherwise).

Theorem 2.1. (Multivariate representation theorem.) *It holds* $u \in T - \mathrm{RV}_\alpha^x(h)$ *if and only if*

$$u(x, \boldsymbol{y}) = c(x) \exp\left\{\int_{e_0}^{x} \alpha(z) \frac{T'(z)}{T(z)}\, \mathrm{d}z\right\} q(x, \boldsymbol{y})$$

for some measurable functions s.t. $c(t) \to c > 0$, $\alpha(t) \to \alpha$, $q(t, \boldsymbol{y}) \to h(\boldsymbol{y})$ *as* $t \uparrow e_1$.

As a consequence, for all $\varepsilon > 0$, there exists $c_1, c_2 > 0$ such that

$$c_1 T(x)^{\alpha-\varepsilon} h(\boldsymbol{y}) < u(x, \boldsymbol{y}) < c_2 T(x)^{\alpha+\varepsilon} h(\boldsymbol{y}) \tag{2.3}$$

for sufficiently large x and $\boldsymbol{y}$ satisfying $h(\boldsymbol{y}) \neq 0$.

Proof of Theorem 2.1. For the direct implication, write $u(x, \boldsymbol{y}) = u(x, \boldsymbol{y})/u(x, \mathbf{1})u(x, \mathbf{1})$. Lemma 2.1 gives $q(t, \boldsymbol{y}) := u(t, \boldsymbol{y})/u(t, \mathbf{1}) \to h(\boldsymbol{y})$ and $u(x, \mathbf{1}) \in T - \mathrm{RV}_\alpha$. The conclusion follows by applying the representation theorem on the regularly varying function $u\{T^{-1}(x), \mathbf{1}\}$. For the reverse, consider $z = t \star \bar{z}$ and $\mathrm{d}z = T(t)T'(\bar{z})/T'(t \star \bar{z})$ to find

$$\int_{t \star e_0}^{t \star x} \frac{\alpha(z)T'(z)}{T(z)}\, \mathrm{d}z = \int_{e_0}^{x} \frac{\alpha(t \star \bar{z})T'(\bar{z})}{T(\bar{z})}\, \mathrm{d}\bar{z} \to \alpha \log T(x),$$

and, thus, $u(t \star x, \boldsymbol{y})/u(t \star e_0, \mathbf{1}) \to T(x)^\alpha h(\boldsymbol{y})$. □

Another important result is Karamata's theorem, which relates the regular variation of a univariate function to that of its integral [7]. We generalize it for one-component regular variation, only treating the case of a negative power index $-\alpha$ for $\alpha > 0$. Suppose that $\bar{U}(x, \boldsymbol{y}) := \int_x^\infty u(z, \boldsymbol{y})\, \mathrm{d}z$ exists for all x, $\boldsymbol{y}$ (it exists for sufficiently large e_0 if $u(x, \boldsymbol{y})/T'(x) \in T - \mathrm{RV}_{-\alpha-1}^x(h)$ as a consequence of (2.3)).

Theorem 2.2. (Multivariate Karamata's theorem.) *It holds that*

$$\frac{u(x, \boldsymbol{y})}{T'(x)} \in T - \mathrm{RV}_{-\alpha-1}^x(h) \quad \Longleftrightarrow \quad \frac{T(t)}{T'(t)} \frac{u(t, \boldsymbol{y})}{\bar{U}(t, \mathbf{1})} \to \alpha h(\boldsymbol{y}).$$

In this case, $\bar{U} \in T - \mathrm{RV}_{-\alpha}^x(h)$, *and its representation has coefficient* $c(x) = \bar{U}(e_0, \mathbf{1})$.

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

Proof. To prove the direct implication, use the change of variable $z = t \star \tilde{z}$ to find

$$\begin{aligned}\frac{T(t\star e_0)}{T'(t\star e_0)}\frac{u(t\star e_0,\boldsymbol{y})}{\int_{t\star e_0}^{\infty}u(z,\mathbf{1})\,\mathrm{d}z} &= \left\{\int_{e_0}^{\infty}\frac{u(t\star\tilde{z},\mathbf{1})}{u(t\star e_0,\mathbf{1})}\frac{T'(t\star e_0)}{T'(t\star\tilde{z})}\frac{T'(\tilde{z})T(t)}{T(t\star e_0)}\,\mathrm{d}\tilde{z}\right\}^{-1}\frac{u(t\star e_0,\boldsymbol{y})}{u(t\star e_0,\mathbf{1})}\\ &\to \left\{\int_{e_0}^{\infty}T(\tilde{z})^{-\alpha-1}T'(\tilde{z})\,\mathrm{d}\tilde{z}\right\}^{-1}h(\boldsymbol{y})\\ &= \alpha h(\boldsymbol{y}),\end{aligned}$$

since $u(\cdot,\mathbf{1})/T'(\cdot) \in T-\mathrm{RV}_{-\alpha-1}$ and $u(t\star e_0,\boldsymbol{y})/u(t\star e_0,\mathbf{1}) \to h(\boldsymbol{y})$ from Lemma 2.1. Limit and integral can be exchanged since $cT(z)^{-\alpha-1+\varepsilon}T'(z)$ dominates the integrand for some $c>0$, $\varepsilon\in(0,\alpha)$, and sufficiently large t thanks to (2.3).

For the reverse implication, let $\alpha(x) := \{u(x,\mathbf{1})/\bar{U}(x,\mathbf{1})\}/\{T'(x)/T(x)\}$, which satisfies $\alpha(t)\to\alpha$. Integrate $u(z,\mathbf{1})/\bar{U}(z,\mathbf{1}) = \alpha(z)T'(z)/T(z)$ from both sides between e_0 and x to obtain $\bar{U}(x,\mathbf{1}) = \bar{U}(e_0,\mathbf{1})\exp\{-\int_{e_0}^{x}\alpha(z)T'(z)/T(z)\,\mathrm{d}z\}$. This is the representation of a T-regularly varying function with $c(x) = \bar{U}(e_0,\mathbf{1})$ thanks to Theorem 2.1. Hence, $T'(t\star e_0)/T'(t\star x)\,u(t\star x,\boldsymbol{y})/u(t\star e_0,\mathbf{1})$ equals

$$\frac{T'(t\star e_0)\bar{U}(t\star e_0,\mathbf{1})}{T(t\star e_0)u(t\star e_0,\mathbf{1})}\frac{T(t\star x)u(t\star x,\boldsymbol{y})}{T'(t\star x)\bar{U}(t\star x,\mathbf{1})}\frac{T(t\star e_0)}{T(t\star x)}\frac{\bar{U}(t\star x,\mathbf{1})}{\bar{U}(t\star e_0,\mathbf{1})}\to T(x)^{-\alpha-1}h(\boldsymbol{y}),$$

which completes the proof of the equivalence in (2.2). Moreover, $\bar{U}\in T-\mathrm{RV}_{\alpha}(h)$ since

$$\begin{aligned}\frac{\bar{U}(t\star x,\boldsymbol{y})}{\bar{U}(t\star e_0,\mathbf{1})} &= \bar{U}(t\star e_0,\mathbf{1})^{-1}\int_{t\star x}^{\infty}u(\tilde{x},\boldsymbol{y})\,\mathrm{d}\tilde{x}\\ &= \frac{u(t\star e_0,\mathbf{1})}{\bar{U}(t\star e_0,\mathbf{1})}\frac{T(t\star e_0)}{T'(t\star e_0)}\int_{x}^{\infty}\frac{u(t\star z,\boldsymbol{y})}{u(t\star e_0,\mathbf{1})}\frac{T'(t\star e_0)}{T'(t\star z)}T'(z)\,\mathrm{d}z\\ &\to T(x)^{-\alpha}h(\boldsymbol{y}).\end{aligned}$$

□

For instance, consider the Gaussian distribution $\mathcal{N}(\mu,\sigma^2)$ with probability density ϕ and survival function $\bar{\Phi}$. Since $2^{-1}t^{-1}\phi(t)/\bar{\Phi}(t)\to 2^{-1}\sigma^{-2}$, Theorem 2.2 gives $\phi(x)/T'(x)\in T-\mathrm{RV}_{-2^{-1}\sigma^{-2}-1}$ and $\bar{\Phi}\in T-\mathrm{RV}_{-2^{-1}\sigma^{-2}}$ for $T(x)=\exp(x^2)$. Let f be the probability density of the log-Cauchy distribution and $\bar{F}$ its survival function. As $t\log t f(t)/\bar{F}(t)\to 1$, it follows that $xf(x)\in\log-\mathrm{RV}_{-2}$, and $\bar{F}\in\log-\mathrm{RV}_{-1}$.

We now extend the standard multivariate regular variation in (1.1) to general decays. We denote the sign of $x\in\mathbb{R}$ by $\sigma_x\in\{-1,0,1\}$ and $\boldsymbol{\sigma_x} = (\sigma_{x_1},\dots,\sigma_{x_d})$ when $\boldsymbol{x}\in\mathbb{R}^d$. Operations between vectors are done componentwise, $x\star\boldsymbol{y} := (x\mathbf{1})\star\boldsymbol{y}$, and $\boldsymbol{e}_0 := e_0\mathbf{1}$. We say that $u\colon\mathbb{R}^d\to[0,\infty)$ satisfying $u(\lambda\mathbf{1})>0$ for all $\lambda\in E$, is *regularly varying with decay T* if there exists v such that $v(\lambda\mathbf{1})>0$ for all $\lambda\in E$ and

$$\frac{u\{(t\star|\boldsymbol{x}|)\boldsymbol{\sigma_x}\}}{u(t\star\boldsymbol{e}_0)}\to v(\boldsymbol{x}) \tag{2.4}$$

on $F := \{(-e_1,-e_0)\cup\{0\}\cup[e_0,e_1)\}^d\setminus\{\mathbf{0}\}$; by convention, $(t\star|x|)1_x = 0$ if $x=0$, where 1_x is the indicator function.

Multivariate regular variation is easily expressed in terms of one-component regular variation by introducing the following change of variable. We say that $\phi\colon C\to(0,\infty)\times\Omega$, for $\Omega\subset\mathbb{R}^{d-1}$, defines a *radial system of coordinates* if it has the form

$$\phi\colon\boldsymbol{x}\mapsto\{r(\boldsymbol{x}),\boldsymbol{\theta}(\boldsymbol{x})\},\qquad \phi^{-1}\colon(r,\boldsymbol{\theta})\mapsto\{r\star|\boldsymbol{\theta}^{-1}(\boldsymbol{\theta})|\}\boldsymbol{\sigma}_{\boldsymbol{\theta}^{-1}(\boldsymbol{\theta})}, \tag{2.5}$$

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

where $r(\cdot)$ and $\boldsymbol{\theta}(\cdot)$ satisfy $r\{(\lambda \star |\boldsymbol{x}|)\boldsymbol{\sigma}_{\boldsymbol{x}}\} = \lambda \star r(\boldsymbol{x}), \boldsymbol{\theta}\{(\lambda \star |\boldsymbol{x}|)\boldsymbol{\sigma}_{\boldsymbol{x}}\} = \boldsymbol{\theta}(\boldsymbol{x})$ for all $\lambda \in E$. Examples include spherical coordinates on $\mathbb{R}^d$ and pseudo-polar coordinates defined by $r(\boldsymbol{x}) = ||\boldsymbol{x}||$, a norm, and $\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{x}_{1:d-1}/r(\boldsymbol{x})$ on $\mathbb{R}^d_+$. When $T = \exp$, the latter translates into $r(\boldsymbol{x}) = \log(||\mathrm{e}^{\boldsymbol{x}}||), \boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{x} - r(\boldsymbol{x})$. It becomes clear that u is regularly varying if and only if $g(r, \boldsymbol{\theta}) := u\{\phi^{-1}(r, \boldsymbol{\theta})\}$ is regularly varying with respect to its first component since

$$\frac{g(t \star r, \boldsymbol{\theta})}{g(t \star \mathbf{1}_r, \mathbf{1}_{\boldsymbol{\theta}})} = \frac{u\{(t \star |\boldsymbol{x}|)\boldsymbol{\sigma}_{\boldsymbol{x}}\}}{u(t \star \boldsymbol{e}_0)} \to v(\boldsymbol{x}),$$

and in this case, Lemma 2.1 gives $v(\boldsymbol{x}) = T\{r(\boldsymbol{x})\}^{\alpha} h\{\boldsymbol{\theta}(\boldsymbol{x})\}$, where $\mathbf{1}_r = r(\boldsymbol{e}_0)$ and $\mathbf{1}_{\boldsymbol{\theta}} = \boldsymbol{\theta}(\boldsymbol{e}_0)$. In particular, v is homogeneous of order α, i.e. $v\{(\lambda \star |\boldsymbol{x}|)\boldsymbol{\sigma}_{\boldsymbol{x}}\} = T(\lambda)^{\alpha} v(\boldsymbol{x})$ for all $\lambda \in E$, and we write (2.4) as $u \in T - \mathrm{RV}_{\alpha}(v)$.

3. One-component regular variation for probability distributions

So far, one-component regular variation has been treated for functions; we develop it further for distributions. A specific representation is found for regularly varying random vectors. Their limits moreover form the class of distributions that are homogeneous with respect to their first component. In a subsequent result, we show that one-component regular variation of the probability density implies, under an extra condition, one-component regular variation of the distribution.

For simplicity, we treat the case $T = \mathrm{id}$. Let $(X, \boldsymbol{Y})$ be a random variable with probability distribution μ on $[1, \infty) \times \mathbb{R}^{d-1}$. We call $\bar{F}(x, \boldsymbol{y}) = \mu\{[x, \infty) \times (-\boldsymbol{\infty}, \boldsymbol{y}]\}$ the *x-survival function*. Subscripts X or $\boldsymbol{Y}$ refer to the corresponding marginal distribution. Suppose that $\bar{F}_X > 0$. We are interested in the weak limit [5] of $(t^{-1}X, \boldsymbol{Y}) \mid X \geq t$, equivalent to the weak limit of its distribution $\mu(tA, B)/\bar{F}_X(t)$, and of its x-survival function $\bar{F}(tx, \boldsymbol{y})/\bar{F}_X(t)$. Let H and ν be two probability distributions on $\mathbb{R}^{d-1}$ and E, respectively, such that $\nu_X \neq \delta_1$. We denote by P_{α} the Pareto distribution on $[1, \infty)$ with shape $\alpha > 0$.

Theorem 3.1. *The following are equivalent:*

(i) $(t^{-1}X, \boldsymbol{Y}) \mid X \geq t \xrightarrow{\mathrm{w}} \nu$;

(ii) $\boldsymbol{Y} \mid X \geq t \xrightarrow{\mathrm{w}} H$ *and* $t^{-1}X \mid X \geq t \xrightarrow{\mathrm{w}} P_{\alpha}$;

(iii) $\bar{F}(x, \boldsymbol{y}) = c(x) \exp\{-\int_{e_0}^{x} \alpha(z) z^{-1} \,\mathrm{d}z\} Q(\boldsymbol{y} \mid x)$, *for measurable* $\alpha(t) \to \alpha, c(t) \to c$, *and a conditional cumulative distribution function* $Q(\boldsymbol{y} \mid t) \xrightarrow{\mathrm{w}} H$;

(iv) $(t^{-1}X, \boldsymbol{Y}) \mid X \geq t \xrightarrow{\mathrm{w}} P_{\alpha} \times H$.

We say that $(X, \boldsymbol{Y})$ is *regularly varying with respect to its first component*, written $(X, \boldsymbol{Y}) \in \mathrm{RV}^{x}_{-\alpha}(H)$, if Theorem 3.1 is satisfied. To see that $\nu_X \neq \delta_1$ is a necessary assumption, consider, for instance, $\bar{F}(x) = \mathrm{e}^{-x}$. The equivalence of Theorem 3.1(i) and 3.1(iv) can be found in [21, Theorem 3.1 with $\{\lambda, (x, \boldsymbol{y})\} \mapsto (\lambda x, \boldsymbol{y})$] or in [16, Proposition 2 with $\alpha = 1, \beta = 0$]. Our contribution is the representation in (iii) and a short proof.

Proof of Theorem 3.1. (i) implies (ii). By [2, Theorem 2.1.4], $t^{-1}X \mid X \geq t \xrightarrow{\mathrm{w}} \nu_X$ implies that $\nu_X = P_{-\alpha}$ for some $\alpha > 0$.

(ii) implies (iii). Since $\bar{F}_X \in \mathrm{RV}_{-\alpha}$ and $\bar{F}_{\boldsymbol{Y} \mid X \geq t} \xrightarrow{\mathrm{w}} H$, write

$$\bar{F}(x, \boldsymbol{y}) = \bar{F}_X(x) \bar{F}_{\boldsymbol{Y} \mid X \geq x}(\boldsymbol{y} \mid x)$$

and apply the representation theorem on $\bar{F}_X$.

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

(iii) implies (iv). We have

$$\frac{\bar{F}(tx, \boldsymbol{y})}{\bar{F}_X(t)} = \frac{c(tx)}{c(t)} \exp\left\{-\int_t^{tx} \alpha(z) z^{-1} \, \mathrm{d}z\right\} Q(\boldsymbol{y} \mid tx) \xrightarrow{\mathrm{w}} x^{-\alpha} H(\boldsymbol{y}),$$

since $\int_t^{tx} \alpha(z) z^{-1} \, \mathrm{d}z = \int_1^x \alpha(t\tilde{z}) \tilde{z}^{-1} \, \mathrm{d}z \to \alpha \log x$.

(iv) implies (i). The proof is clear. □

Furthermore, suppose that $(X, \boldsymbol{Y})$ admits a probability density f with respect to the Lebesgue measure such that f_X and $f(\cdot, \mathbf{1})$ are positive. We want to know what the relation is between regular variation of f and of $(X, \boldsymbol{Y})$.

We start by answering the question in the univariate case. From Karamata's theorem (Theorem 2.2), $f_X \in \mathrm{RV}_{-\alpha-1}$ implies $\bar{F}_X \in \mathrm{RV}_{-\alpha}$ and its representation has coefficient $c(\cdot) = 1$. Thus, $X \in \mathrm{RV}_{-\alpha}$ and its representation now determines exactly a probability distribution: $\bar{F}_X(x) = \exp\{-\int_1^x \alpha(z) z^{-1} \, \mathrm{d}z\}$ for all $x \geq 1$, for a positive and measurable $\alpha(t) \to \alpha > 0$. The following example shows that regular variation of X does not imply regular variation of f_X in general. Equalities and pointwise convergences between densities hold almost everywhere.

Example 3.1. Let $\bar{F}_X(x) = \exp\{-\int_1^x \alpha(y) y^{-1} \, \mathrm{d}y\}$ on $[1, \infty)$ and $\alpha(x) = \sin(x) + 2$. Since $\bar{F}_X(x) = c(x) x^{-2}$ for $c(x) = \exp\{-\int_1^x y^{-1} \sin y \, \mathrm{d}y\}$ satisfying $c(t) \to 1$, the representation theorem gives $\bar{F}_X \in \mathrm{RV}_{-2}$, and, thus, $X \in \mathrm{RV}_{-2}$. However, $t f_X(t)/\bar{F}_X(t) = \alpha(t) \not\to 2$, and from Karamata's theorem f_X is not regularly varying.

We now provide an answer in the multivariate case. If $f \in \mathrm{RV}^x(v)$ such that the limit v is nonnull and the sequence is dominated by an integrable function, then by the dominated convergence theorem

$$f_{t^{-1}X, \boldsymbol{Y} \mid X \geq t}(x, \boldsymbol{y}) = c_t^{-1} \frac{f(tx, \boldsymbol{y})}{f(t, \mathbf{1})} \to c^{-1} v(x, \boldsymbol{y}), \tag{3.1}$$

where $c_t = \int_1^\infty \int_{\mathbb{R}^{d-1}} f(tx, \boldsymbol{y})/f(t, \mathbf{1}) \, \mathrm{d}x \, \mathrm{d}\boldsymbol{y} \to c = \int_1^\infty \int_{\mathbb{R}^{d-1}} v(x, \boldsymbol{y}) \, \mathrm{d}x \, \mathrm{d}\boldsymbol{y} < \infty$. Alternatively, we can assume v integrable and the sequence monotone instead of dominated. Let $\alpha > 0$ and H be a probability distribution on $\mathbb{R}^{d-1}$ with density h satisfying $h(\mathbf{1}) > 0$. According to Lemma 2.1, the limit in (3.1) has the form $\alpha x^{-\alpha-1} h(\boldsymbol{y})$, and, thus, $(X, \boldsymbol{Y}) \in \mathrm{RV}_{-\alpha}(H)$. This is an important result: we can guarantee regular variation of $(X, \boldsymbol{Y})$ simply by computing the limit of $f(tx, \boldsymbol{y})/f(t, \mathbf{1})$, a useful approach when $\bar{F}$ is intractable or the weak convergences in Theorem 3.1 are difficult to check. Whereas monotonicity or domination is sufficient to obtain (3.1), our next result reveals necessary and sufficient conditions.

Theorem 3.2. *The following are equivalent:*

(i) $f_{\boldsymbol{Y} \mid X}(\mathbf{1} \mid t) \to h(\mathbf{1})$ *and* $f \in \mathrm{RV}^x_{-\alpha-1}\{h(\cdot)/h(\mathbf{1})\}$;

(ii) $f_{\boldsymbol{Y} \mid X}(\boldsymbol{y} \mid t) \to h(\boldsymbol{y})$ *and* $f_X \in \mathrm{RV}_{-\alpha-1}$;

(iii) $f_{\boldsymbol{Y} \mid X}(\boldsymbol{y} \mid t) \to h(\boldsymbol{y})$ *and* $f_{X \mid \boldsymbol{Y}=\mathbf{1}} \in \mathrm{RV}_{-\alpha-1}$;

(iv) $f_{\boldsymbol{Y} \mid X}(\boldsymbol{y} \mid t) \to h(\boldsymbol{y})$ *and* $f_{X \mid \boldsymbol{Y}=\boldsymbol{y}} \in \mathrm{RV}_{-\alpha-1}$, *for all* $\boldsymbol{y}$ *s.t.* $h(\boldsymbol{y}) > 0$;

(v) $f(x, \boldsymbol{y}) = \alpha(x) x^{-1} \exp\{-\int_1^x \alpha(z) z^{-1} \, \mathrm{d}z\} q(\boldsymbol{y} \mid x)$, *for a positive and measurable* $\alpha(t) \to \alpha$ *and a conditional probability density* $q(\boldsymbol{y} \mid t) \to h(\boldsymbol{y})$, *for all* $x \geq 1$, *for all* $\boldsymbol{y}$;

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

(vi) $f_{t^{-1}X,\boldsymbol{Y}\mid X\geq t}(x,\boldsymbol{y})\to\alpha x^{-\alpha-1}h(\boldsymbol{y})$.

In this case, $(X,\boldsymbol{Y})\in \mathrm{RV}^{x}_{-\alpha}(H)$.

In (i), the first condition is necessary and can be replaced by $tf(t,\mathbf{1})/\bar{F}(t)\to\alpha h(1)$, and in (ii) by $tf(t,\boldsymbol{y})/\bar{F}_X(t)\to\alpha h(\boldsymbol{y})/h(\mathbf{1})$.

Proof of Theorem 3.2. (i) It is straightforward to see that (i) is equivalent to (iv). Use Lemma 2.1 to show that (iii) implies (iv) and (iv) implies (i).

(ii) implies (v). Write $f(x,\boldsymbol{y})=f_X(x)f_{\boldsymbol{Y}\mid X}(\boldsymbol{y}\mid x)$. On the one hand, $f_X\in\mathrm{RV}_{-\alpha}$ and thus from Karamata's theorem (Theorem 2.2) $\bar{F}_X$ has a representation with $c(\cdot)=1$. Differentiate this representation to find one for f_X. On the other hand, $f_{\boldsymbol{Y}\mid X}(\boldsymbol{y}\mid t)\to h(\boldsymbol{y})$ by assumption. Altogether, this gives (v), a well-defined probability density because ε and q are measurable, and Tonelli's theorem ensures it integrates to 1.

(v) implies (vi). Integrate (v) to find $\bar{F}(x)=\exp\{-\int_1^x\alpha(z)z^{-1}\,\mathrm{d}z\}$, and, thus,

$$\frac{tf(tx,\boldsymbol{y})}{\bar{F}_X(t)}=\alpha(tx)x^{-1}\exp\left\{-\int_1^x\alpha(tz)z^{-1}\,\mathrm{d}z\right\}q(\boldsymbol{y}\mid tx)\to\alpha x^{-\alpha-1}h(\boldsymbol{y}).$$

(vi) implies (i). First,

$$\frac{f(tx,\boldsymbol{y})}{f(t,\mathbf{1})}=\frac{tf(tx,\boldsymbol{y})}{\bar{F}_X(t)}\frac{\bar{F}_X(t)}{tf(t,\mathbf{1})}\to x^{-\alpha-1}\frac{h(\boldsymbol{y})}{h(\mathbf{1})}.$$

Second, apply Scheffé's lemma [12] on (vi) to obtain $(X,\boldsymbol{Y})\in\mathrm{RV}^{x}_{-\alpha}(H)$ and, in particular, $X\in\mathrm{RV}_{-\alpha}$. Karamata's theorem ensures that

$$f_{\boldsymbol{Y}\mid X}(\mathbf{1}\mid t)=\frac{tf(t,\mathbf{1})}{\bar{F}_X(t)}\frac{\bar{F}_X(t)}{tf_X(t)}\to h(\mathbf{1}).\qquad\square$$

The following example illustrates the use of Theorem 3.2, which holds also when the convergence is not uniform in $\boldsymbol{y}$.

Example 3.2. Consider

$$f(x,y)=\begin{cases}\frac{6}{5}x^{-2} & \text{if } y\in(0,x^{-1}),\\ \frac{6}{5}x^{-2}y & \text{otherwise,}\end{cases}\qquad\text{on } [1,\infty)\times(0,1].$$

The convergence $f(tx,y)/f(t,1)\to x^{-2}y$ gives $f\in\mathrm{RV}_{-2}\{h(\cdot)/h(1)\}$ for the cumulative distribution function $H(y)=y^2$ with probability density $h(y)=2y$ on $(0,1]$. Since $f(t,1)/f_X(t)\to 2=h(1)$ (Theorem 3.2(i)), $(X,\boldsymbol{Y})\in\mathrm{RV}_{-\alpha}(H)$, i.e. $\bar{F}(tx,y)/\bar{F}_X(t)\to x^{-1}y^2$. The same conclusion is drawn by showing that $f_X\in\mathrm{RV}_{-2}$ and $f_{Y\mid X}(y\mid t)\to 2y=h(y)$ (Theorem 3.2(ii)).

The results of this section are easily extended to hold for the arbitrary decay T defined in (2.1). We write $(X,\boldsymbol{Y})\in T-\mathrm{RV}^{x}_{-\alpha}(H)$ if $\{T(X),\boldsymbol{Y}\}\in\mathrm{RV}^{x}_{-\alpha}(H)$. From the continuous mapping theorem, this is equivalent to $\mu_{t\star}\overset{\mathrm{w}}{\to}\nu$, where $\mu_{\lambda\star}(A,B):=\mu(\lambda\star A,B)/\mu_X(\lambda\star E)$ is a probability distribution on $E\times\mathbb{R}^d$ for every $\lambda\in E$. In this case, $\nu=P_\alpha\circ T\times H$, and in particular ν_X has survival function $T(x)^{-\alpha}$. We say that μ is homogeneous with respect to its first component if $\mu(\lambda\star A,B)=T(\lambda)^{-\alpha}\mu(A,B)$ for all $\lambda\in E$. Theorem 3.1 characterizes such distributions.

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

Corollary 3.1. *The following are equivalent:*

(i) μ *is homogeneous of order* $-\alpha$ *with respect to its first component;*

(ii) $\bar{F}$ *is homogeneous of order* $-\alpha$ *with respect to its first component;*

(iii) $\mu_{\lambda\star} = \mu$, *for all* $\lambda \in E$; $X \perp\!\!\!\perp \boldsymbol{Y}$ *and* $X \sim P_\alpha \circ T$;

(iv) *(under its existence)* $f_{X,\boldsymbol{Y}}$ *is homogeneous of order* $-\alpha - 1$ *with respect to its first component.*

We now focus on one-component regular variation when the first component is the radius of a random vector $\boldsymbol{X}$ and the rest of the vector is the angle—note the change in the notation. Let $(R, \boldsymbol{\Theta}) = \phi(\boldsymbol{X}) = \{r(\boldsymbol{X}), \boldsymbol{\theta}(\boldsymbol{X})\}$ be its expression in radial coordinates defined in (2.5) with $T = \mathrm{id}$ and suppose that $\bar{F}_R > 0$. Let $\alpha > 0$ and $\boldsymbol{Y} \sim \nu$, a probability distribution on $C_r = \{r(\boldsymbol{x}) \geq 1\}$, such that $r(\boldsymbol{Y}) \not\sim \delta_1$.

Analogously to functions, $\boldsymbol{X}$ is multivariate regularly varying in the sense (1.2) if and only if $(R, \boldsymbol{\Theta})$ is one-component regularly varying, i.e.

$$t^{-1}\boldsymbol{X} \mid \boldsymbol{X} \in tC_r \xrightarrow{\mathrm{w}} \nu \text{ on } C_r \quad \Longleftrightarrow \quad (R, \boldsymbol{\Theta}) \in \mathrm{RV}_{-\alpha}(H) \quad \text{on } [1, \infty) \times \Omega. \tag{3.2}$$

This relation is well-known when, for instance, ϕ is the change of variables into pseudo polar coordinates [3] and is easily proved in the general case using the continuous mapping theorem and Corollary 3.1. We denote the multivariate regular variation in (3.2) by $\boldsymbol{X} \in \mathrm{RV}_{-\alpha}(\nu)$ or $\mathrm{RV}_{-\alpha}(\boldsymbol{Y})$. Regular variation with respect to another radial function $\tilde{r}(\cdot)$ is equivalent if there exists $c_1, c_2 > 0$ such that $c_1 r(\boldsymbol{x}) \leq \tilde{r}(\boldsymbol{x}) \leq c_2 r(\boldsymbol{x})$ for all $\boldsymbol{x}$, and in this case, the limits are related as follows: $\nu_r(A) = \nu_{\tilde{r}}(c_2 A)/\nu_{\tilde{r}}(c_2 C_r)$.

Suppose that $\boldsymbol{X}$ admits a probability density $f_{\boldsymbol{X}}$ with respect to the Lebesgue measure and let $f_{\boldsymbol{Y}}$ be a probability on C_r such that $f_{\boldsymbol{X}}(\lambda \mathbf{1}), f_{\boldsymbol{Y}}(\lambda \mathbf{1}) > 0$ for all $\lambda \geq 1$. Similarly to what we did in the previous section, we now study the relation between regular variation of $f_{\boldsymbol{X}}$ in the sense (2.4) and of $\boldsymbol{X}$. If $f_{\boldsymbol{X}} \in \mathrm{RV}(v)$ for some nonnull function v and the sequence is dominated by an integrable function, then

$$f_{t^{-1}\boldsymbol{X} \mid \boldsymbol{X} \in tC_r}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}}(t\boldsymbol{x})}{f_{\boldsymbol{X}}(t\mathbf{1})} \left(\int_{C_r} \frac{f_{\boldsymbol{X}}(t\tilde{\boldsymbol{x}})}{f_{\boldsymbol{X}}(t\mathbf{1})} \, \mathrm{d}\tilde{\boldsymbol{x}} \right)^{-1} \to \frac{v(\boldsymbol{x})}{\int_{C_r} v(\tilde{\boldsymbol{x}}) \, \mathrm{d}\tilde{\boldsymbol{x}}} =: f_{\boldsymbol{Y}}(\boldsymbol{x}); \tag{3.3}$$

thus, the limiting probability density $f_{\boldsymbol{Y}}$ is homogeneous and by Scheff's Lemma $\boldsymbol{X} \in \mathrm{RV}_{-\alpha}(\boldsymbol{Y})$ on C_r. As a comparison, [24, Proposition 5.20] has the same conclusion; however, its assumptions require the convergence to be uniform and the limiting density to be bounded on $\{\boldsymbol{x} \in \mathbb{R}^d : r(\boldsymbol{x}) = 1\}$.

Relations (3.2) and (3.3) are easily extended in the case of the general decay T. We write $\boldsymbol{X} \in T\text{–}\mathrm{RV}^x_{-\alpha}(H)$ if $\{T(X_1), \ldots, T(X_d)\} \in \mathrm{RV}^x_{-\alpha}(\nu)$ or, equivalently, $\mu_{t\star} \xrightarrow{\mathrm{w}} \nu$, where $\mu_{\lambda\star}(A) := \mu(\lambda \star A)/\mu(\lambda \star C_r)$ defines a probability distribution on C_r for all $\lambda \in E$. In this case, ν is homogeneous of order $-\alpha$, i.e. $\mu(\lambda \star A) = T(\lambda)^{-\alpha}\mu(A)$ for all $\lambda \in E$.

4. Multivariate exceedances

Extreme-value theory aims at describing the distribution of an $\mathbb{R}^d$-random vector $\boldsymbol{X}$ when at least one of its marginal is large. Typically, it is assumed that $\boldsymbol{X}$ is regularly varying on $C_{||\cdot||_\infty} = \{\boldsymbol{x} \in \mathbb{R}^d : ||\boldsymbol{x}||_\infty\}$, i.e. $t^{-1}\boldsymbol{X} \mid ||\boldsymbol{X}||_\infty \geq t \xrightarrow{\mathrm{w}} \boldsymbol{Y}$, and that $\boldsymbol{Y}$ admits a probability density $f_{\boldsymbol{Y}}$, known to be homogeneous. This excludes the case of asymptotically independent

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

marginals [24], which is to be treated separately. The tail distribution of $\boldsymbol{X}$ is commonly inferred by censoring the marginals whose absolute values fall under some large threshold [20], requiring a model for $f_{\boldsymbol{Y}}$ and its censored versions. In this section we first suggest a general procedure to obtain parametric homogeneous densities whose multivariate marginals and censored densities are known. Second, we explain how to simplify $f_{\boldsymbol{Y}}$ when d is large using graphical models.

We start by asking whether the marginals of a regularly varying vector are also regularly varying and, in this case, if the marginals of the limit are the limits of the marginals. The following lemma provides an answer. Let $A \subseteq \{1, \dots, d\}$, $C^A_{||\cdot||_\infty} := \{\boldsymbol{x}_A \in \mathbb{R}^{|A|} : ||\boldsymbol{x}_A||_\infty \geq 1\}$ and $v \neq 0$.

Lemma 4.1. *Suppose that*

$$\mathbb{P}(bX_i \geq t \mid ||\boldsymbol{X}||_\infty \geq t) \to c_{bi} > 0, \qquad b \in \{-1, 1\},\ i = 1, \dots, d. \tag{4.1}$$

If $\boldsymbol{X} \in \mathrm{RV}_{-\alpha}(\boldsymbol{Y})$ *on* $C_{||\cdot||_\infty}$ *then* $\boldsymbol{X}_A \in \mathrm{RV}_{-\alpha}(\boldsymbol{Y}_A \mid ||\boldsymbol{Y}_A||_\infty \geq 1)$ *on* $C^A_{||\cdot||_\infty}$. *If* $f_{\boldsymbol{X}} \in \mathrm{RV}_{-\alpha}(v)$ *and the sequence is either monotone and* v *is integrable or is dominated by an integrable function, then* $f_{t^{-1}\boldsymbol{X}_A \mid ||\boldsymbol{X}_A||_\infty \geq t} \to f_{\boldsymbol{Y}_A \mid ||\boldsymbol{Y}_A||_\infty \geq 1}$.

This result is specific to the Pareto decay ($T = \mathrm{id}$): $\bar{F}(x, y) = 2/(\mathrm{e}^{|x|} + \mathrm{e}^{|y|})$ is homogeneous on $\mathbb{R}^2$ with respect to addition but its marginal $\bar{F}_X(x) = 2/(\mathrm{e}^{|x|} + 1)$, although regularly varying, is not homogeneous.

Example 4.1. Let $\boldsymbol{X}$ be multivariate Student t-distributed with mean $\boldsymbol{0}$, covariance matrix $(v/(v-2))\Sigma \subset \mathbb{R}^{d\times d}$, and degree of freedom $v > 2$. The sequence $f_{\boldsymbol{X}}(t\boldsymbol{x})/f_{\boldsymbol{X}}(t\boldsymbol{1})$ is monotone and converges to $v(\boldsymbol{x})/v(\boldsymbol{1})$ for $v(\boldsymbol{x}) = (\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x})^{-(v+d)/2}$ which integrates to $c < \infty$ on $C_{||\cdot||_\infty}$. As explained in (3.3), this implies

$$f_{t^{-1}\boldsymbol{X} \mid ||\boldsymbol{X}||_\infty \geq t}(\boldsymbol{y}) \to c^{-1}v(\boldsymbol{y}) =: f_{\boldsymbol{Y}}(\boldsymbol{y}),$$

and thus, $\boldsymbol{X} \in \mathrm{RV}_{-\alpha}(\boldsymbol{Y})$. In addition, the censored limiting density is the limit of the censored density; that is, $f_{t^{-1}\boldsymbol{X}_A, |\boldsymbol{X}_A|<t \mid ||\boldsymbol{X}||_\infty \geq t} \to f_{\boldsymbol{Y}_A, |\boldsymbol{Y}_A|<1}$, and Lemma 4.1 ensures that the marginals of the limit are the limits of the marginals, i.e. $f_{t^{-1}\boldsymbol{X}_A \mid ||\boldsymbol{X}_A||_\infty \geq t} \to f_{\boldsymbol{Y}_A}$. These quantities can thus be computed from the censored densities and marginals of the multivariate Student by passing to the limit. Up to a constant and a transformation of the univariate marginals, $f_{\boldsymbol{Y}}$ coincides with the density of the exponent measure of the extremal-t distribution derived in [26].

As the previous example suggests, we can obtain parametric families for homogeneous distributions by computing limits of well-known multivariate distributions that are regularly varying with decay T. We now focus on a complementary approach for modeling $f_{\boldsymbol{Y}}$ using graphical models. In short, graphical models offer a way to simplify a joint density by assuming some conditional independence between the marginals. Conditional independence—and independence—between marginals is only meaningful on a product space [10]. Since $\boldsymbol{Y}$ has values in $C_{||\cdot||_\infty}$, the range of Y_i depends on the other marginals. To remedy this problem, we work with the censored vector

$$\boldsymbol{Y}^{\mathrm{C}} = \begin{cases} (Y_1 1_{|Y_1|\geq 1}, \dots, Y_d 1_{|Y_d|\geq 1}) & \text{with probability } p, \\ \boldsymbol{0} & \text{with probability } 1-p, \end{cases}$$

whose values lie in $F^d = \{(-\infty, -1] \times \{0\} \times [1, \infty)\}^d$, where $p = \mathbb{P}(||\boldsymbol{X}||_\infty \geq u) \in (0, 1)$ for some large threshold u. This stays in line with statistical inference, which typically estimates a censored version of the limiting distribution.

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

The Hammersley–Clifford theorem states that if $f_{\boldsymbol{Y}^{\mathrm{C}}}$ is positive, it factorizes with respect to a graph $\mathcal{G} = (\{1, \dots, d\}, E)$ if and only if $Y_i^{\mathrm{C}} \perp\!\!\!\perp Y_j^{\mathrm{C}} \mid \boldsymbol{Y}^{\mathrm{C}}_{\{1,\dots,d\}\setminus\{i,j\}}$ for all $(i, j) \notin E$ [19]. The latter denotes conditional independence between Y_i^{C} and Y_j^{C} given the rest of the vector. Let us illustrate an application of the theorem.

Example 4.2. Let $\boldsymbol{Y}$ have values in $\mathbb{R}^3_+$ with

$$f_{\boldsymbol{Y}_{12}\,|\,||\boldsymbol{Y}_{12}||_\infty \geq 1}(x, y) = \tfrac{4}{3}(x + y)^{-3} = f_{\boldsymbol{Y}_{23}\,|\,||\boldsymbol{Y}_{23}||_\infty \geq 1}(x, y) \quad \text{and} \quad p = \mathbb{P}(||\boldsymbol{Y}_{12}||_\infty \geq 1).$$

From the Hammersley–Clifford theorem, $Y_1^{\mathrm{C}} \perp\!\!\!\perp Y_3^{\mathrm{C}} \mid Y_2^{\mathrm{C}}$ if and only if

$$f_{\boldsymbol{Y}^{\mathrm{C}}}(x, y, z) = \frac{f_{\boldsymbol{Y}_{12}^{\mathrm{C}}}(x, y) f_{\boldsymbol{Y}_{23}^{\mathrm{C}}}(y, z)}{f_{Y_2^{\mathrm{C}}}(y)} \quad \text{for all } (x, y, z) \in (\{0\} \cup [1, \infty))^3,$$

where $f_{\boldsymbol{Y}_{12}^{\mathrm{C}}}(x, 0) = \frac{2}{3} p\{x^{-2} - (x + 1)^{-2}\}$, $f_{Y_2^{\mathrm{C}}}(x) = \frac{2}{3} p x^{-2}$, etc.

From a statistical perspective, the distribution of $\boldsymbol{Y}^{\mathrm{C}}$ is to be estimated from samples of $\boldsymbol{X}$. To that end, we consider a sequence $\boldsymbol{X}_t^{\mathrm{C}}$ which is function of $\boldsymbol{X}$ and satisfies $\boldsymbol{X}_t^{\mathrm{C}} \xrightarrow{\mathrm{w}} \boldsymbol{Y}^{\mathrm{C}}$ if $\boldsymbol{X} \in \mathrm{RV}_{-\alpha}(\boldsymbol{Y})$. Crucially, the sequence of marginals $(\boldsymbol{X}_t^{\mathrm{C}})_D$ for $D \subseteq \{1, \dots, d\}$ coincides with the censored sequence $(\boldsymbol{X}_{t,D})^{\mathrm{C}}$ converging to $\boldsymbol{Y}^{\mathrm{C}}_{t,D}$. It is defined for all $\boldsymbol{x} \in \mathbb{R}^{|D\cap A|}$ s.t. $|\boldsymbol{x}| \geq 1$, $A \subseteq D \setminus \varnothing$, by

$$\mathbb{P}(\boldsymbol{\sigma_x} \boldsymbol{X}^{\mathrm{C}}_{t,D\cap A} \geq |\boldsymbol{x}|,\ \boldsymbol{X}^{\mathrm{C}}_{t,D\cap A^{\mathrm{c}}} = \boldsymbol{0}) = p \frac{\mathbb{P}(\boldsymbol{\sigma_x} \boldsymbol{X}_{D\cap A} \geq t|\boldsymbol{x}|,\ |\boldsymbol{X}_{D\cap A^{\mathrm{c}}}| < t)}{\mathbb{P}(||\boldsymbol{X}||_\infty \geq t)},$$

and $\mathbb{P}(\boldsymbol{X}^{\mathrm{C}}_{t,D} = \boldsymbol{0}) = 1 - p\mathbb{P}(||\boldsymbol{X}_D||_\infty \geq t \mid ||\boldsymbol{X}||_\infty \geq t)$, where $\boldsymbol{\sigma_x}$ denotes the sign of $\boldsymbol{x}$. Both $\boldsymbol{X}_t^{\mathrm{C}}$ and $\boldsymbol{Y}^{\mathrm{C}}$ admit probability densities $f_{\boldsymbol{X}_t^{\mathrm{C}}}$ and $f_{\boldsymbol{Y}^{\mathrm{C}}}$ with respect to the measure $\mu_0(A_1 \times \cdots \times A_d) := \sum_{D\subseteq\{1,\dots,d\}} \lambda^{|D|}(\prod_{i\in D} A_i)\delta_{\boldsymbol{0}}^{|D^{\mathrm{c}}|}(\prod_{i\in D^{\mathrm{c}}} A_i)$ on F^d, where λ^d is the d-dimensional Lebesgue measure and $\delta_{\boldsymbol{0}}^d$ the Dirac measure on $\boldsymbol{0} \in \mathbb{R}^d$.

We want to find conditions on $\boldsymbol{X}$ that impose a factorization of $f_{\boldsymbol{Y}^{\mathrm{C}}}$. This brings us to extend the Hammersley–Clifford theorem for densities that factorize in the limit. Conditional independence is translated into *asymptotic conditional independence*, a notion little mentioned in the literature.

5. Asymptotic graphical models

Let $\boldsymbol{X}_t$ for $t \geq 1$ and $\boldsymbol{X}$ be $\mathbb{R}^d$-valued random vectors with almost everywhere (a.e.) continuous probability densities $f_{\boldsymbol{X}_t}$ and $f_{\boldsymbol{X}}$ with respect to a base measure μ_0, typically the Lebesgue, the counting measure, or a combination of the two. For disjoint subsets $A, B, C \subseteq \{1, \dots, d\}$, we call the marginals $\boldsymbol{X}_{t,A}$ and $\boldsymbol{X}_{t,B}$ *asymptotically conditionally independent* with respect to $\boldsymbol{X}_{C,t}$, written $\boldsymbol{X}_{t,A} \tilde{\perp\!\!\!\perp} \boldsymbol{X}_{t,B} \mid \boldsymbol{X}_{t,C}$, if

$$(f_{\boldsymbol{X}_{t,ABC}} f_{\boldsymbol{X}_{t,C}} - f_{\boldsymbol{X}_{t,AC}} f_{\boldsymbol{X}_{t,BC}})\,\mathrm{d}\mu_0 \xrightarrow{\mathrm{w}} 0. \tag{5.1}$$

Here, '$f_{\boldsymbol{X}_t}\,\mathrm{d}\mu_0 \xrightarrow{\mathrm{w}}$' stands for weak convergence of $\boldsymbol{X}_t$, i.e. the convergence of $\int g f_{\boldsymbol{X}_t}\,\mathrm{d}\mu_0$ for every g a.e. continuous and bounded. When $\boldsymbol{X}_t = \boldsymbol{X}_1$ for all t, (5.1) coincides with conditional independence of random variables [19].

Let $\mathcal{G} = (V, E)$ be an undirected graph with set of nodes $V = \{1, \dots, d\}$ and set of edges E. We say that $\boldsymbol{X}_t$ satisfies the *asymptotic Markov properties* according to $\mathcal{G}$ if $\{i, j\} \notin E$ implies $X_{t,i} \tilde{\perp\!\!\!\perp} X_{t,j} \mid \boldsymbol{X}_{t,V\setminus\{i,j\}}$. Moreover, we say that $f_{\boldsymbol{X}_t}$ *asymptotically factorizes* with respect to $\mathcal{G}$ if $\boldsymbol{X}_t \xrightarrow{\mathrm{w}} \boldsymbol{X}$ and $f_{\boldsymbol{X}}$ factorizes with respect to $\mathcal{G}$.

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

Theorem 5.1. (Asymptotic Hammersley–Clifford theorem.) *Suppose that* $\boldsymbol{X}_t \xrightarrow{\mathrm{w}} \boldsymbol{X}$, $f_{\boldsymbol{X}} > 0$, *and* $f_{\boldsymbol{X}_A}$ *bounded for all* $A \subseteq V$ *a.e. The error* $\varepsilon_t := f_{\boldsymbol{X}_t} - f_{\boldsymbol{X}}$ *satisfies*

$$(\varepsilon_{t,V}\varepsilon_{t,V\setminus\{i,j\}} - \varepsilon_{t,V\setminus\{i\}}\varepsilon_{t,V\setminus\{j\}})\,\mathrm{d}\mu_0 \xrightarrow{\mathrm{w}} 0, \quad \textit{for all } (i,j) \in E, \tag{5.2}$$

if and only if the following are equivalent:

(i) $f_{\boldsymbol{X}_t}$ *asymptotically factorizes with respect to* $\mathcal{G}$;

(ii) $\boldsymbol{X}$ *satisfies the Markov properties according to* $\mathcal{G}$;

(iii) $\boldsymbol{X}_t$ *satisfies the asymptotic Markov properties according to* $\mathcal{G}$.

Proof. (i) if and only if (ii). Apply the Hammersley–Clifford theorem.

(ii) if and only if (iii). It suffices to show relations of the form $f_{XY\mathbf{Z}}f_{\mathbf{Z}} = f_{X\mathbf{Z}}f_{Y\mathbf{Z}}$ a.e. if and only if $(f_{X_tY_t\mathbf{Z}_t}f_{\mathbf{Z}_t} - f_{X_t\mathbf{Z}_t}f_{Y_t\mathbf{Z}_t})\,\mathrm{d}\mu_0 \xrightarrow{\mathrm{w}} 0$. Write the latter as

$$\begin{aligned}&(f_{XY\mathbf{Z}}f_{\mathbf{Z}} - f_{X\mathbf{Z}}f_{Y\mathbf{Z}})d\mu_0 + (\varepsilon_{X_t,Y_t,\mathbf{Z}_t}\varepsilon_{\mathbf{Z}_t} - \varepsilon_{X_t,\mathbf{Z}_t}\varepsilon_{Y_t,\mathbf{Z}_t})\,\mathrm{d}\mu_0\\ &+ (\varepsilon_{X_t,Y_t,\mathbf{Z}_t}f_{\mathbf{Z}} + \varepsilon_{\mathbf{Z}_t}f_{X,Y,\mathbf{Z}} - \varepsilon_{X_t,\mathbf{Z}_t}f_{Y,\mathbf{Z}} - \varepsilon_{Y_t,\mathbf{Z}_t}f_{X,\mathbf{Z}})\,\mathrm{d}\mu_0\\ &\quad\xrightarrow{\mathrm{w}} 0.\end{aligned} \tag{5.3}$$

Since $(X_t, Y_t, \mathbf{Z}_t) \xrightarrow{\mathrm{w}} (X, Y, \mathbf{Z})$, ε_t converges weakly to 0 and so does the last term in (5.3) since $f_{X,Y,\mathbf{Z}}$ and its marginals are a.e. continuous and bounded. Moreover, the middle term vanishes by (5.2), proving the equivalence. □

As mentioned in [19, Example 3.11], the Markov property is in general not satisfied under weak convergence; Theorem 5.1 provides conditions under which it is. If $f_{\boldsymbol{X}_t} \to f_{\boldsymbol{X}}$ and the sequence is dominated by an integrable function, then $f_{\boldsymbol{X}_{t,A}} \to f_{\boldsymbol{X}_A}$ for all $A \subseteq \{1,\dots,d\}$. Hence, $f_{\boldsymbol{X}}$ need not be bounded and, crucially, (5.2) holds directly. Similarly, if $f_{\boldsymbol{X}_t} \to f_{\boldsymbol{X}}$ in L_2, then $f_{\boldsymbol{X}_{t,A}} \to f_{\boldsymbol{X}_A}$ in L_2 by Jensen's inequality, and we apply Hölder's inequality to show that (5.2) is satisfied.

6. Asymptotic graphical modeling of exceedances

We explain how to model the exceedances of a random vector $\boldsymbol{X}$ with probability density $f_{\boldsymbol{X}}$ using asymptotic graphical models. Let $\mathcal{G}$ be a decomposable graph with set of maximal cliques $\mathcal{C}$. Suppose that $f_{\boldsymbol{X}} \in \mathrm{RV}_{-\alpha}(v)$ for $v \neq 0$ and that the sequence is either monotone and v is integrable or dominated by an integrable function. As seen in (3.1), $f_{t^{-1}\boldsymbol{X}\,|\,||\boldsymbol{X}||_\infty \geq t} \to f_{\boldsymbol{Y}}$ and we assume that the probability density $f_{\boldsymbol{Y}}$ is positive. It follows that $f_{\boldsymbol{X}_t^{\mathrm{C}}} \to f_{\boldsymbol{Y}^{\mathrm{C}}} > 0$, where $\boldsymbol{X}_t^{\mathrm{C}}$ and $\boldsymbol{Y}^{\mathrm{C}}$ are defined in Section 4. From Theorem 5.1, $\boldsymbol{X}_t^{\mathrm{C}}$ satisfies the asymptotic Markov properties according to $\mathcal{G}$ if and only if $f_{\boldsymbol{Y}^{\mathrm{C}}}$ factorizes with respect to $\mathcal{G}$, i.e.

$$f_{\boldsymbol{Y}^{\mathrm{C}}}(\boldsymbol{y}) = \frac{\prod_{C\in\mathcal{C}} f_{\boldsymbol{Y}_C^{\mathrm{C}}}(\boldsymbol{y}_C)}{\prod_{D\in\mathcal{D}} f_{\boldsymbol{Y}_D^{\mathrm{C}}}(\boldsymbol{y}_D)}, \qquad \boldsymbol{y} \in F^d,$$

where $\mathcal{D}$ is a multiset containing intersections between some of the maximal cliques. Hence, for all $A \subseteq \{1,\dots,d\}$ and $\boldsymbol{y}_A$ s.t. $|\boldsymbol{y}_A| \geq \mathbf{1}$,

$$f_{\boldsymbol{Y}_A,|\boldsymbol{Y}_{A^{\mathrm{C}}}|<1}(\boldsymbol{y}_A) = \frac{\prod_{C\in\mathcal{C}:\,C\cap A\neq\varnothing} f(\boldsymbol{y}_{C\cap A})}{\prod_{D\in\mathcal{D}:\,D\cap A\neq\varnothing} f(\boldsymbol{y}_{D\cap A})}\frac{\prod_{C\in\mathcal{C}} q_{C,A}}{\prod_{D\in\mathcal{C}} q_{D,A}}, \tag{6.1}$$

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

where $f(\boldsymbol{y}_{S\cap A})$ denotes the density of $\boldsymbol{Y}_{S\cap A}, \boldsymbol{Y}_{S\cap A^c} < 1 \mid ||\boldsymbol{Y}_S||_\infty \geq 1$ and

$$q_{S,A} = \begin{cases} p_S := p\mathbb{P}(||\boldsymbol{Y}_S||_\infty \geq 1) & \text{if } S \cap A \neq \varnothing, \\ 1 - p_S & \text{otherwise,} \end{cases} \quad \text{for all } S \in \mathcal{C} \cup \mathcal{D}.$$

By positivity of $f_{\boldsymbol{Y}}$, (4.1) is satisfied and, thus, Lemma 4.1 gives $\boldsymbol{X}_S \in \mathrm{RV}_{-\alpha}(\boldsymbol{Y}_S \mid ||\boldsymbol{Y}_S||_\infty \geq 1)$ on $C^S_{||\cdot||_\infty}$ with censored limiting density $f(\boldsymbol{y}_{S\cap A})$. This is a key step as it means that the latter can be estimated from samples of $\boldsymbol{X}_S$.

This suggests the following strategy to model exceedances of an independent and identically distributed sample of $\boldsymbol{X}$. First, transform the positive and negative part of each marginal that exceeds a large threshold to unit Pareto to obtain the censored vector $\boldsymbol{X}^{\mathrm{C}}_1$. Second, select a decomposable graph $\mathcal{G}$ by testing conditional independence between the marginals of $\boldsymbol{X}^{\mathrm{C}}_1 = (X_1 1_{|X_1|\geq 1}, \ldots, X_d 1_{|X_d|\geq 1})$. Third, for all $S \in \mathcal{C} \cup \mathcal{D}$, select a model for the homogeneous density $f_{\boldsymbol{Y}_S \mid ||\boldsymbol{Y}_S||\geq 1}$ with unit Pareto marginals; estimate it using a sample from $\boldsymbol{X}^{\mathrm{C}}_{1,S}$, and estimate $p_S \approx \mathbb{P}(||\boldsymbol{X}_{1,S}||_\infty \geq 1)$. This determines the asymptotic approximation of $f_{\boldsymbol{X}^{\mathrm{C}}_1}$ in (6.1).

In low dimensions, there is a rich class of models for extreme-value distributions [15]. If the later have Fréchet marginals, we have seen that their exponent measure density, when it exists, is homogeneous and, thus, provides a model for $f_{\boldsymbol{Y}_S \mid ||\boldsymbol{Y}_S||_\infty \geq 1}$. In a parametric approach, a refinement is to impose the parameters of $f_{\boldsymbol{Y}_C}(\cdot; \boldsymbol{\theta}_C)$ to be consistent with the ones of $f_{\boldsymbol{Y}_D}(\cdot; \boldsymbol{\theta}_D)$ whenever $D \subset C$ and to estimate them from sufficient statistics; Example 4.1 showed a way to find consistent parametric families. An extreme-value distribution often used in practice is the Hüsler–Reiss [13], [17].

Example 6.1. The Hüsler–Reiss exponent measure has homogeneous probability density

$$f_{\boldsymbol{Y} \mid Y_k \geq 1}(\boldsymbol{y}) = y_k^{-2} \phi(\boldsymbol{y}_{-k} \mid y_k) \quad \text{on} \quad C_k = \{\boldsymbol{y} \geq \boldsymbol{0} \colon y_k \geq 1\} \quad \text{for all } k = 1, \ldots, d,$$

where ϕ is the density of the multivariate log-normal distribution on $\mathbb{R}^{d-1}$ with mean $(\log y_k - \frac{1}{2}\Gamma_{jk})_{j\neq k}$ and covariance matrix Σ_ϑ^{-1}, where $\Sigma_\vartheta = \frac{1}{2}\{\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}\}_{2\leq i,\, j\leq d}$, and $\Gamma_{ij} = \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{i,j}$ is the incremental variance defined by a correlation matrix $\Sigma \subset \mathbb{R}^{d\times d}$. The marginals $\boldsymbol{Y}_A \mid ||\boldsymbol{Y}_A||_\infty \geq 1$ and the censored densities can be computed. Moreover, conditional independence corresponds to a specific constraint on the covariance matrix, analogously to the Gaussian case:

$$Y_i \perp\!\!\!\perp Y_j \mid \boldsymbol{Y}_{\{1,\ldots,d\}\setminus\{i,j\}} \text{ on } C_k \quad \Longleftrightarrow \quad \Sigma_{\vartheta,ij} = \Gamma_{ij} - \Gamma_{ik} - \Gamma_{jk} = 0. \tag{6.2}$$

Asymptotic graphical modeling for extremes necessitates testing asymptotic conditional independence, which is difficult in a nonparametric setting. As a first approximation, we can easily test conditional independence of the binary vector $\boldsymbol{B} = (1_{|X_1|\geq t}, \ldots, 1_{|X_d|\geq t})$ or a more refined discretization [23]. Further efforts are needed to build a test for tail conditional independence in the parametric continuous case—for instance, by relying upon (6.2). Lastly, we mention the success of the Gaussian graphical lasso [14], which has been able to tackle high-dimensional problems by imposing sparsity of the inverse covariance matrix. Future work could allow a similar approach for the Hüsler–Reiss distribution.

Acknowledgements

The first author is grateful to Anthony Davison for suggesting graphical modeling of extreme-value distributions, Richard Davis who proposed the notion of asymptotic conditional independence, Thomas Mikosch for a stimulating lecture on regular variation, Phyllis Wan, Sebastian

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press

Engelke for helpful discussions, and the anonymous referee for their pertinent comments. He would also like to thank the Marquise and Marquis de Amodio for their funding.

References

[1] Ballani, F. and Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika* **98,** 633–645.
[2] Basrak, B. (2000). The sample autocorrelation function of non-linear time series. *Doctoral Thesis*, University of Groningen.
[3] Basrak, B., Davis, R. A. and Mikosch, T. (2002). A characterization of multivariate regular variation. *Ann. Appl. Prob.* **12,** 908–920.
[4] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley, Chichester.
[5] Billingsley, P. (1995). *Probability and Measure*, 3rd edn. John Wiley, New York.
[6] Bingham, N. H. and Ostaszewski, A. J. (2010). Topological regular variation: III. Regular variation. *Topology Appl.* **157,** 2024–2037.
[7] Bingham, N. H., Goldie, C. M. and Teugles, J. L. (1989). *Regular Variation* (Encyclopedia Math. Appl. **27**). Cambridge University Press.
[8] Boldi, M.-O. and Davison, A. C. (2007). A mixture model for multivariate extremes. *J. R. Statist. Soc. B* **69,** 217–229.
[9] Cooley, D., Davis, R. A. and Naveau, P. (2010). The pairwise beta distribution: a flexible parametric multivariate model for extremes. *J. Multivariate Anal.* **101,** 2103–2117.
[10] Dawid, A. P. (2001). Separoids: a mathematical framework for conditional independence and irrelevance. Representations of uncertainty. *Ann. Math. Artif. Intell.* **32,** 335–372.
[11] De Haan, L. and Resnick, S. (1987). On regular variation of probability densities. *Stoch. Process. Appl.* **25,** 83–93.
[12] Durrett, R. (2010). *Probability: Theory and Examples*, 4th edn. Cambridge University Press.
[13] Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Statist. Soc. B* **77,** 239–265.
[14] Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9,** 432–441.
[15] Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications*, Springer, Heidelberg, pp. 127–145.
[16] Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Prob.* **17,** 537–571.
[17] Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statist. Prob. Lett.* **7,** 283–286.
[18] Jaroš, J. and Kusano, T. (2004). Self-adjoint differential equations and generalized Karamata functions. *Bull. Cl. Sci. Math. Nat. Sci. Math.* **29,** 25–60.
[19] Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.
[20] Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83,** 169–187.
[21] Lindskog, F., Resnick, S. I. and Roy, J. (2014). Regularly varying measures on metric spaces: hidden regular variation and hidden jumps. *Prob. Surv.* **11,** 270–314.
[22] Meerschaert, M. M. (1993). Regular variation and generalized domains of attraction in $\mathbb{R}^k$. *Statist. Prob. Lett.* **18,** 233–239.
[23] Nagarajan, R., Scutari, M. and Lèbre, S. (2013). *Bayesian Networks in R with Applications in Systems Biology*. Springer, New York.
[24] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes* (Appl. Prob. Ser. Appl. Prob. Trust **4**). Springer, New York.
[25] Resnick, S. I. and Zeber, D. (2014). Transition kernels and the conditional extreme value model. *Extremes* **17,** 263–287.
[26] Ribatet, M. (2013). Spatial extremes: max-stable processes at work. *J. Soc. Française Statist.* **154,** 156–177.
[27] Seneta, E. (1973). An interpretation of some aspects of Karamata's theory of regular variation. *Publ. Inst. Math. (Beograd) (N.S.)* **15,** 111–119.
[28] Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77,** 245–253.
[29] Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1,** 1–305.

https://doi.org/10.1017/jpr.2016.37 Published online by Cambridge University Press