CrossMark

Adv. Appl. Prob. **52,** 855–878 (2020)
*doi:*10.1017/apr.2020.22
© *Applied Probability Trust* 2020

ONE- VERSUS MULTI-COMPONENT REGULAR VARIATION AND EXTREMES OF MARKOV TREES

JOHAN SEGERS,* *Université catholique de Louvain*

Abstract

A Markov tree is a random vector indexed by the nodes of a tree whose distribution is determined by the distributions of pairs of neighbouring variables and a list of conditional independence relations. Upon an assumption on the tails of the Markov kernels associated to these pairs, the conditional distribution of the self-normalized random vector when the variable at the root of the tree tends to infinity converges weakly to a random vector of coupled random walks called a tail tree. If, in addition, the conditioning variable has a regularly varying tail, the Markov tree satisfies a form of one-component regular variation. Changing the location of the root, that is, changing the conditioning variable, yields a different tail tree. When the tails of the marginal distributions of the conditioning variables are balanced, these tail trees are connected by a formula that generalizes the time change formula for regularly varying stationary time series. The formula is most easily understood when the various one-component regular variation statements are tied up into a single multi-component statement. The theory of multi-component regular variation is worked out for general random vectors, not necessarily Markov trees, with an eye towards other models, graphical or otherwise.

Keywords: Conditional independence; graphical model; Hüsler–Reiss distribution; max-linear model; Markov tree; multivariate Pareto distribution; Pickands dependence function; regular variation; root change formula; tail measure; tail tree; time change formula

1. Introduction

Imagine a random vector $X = (X_1, \ldots, X_d)$ of nonnegative variables. One of the components, say X_i, is known to have exceeded a large threshold. How does this information affect the conditional distribution of the whole vector X? There could be a causal link from X_i to the other variables X_j, perhaps via a network of dependence relations, so that tampering with X_i would affect the whole system. Another possibility is that a large value of X_i is merely the result of a large value of some other variable X_j. The latter event, however, could have consequences for still other variables X_k.

Depending on which one of the d components is known to have been exceptionally large, the conditional distribution of X is likely to be different. Still, if high values of two variables X_i and X_j are not unlikely to arrive together, the conditional distribution of X given that X_i is large must be connected to the one given that X_j is large.

Received 6 February 2019; revision received 2 March 2020; accepted 10 March 2020.
* Postal address: Université catholique de Louvain, LIDAM/ISBA, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. Email: johan.segers@uclouvain.be

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

In this paper, these questions are studied for general random vectors using the language of regular variation. The answers are worked out for the particular case that X is a Markov tree. A large value at a particular node is found to spread through the tree via independent increments along the edges. The joint limit distribution is that of a vector of coupled geometric random walks. The couplings occur through the common edges of different paths starting at the same root node.

Graphical models, of which Markov trees are a special case, bring structure and sparsity to the web of dependence relations between many random variables [23] and [28]. Extreme value theory for such models is a fairly recent subject. In [1], a metric that takes the distance along a river into account underlies a spatial model for extremes of river networks. Recursive max-linear models on directed acyclic graphs are proposed in [13] and put to work in [9] and [14]. In [17], the density of a multivariate Pareto distribution is factorized through a version of the Hammersley–Clifford theorem. Such factorizations are also the theme in [10], where they form the basis of new inference methods for extremes of graphical models, including the identification of the graphical structure itself. Multivariate Hüsler–Reiss extreme-value copulas based on Gaussian Markov trees and higher-order truncated vines are introduced in [24], which proposes composite likelihood methods based on bivariate margins to estimate the parameters.

Multivariate Pareto distributions arise as weak limits of normalized random vectors conditionally on the event that at least one component exceeds a high threshold. Although such conditioning events are covered by Theorem 4 below, the focus of this paper is rather on the case where the exceedance is known to have occurred at a specific variable. The message hinted at in the title is that both points of view are mathematically equivalent, but that, at least for Markov trees, the one-component limit is particularly elegant, as will be explained next.

1.1. Tail tree of a Markov tree

For a Markov chain, it was discovered in [35] that, conditionally on the event that the series is large at some time instant, the conditional distribution of the future of the system is that of a random walk, a process called a tail chain in [26]. For light-tailed marginal distributions, this random walk is additive, and for heavy-tailed margins it is geometric, i.e., multiplicative, which is the convention used in this paper.

A Markov tree can be viewed as a coupled collection of Markov chains with common stretches. Take for instance the four-variate Markov tree in Figure 1. The nodes of the tree are $\{1, 2, 3, 4\}$ and the three pairs of neighbours are $\{1, 2\}$, $\{2, 3\}$, and $\{2, 4\}$. The vector (X_1, X_2, X_3) is a Markov chain, and so is (X_1, X_2, X_4). These two chains are coupled via the common pair (X_1, X_2). Conditionally on X_2, the variables X_1, X_3, and X_4 are independent, since any path that connects two of the three nodes 1, 3, and 4 passes through node 2. This conditional independence property together with the distributions of the three pairs (X_1, X_2), (X_2, X_3), and (X_2, X_4) determines the joint distribution of (X_1, X_2, X_3, X_4).

For the moment, assume that the four variables have the same, regularly varying tail function. The set-up involving regular variation will be further motivated in Section 1.2. The effect (not necessarily causal) on X_2 of a large value at X_1 is via a multiplicative increment $M_{1,2}$ whose distribution is equal to the weak limit of X_2/X_1 conditionally on $X_1 = t$ as $t \to \infty$. The existence of this limit is an assumption on the Markov kernel induced by the distribution of the pair (X_1, X_2). Similarly, a large value at X_2 affects X_3 and X_4 via the increments $M_{2,3}$ and $M_{2,4}$, respectively. The effect of X_1 on X_3 is then through the composite increment $M_{1,2}M_{2,3}$, whereas on X_4 it is through $M_{1,2}M_{2,4}$. The conditional independence property ensures that the increments $M_{1,2}$, $M_{2,3}$, and $M_{2,4}$ are mutually independent. The common edge (1, 2) on

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

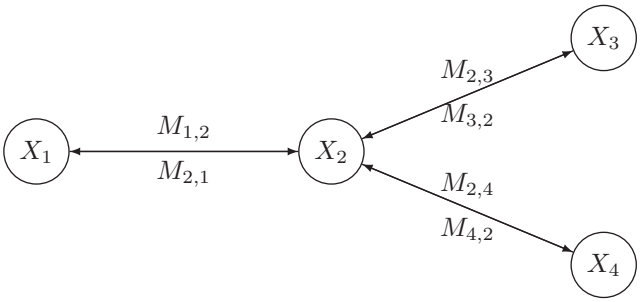


FIGURE 1: A four-variate Markov tree $(X_1, \ldots, X_4)$ and the weak limits $M_{a,b}$ of X_b/X_a given $X_a = t$ as $t \to \infty$ for edges (a, b) in the tree. These limits constitute the multiplicative increments of the tail tree Θ_u starting at a given root, for instance at node $u = 1$ in (1) or at node $u = 3$ in (2).

the paths from node 1 to node 3 and from node 1 to node 4 induces dependence between the two tail chains $(M_{1,2}, M_{1,2}M_{2,3})$ and $(M_{1,2}, M_{1,2}M_{2,4})$ via the common increment $M_{1,2}$. In this paper, the random vector

$$(\Theta_{1,2}, \Theta_{1,3}, \Theta_{1,4}) = (M_{1,2}, M_{1,2}M_{2,3}, M_{1,2}M_{2,4}) \tag{1}$$

is called the *tail tree* induced by X with root at node $u = 1$.

The tail tree represents a network of stochastic dependence relations that are not necessarily causal. Suppose the Markov tree in Figure 1 represents water levels at four locations on a river network. If water flows from left to right, node 2 represents a point where the stream branches into two channels, as occurs for instance in a river delta. If water flows from right to left, however, node 2 represents the junction of two branches coming from nodes 3 and 4 into a larger stream flowing towards node 1. In the first case, the tail tree describes how a high water level at the upstream node 1 may cause high water levels at various locations in the delta further downstream. In the second case, however, it is nodes 3 and 4 that are situated upstream, and the tail tree models the sources of a high water volume at the downstream site 1. Still other set-ups are possible, such as for instance node 3 being upstream and nodes 1 and 4 being downstream: high water levels at nodes 1 and 4 are then related through a common cause at node 2, which can itself perhaps be traced back to node 3.

Whatever the causal relationships within X, it may make sense to change the conditioning variable. In Figure 1, for instance, suppose it is known that a large value has occurred at node 3 rather than at node 1. Tracing the paths from node 3 to the three other nodes yields the tail tree with root at node $u = 3$:

$$(\Theta_{3,2}, \Theta_{3,1}, \Theta_{3,4}) = (M_{3,2}, M_{3,2}M_{2,1}, M_{3,2}M_{2,4}). \tag{2}$$

The tail trees in (1) and (2) have a similar structure. The two edges on the path between the root nodes 1 and 3 have changed direction, however. The edge from node 2 to node 4 is common to both tail trees.

For each pair $\{a, b\}$ of neighbouring nodes, the choice of the root node u determines which of the two increments appears in the tail tree: $M_{a,b}$ from X_a to X_b or $M_{b,a}$ from X_b to X_a. The distributions of $M_{a,b}$ and $M_{b,a}$ are connected by an expression that involves the marginal distributions of X_a and X_b. For stationary and reversible Markov chains, this relation underlies a sufficiency property discovered in [4]. For tail chains of not necessarily reversible Markov

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

chains, it was described in [21] and [33] and for tail processes of regularly varying stationary time series in [2] via the time change formula. This formula can be understood most easily through the connection between the tail process and the tail measure [7], [27], [32], and this is also the way in which the root change formula in Corollary 2 below will be derived, but there without the assumption of stationarity and for general random vectors, not necessarily Markov trees.

1.2. Regular variation

The language of regularly varying functions and measures provides a rich medium through which to express limit theorems. Recall that a positive, Lebesgue measurable function f defined on a neighbourhood of infinity is regularly varying with index $\tau \in \mathbb{R}$ if $\lim_{t\to\infty} f(\lambda t)/f(t) = \lambda^\tau$ for all $\lambda \in (0, \infty)$. If X is a nonnegative random variable with unbounded support, cumulative distribution function $F(x) = \mathbb{P}(X \leqslant x)$, and tail function $\overline{F} = 1 - F$, then regular variation of $\overline{F}$ with index $-\alpha < 0$ is equivalent to weak convergence of the conditional distribution of X/t given that $X > t$ to a Pareto random variable Y with index α (i.e., $\mathbb{P}(Y > y) = y^{-\alpha}$ for all $y \in [1, \infty)$). We write $\mathcal{L}(X/t \mid X > t) \xrightarrow{d} \mathrm{Pa}(\alpha)$ as $t \to \infty$, where $\mathcal{L}(Z \mid A)$ denotes the conditional distribution of the random object Z given the event A, the arrow $\xrightarrow{d}$ denotes convergence in distribution, and $\mathrm{Pa}(\alpha)$ denotes the said Pareto distribution.

For multivariate distributions, regular variation can be described via multivariate cumulative distribution functions as well, but an approach via convergence of Borel measures is more versatile. Let the state space be $\mathbb{S} = [0, \infty)^d$. (Generalizations to star-shaped metric spaces or abstract cones as in [7], [18], [25], and [34] are left for further work.) Let $I \subset \{1, \ldots, d\}$ denote the non-empty set of indices i of variables for which the conditioning event $X_i > t$ is of possible interest. The marginal distributions of X_i for $i \in I$ are assumed to be regularly varying and the ratios of their tail functions are assumed to converge to positive constants. This set-up is a bit more general than the one of identical margins and comes at little technical or notational cost.

The measures involved may have infinite mass but need to assign finite values to sets that remain bounded away from $\{x \in \mathbb{S} : \forall i \in I, x_i = 0\}$ or $\{x \in \mathbb{S} : x_i = 0\}$, depending on the conditioning event. The topology on the space of such measures will be the one proposed in [25], extending [18], and resembles that of vague convergence of measures, but avoiding the need to consider artificially compactified spaces. Regular variation is defined as convergence of $b(t)\,\mathbb{P}(X/t \in \cdot\,)$ to a limit measure called tail measure. Here, $b(t) > 0$ is a scale function tending to infinity and calibrated to the marginal distributions of X_i for $i \in I$.

It is instructive to formulate statements in terms of weak convergence of distributions. For a high threshold t tending to infinity and for a component $i \in I$, consider the asymptotic distribution of the rescaled random vector X/t given that $X_i > t$. Decompose X/t as $(X_i/t, X/X_i)$. Here, X_i/t represents the overall level of X with respect to t, whereas X/X_i represents a self-normalized version of X. Convergence in distribution of $(X_i/t, X/X_i)$ given $X_i > t$ as $t \to \infty$ is a special case of what is called one-component regular variation in [17], explored already in [16] and [30] for the bivariate case but allowing for affine normalizations. The random variable X_i/t is asymptotically $\mathrm{Pa}(\alpha)$ distributed and independent of X/X_i, whose weak limit, denoted by $\Theta_i = (\Theta_{i,j})_{j=1}^d$, captures extremal dependence within X given that X_i is large. Letting the index i run through I produces multiple such one-component regular variation statements, which, together, are equivalent to what can be called multi-component regular variation. The limit distributions Θ_i that arise for various indices i must be mutually consistent, and the tail measure mentioned at the end of the previous paragraph embraces them all at once.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

In Section 3, the focus is on tying together multiple one-component regular variation limits. The theory is worked out for general random vectors, not necessarily Markov trees. A number of results in that section have already been formulated in the literature in one way or another, in slightly different settings. Some of the equivalence relations in Theorem 3, for instance, resemble those in [17, Theorem 1.4] and [34, Proposition 3.1]. The model consistency property between limit measures in Theorem 2(ii) is formulated in [6, Section 2] for the bivariate case. The root change formula in Corollary 3 extends the time change formula for regularly varying stationary time series stemming from [2] and studied extensively in [7] and [20]. Multivariate Pareto distributions as in Theorem 3 are foreshadowed in [29, Section 6.3]; they appear in [11] and [31] when $\rho(x) = \max(x_1, \ldots, x_d)$ and in [8] for more general functionals ρ. These are just a few connections, and the above list is by no means intended to be complete.

The set-up involving regular variation is intended to serve two purposes: first, to model tail dependence within a vector of random variables which have been transformed to the same, heavy-tailed distribution, such as the unit-Fréchet distribution, as is common in multivariate extreme value theory; and second, to model the joint distribution of a vector of regularly varying random variables, not necessarily identically distributed, but with equivalent tails, such as returns on financial portfolios composed of the same basket of underlying assets. The latter framework is more general than the former and comes at little additional notational cost.

1.3. Outline

For a Markov tree X, convergence as $t \to \infty$ of the conditional distribution of X/X_u given that $X_u = t$ is proved in Section 2. The main assumption is that, for edges $e = (a, b)$ directed away from the root u, the conditional distribution of X_b/X_a given $X_a = t$ converges as $t \to \infty$. No regular variation is needed yet.

The tail trees pertaining to different roots u can be linked up thanks to the theory of one- and multi-component regular variation developed in Section 3. The results do not rely on the Markov property and cover quite general random vectors X on $[0, \infty)^d$, as is illustrated briefly for max-linear models. An interesting special case of these is that of the recursive max-linear structural equation models introduced in [13], featuring a causal structure induced by a directed acyclic graph. Most of the proofs of this section are deferred to the appendix.

When combined, the results in Sections 2 and 3 serve to uncover the regular variation properties of Markov trees in Section 4. The common special case that the joint distribution of the Markov tree is absolutely continuous with respect to Lebesgue measure is the subject of Section 5. The theory then simplifies considerably and the limit distribution with respect to a single root u is already sufficient to reconstruct the limit distributions with respect to all other possible roots $\bar{u}$.

In Sections 4 and 5, the distributions of the increments of the tail trees are calculated in case the pair distributions are max-stable, not necessarily absolutely continuous. For the Hüsler–Reiss max-stable distribution, the tail tree is multivariate log-normal, constructed from partial sums of independent normal random variables along the edges of the tree.

2. The spectral tail tree of a Markov tree

A (finite) graph is a pair (V, E) where V is a non-empty finite set of vertices or nodes and where $E \subset V \times V$ is a set of edges. Self-loops are excluded; i.e., $(u, u) \notin E$ for all $u \in V$. To avoid trivialities, V is assumed to have at least two elements. Two nodes are neighbours if they are joined by an edge. A graph is undirected if $(a, b) \in E$ implies $(b, a) \in E$. A path from a node u to a node v is a collection $\{e_1, \ldots, e_n\} \subset E$ of edges such that $e_k = (u_{k-1}, u_k)$ for

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

all $k = 1, \ldots, n$, for $n+1$ *distinct* nodes $u_0, u_1, \ldots, u_n \in V$ such that $u_0 = u$ and $u_n = v$. An undirected tree $\mathcal{T} = (V, E)$ is an undirected graph such that for any pair of distinct nodes u and v, there exists a unique path from u to v; this path is then denoted by $[u \rightsquigarrow v]$.

Let $\mathcal{T} = (V, E)$ be an undirected tree and let $X = (X_v)_{v \in V}$ be a random vector indexed by the nodes of the tree. The pair $(X, \mathcal{T})$ is a Markov tree if it satisfies the global Markov property [23]: whenever A, B, S are disjoint, non-empty subsets of V such that S separates A and B (i.e., any path between a node $a \in A$ and a node $b \in B$ passes through some node in S), the conditional independence relation

$$X_A \perp\!\!\!\perp X_B \mid X_S \tag{3}$$

holds, where X_W denotes the random vector $(X_v)_{v \in W}$ for $W \subset V$.

For an undirected tree $\mathcal{T} = (V, E)$ and a node $u \in V$, let $\mathcal{T}_u = (V, E_u)$ denote the directed, rooted tree formed by directing the edges in E outward starting from u. Formally, E_u is the subset of E that is obtained by choosing from every pair of edges (a, b) and (b, a) in E the one such that the first node separates the second one from u. If $(a, b) \in E_u$, then a is the (necessarily unique) parent of b in $\mathcal{T}_u$, whereas b is a child of a in $\mathcal{T}_u$.

Let $(X, \mathcal{T})$ be a nonnegative Markov tree, where $\mathcal{T} = (V, E)$ is an undirected tree.

Condition 1. There exists $u \in V$ with the following two properties.

(i) For every directed edge $e = (a, b) \in E_u$, there exists a version of the conditional distribution of X_b given X_a and a probability measure μ_e on $[0, \infty)$ such that

$$\mathcal{L}(X_b / x_a \mid X_a = x_a) \xrightarrow{d} \mu_e, \qquad x_a \to \infty. \tag{4}$$

(ii) For edges $e = (a, b) \in E_u$ such that $a \neq u$ and such that there exists an edge $\bar{e} \in [u \rightsquigarrow a]$ for which $\mu_{\bar{e}}(\{0\}) > 0$, we have

$$\forall \eta > 0, \qquad \lim_{\delta \downarrow 0} \limsup_{x \to \infty} \sup_{\varepsilon \in [0, \delta]} \mathbb{P}(X_b / x > \eta \mid X_a = \varepsilon x) = 0. \tag{5}$$

Condition 1(ii) is similar to [26, Equation (3.4)] and prevents non-extreme values from causing extreme ones. A similar assumption is [33, Equation (2.4)]; [33, Example 7.5] illustrates what can go wrong without it.

Theorem 1. *Let $(X, \mathcal{T})$ be a nonnegative Markov tree on $\mathcal{T} = (V, E)$. Assume Condition 1. Let $(M_e : e \in E_u)$ be a vector of independent random variables such that the law of M_e is μ_e for all $e \in E_u$. Then*

$$\mathcal{L}(X / X_u \mid X_u = t) \xrightarrow{d} \Theta_u = (\Theta_{u,v})_{v \in V}, \qquad t \to \infty, \tag{6}$$

where $\Theta_{u,u} = 1$ and

$$\forall v \in V \setminus \{u\}, \qquad \Theta_{u,v} = \prod_{e \in [u \rightsquigarrow v]} M_e. \tag{7}$$

The random vector $(\Theta_{u,v})_{v \in V}$ is called the *tail tree* of the Markov tree $(X_v)_{v \in V}$, adapting terminology for Markov chains in [26]. In Figure 2, the tail tree is illustrated for a tree with seven nodes. For subvectors $(\Theta_{u,w})_{w \in W}$ where all nodes in W lie on the same path starting at u, the structure of the tail tree is that of a geometric random walk; take for instance $u = 1$ and $W = \{1, 4, 5, 7\}$ in Figure 2. The tail tree couples several geometric random walks together through the common edges in the underlying paths: in the same figure, consider for instance the vectors indexed by $\{1, 4, 5, 7\}$ and by $\{1, 4, 6\}$, respectively, which share the initial edge $(1, 4)$.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

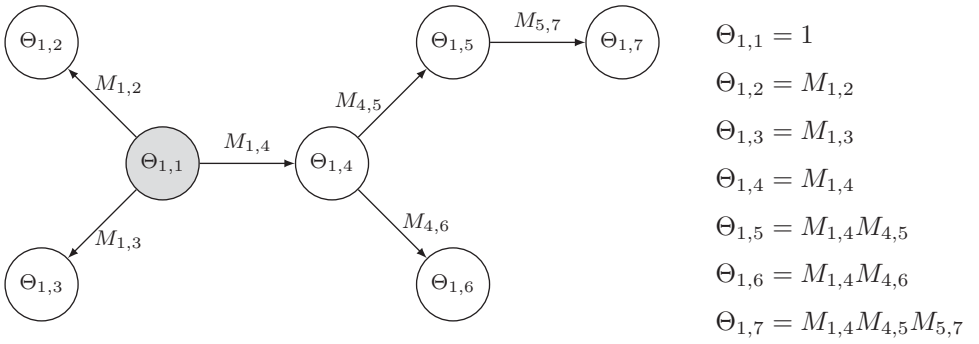


FIGURE 2: Illustration of the tail tree in (7). To each edge $e = (a, b)$ in the tree rooted at $u = 1$, a random variable $M_{a,b}$ is associated. These variables are independent. The limit variable $\Theta_{u,v}$ is the product of the variables M_e along the edges e on the path from u to v. The joint distribution of $\Theta_u = (\Theta_{u,v})_{v \in V}$ is the tail tree with root u.

Proof of Theorem 1. Put $d = |V| - 1 \geqslant 1$. The proof is by induction on d.

If V has only two elements, i.e. $d = 1$, then Condition 2(i) already confirms the convergence stated in (6) and (7). Therefore, we can henceforth assume that V has at least three elements, i.e. $d \geqslant 2$. Identify V with $\{0, 1, \ldots, d\}$ in such a way that the root is $u = 0$ and such that if $(a, b) \in E_u$ then $a < b$. Since $X_0/X_0 = 1 = \Theta_{0,0}$, we do not need to consider the components X_0 and $\Theta_{0,0}$ in (6).

Step 1. — Let k denote the parent of d in the directed tree $\mathcal{T}_0$; that is, k is the unique node in $\{0, 1, \ldots, d-1\}$ such that (k, d) is an edge in E_0. Our way of numbering nodes implies that d cannot be the parent of every other node. Condition 1 is then satisfied also for the nonnegative Markov tree $X_{0:(d-1)} = (X_0, X_1, \ldots, X_{d-1})$ on the tree that is obtained from $\mathcal{T}$ by removing node d from V and edges (k, d) and (d, k) from E. The induction hypothesis then means that, for every bounded, continuous function $g : [0, \infty)^d \to \mathbb{R}$, we have

$$\lim_{t \to \infty} \mathbb{E}[g(X_{1:(d-1)}/t) \mid X_0 = t] = \mathbb{E}[g(\Theta_{0,1:(d-1)})],$$

the joint distribution of $\Theta_{0,1:(d-1)} = (\Theta_{0,1}, \ldots, \Theta_{0,d-1})$ being given by (7).

Let $f : \mathbb{R}^d \to [0, 1]$ be a Lipschitz function. We will show that

$$\lim_{t \to \infty} \mathbb{E}[f(X_{1:d}/t) \mid X_0 = t] = \mathbb{E}[f(\Theta_{0,1:d})].$$

Recall that $k \in \{0, 1, \ldots, d-1\}$ denotes the parent node of d. We need to distinguish between two cases: $k = 0$ is the root or $k \in \{1, \ldots, d-1\}$ is a non-root vertex. The case $k = 0$ is similar to but easier than the case $k \in \{1, \ldots, d-1\}$ and is left to the reader. We assume henceforth that $k \in \{1, \ldots, d-1\}$.

Step 2. — Let $\delta > 0$ be such that $\mathbb{P}(\Theta_{0,k} = \delta) = 0$. We have

$$\left|\mathbb{E}[f(X_{1:d}/t) \mid X_0 = t] - \mathbb{E}[f(\Theta_{0,1:d})]\right| \tag{8}$$

$$\leqslant \left|\mathbb{E}[\mathbb{1}(X_k/t \geqslant \delta) f(X_{1:d}/t) \mid X_0 = t] - \mathbb{E}[\mathbb{1}(\Theta_k \geqslant \delta) f(\Theta_{0,1:d})]\right| \tag{9}$$

$$+ \left|\mathbb{E}[\mathbb{1}(X_k/t < \delta) f(X_{1:d}/t) \mid X_0 = t] - \mathbb{E}[\mathbb{1}(\Theta_k < \delta) f(\Theta_{0,1:d})]\right|. \tag{10}$$

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

We will show that the expression (9) converges to zero as $x_0 \to \infty$ (Step 3). Moreover, we will find a bound for the limit superior of (10) as $x \to \infty$. The bound will depend on δ but will converge to zero as $\delta \downarrow 0$ (Step 4). Together, these properties of (9) and (10) are sufficient to prove the theorem (Step 5).

Step 3*: The term* (9). — The vertex k is the parent of d in $\mathcal{T}_0$, and therefore it separates d from the other vertices. By the conditional independence property (3),

$$\mathbb{E}\biggl[\mathbb{1}\biggl(\frac{X_k}{t} \geqslant \delta\biggr) f\biggl(\frac{X_{1:d}}{t}\biggr) \mid X_0 = t\biggr] = \int_{x_{1:(d-1)}} \mathbb{1}\biggl(\frac{x_k}{t} \geqslant \delta\biggr)\, \mathbb{E}\biggl[f\biggl(\frac{x_{1:(d-1)}}{t}, \frac{X_d}{t}\biggr) \mid X_k = x_k\biggr] \mathbb{P}[X_{1:(d-1)} \in \mathrm{d}x_{1:(d-1)} \mid X_0 = t].$$

To explain our notation: the integral is over $x_{1:(d-1)} = (x_1, \ldots, x_{d-1})$ and is with respect to the conditional distribution of $X_{1:(d-1)} = (X_1, \ldots, X_{d-1})$ given that $X_0 = t$. The integrand involves the conditional expectation of a function of X_d given that $X_k = x_k$.

We change variables and integrate with respect to the conditional distribution of $X_{1:(d-1)}/t$ given that $X_0 = t$: we get

$$\mathbb{E}\biggl[\mathbb{1}\biggl(\frac{X_k}{t} \geqslant \delta\biggr) f\biggl(\frac{X_{1:d}}{t}\biggr) \mid X_0 = t\biggr] = \int_{y_{1:(d-1)}} g_t(y_{1:(d-1)}) \mathbb{P}\biggl[\frac{X_{1:(d-1)}}{t} \in \mathrm{d}y_{1:(d-1)} \mid X_0 = t\biggr], \tag{11}$$

where the integrand in (11) is given by

$$g_t(y_{1:(d-1)}) = \mathbb{1}(y_k \geqslant \delta) \mathbb{E}\biggl[f\biggl(y_{1:(d-1)}, y_k \frac{X_d}{t y_k}\biggr) \mid X_k = t y_k\biggr].$$

By Condition 1(i), we have $X_d/x_k \mid X_k = x_k \xrightarrow{d} M_{k,d}$ as $x_k \to \infty$. Define

$$g(y_{1:(d-1)}) = \mathbb{1}(y_k \geqslant \delta) \mathbb{E}[f(y_{1:(d-1)}, y_k M_{k,d})].$$

Recall that f is bounded and (Lipschitz) continuous. By the extended continuous mapping theorem [37, Theorem 18.11], we have, for all vectors $y_{1:(d-1)}$ such that $y_k \neq \delta$ and for all functions $y_{1:(d-1)}(\,\cdot\,)$ such that $y_{1:(d-1)}(t) \to y_{1:(d-1)}$ as $t \to \infty$, the limit relation

$$\lim_{t \to \infty} g_t\bigl(y_{1:(d-1)}(t)\bigr) = g(y_{1:(d-1)}).$$

Moreover, $\mathcal{L}(X_{1:(d-1)}/t \mid X_0 = t) \xrightarrow{d} \mathcal{L}(\Theta_{1:(d-1)})$ as $t \to \infty$ by the induction hypothesis. By the same extended continuous mapping theorem, the integral (11) converges to

$$\int_{y_{1:(d-1)}} g(y_{1:(d-1)}) \mathbb{P}[\Theta_{1:(d-1)} \in \mathrm{d}y_{1:(d-1)}] = \int_{y_{1:(d-1)}} \mathbb{1}(y_k \geqslant \delta) \mathbb{E}[f(y_{1:(d-1)}, y_k M_{k,d})] \mathbb{P}[\Theta_{1:(d-1)} \in \mathrm{d}y_{1:(d-1)}].$$

Recall that $(M_e)_{e \in E_u}$ is a vector of independent random variables such that the law of M_e is μ_e for $e \in E_u$. By construction, $M_{k,d}$ and $\Theta_{1:(d-1)}$ are then independent too: each component $\Theta_{0,j}$ of $\Theta_{1:(d-1)}$ is a product of random variables $M_{a,b}$ with $a, b \in \{0, \ldots, d-1\}$ and thus $e = (a, b) \neq (k, d)$. The above integral may therefore be simplified to

$$\mathbb{E}[\mathbb{1}(\Theta_k \geqslant \delta) f(\Theta_{1:(d-1)}, \Theta_k M_{k,d})] = \mathbb{E}[\mathbb{1}(\Theta_k \geqslant \delta) f(\Theta_{1:d})]$$

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

since $\Theta_d = \Theta_k M_{k,d}$. It follows that the limit of (9) as $t \to \infty$ is equal to zero.

Step 4*: The term* (10). — We consider two cases: $\mathbb{P}(\Theta_k = 0) = 0$ (Step 4.a) and $\mathbb{P}(\Theta_k = 0) > 0$ (Step 4.b).

Step 4.*a:* The case $\mathbb{P}(\Theta_k = 0) = 0$. — Since $0 \leqslant f \leqslant 1$, the integral (10) is bounded by

$$\mathbb{P}(X_k/t < \delta \mid X_0 = t) + \mathbb{P}(\Theta_k < \delta).$$

By the induction hypothesis, this sum converges to $2\mathbb{P}(\Theta_k < \delta)$ as $x \to \infty$. The latter probability converges to zero as $\delta \downarrow 0$, as required.

Step 4.*b: The case* $\mathbb{P}(\Theta_k = 0) > 0$. — We decompose (10) into three terms:

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathbb{1}\left(\frac{X_k}{t} < \delta\right) f\left(\frac{X_{1:d}}{t}\right) \mid X_0 = t\right] - \mathbb{E}[\mathbb{1}(\Theta_k < \delta) f(\Theta_{1:d})]\right| \\
&\quad \leqslant \left|\mathbb{E}\left[\mathbb{1}\left(\frac{X_k}{t} < \delta\right) f\left(\frac{X_{1:d}}{t}\right) \mid X_0 = t\right] - \mathbb{E}\left[\mathbb{1}\left(\frac{X_k}{t} < \delta\right) f\left(\frac{X_{1:(d-1)}}{t}, 0\right) \mid X_0 = t\right]\right| \quad (12) \\
&\qquad + \left|\mathbb{E}\left[\mathbb{1}\left(\frac{X_k}{t} < \delta\right) f\left(\frac{X_{1:(d-1)}}{t}, 0\right) \mid X_0 = t\right] - \mathbb{E}[\mathbb{1}(\Theta_k < \delta) f(\Theta_{1:(d-1)}, 0)]\right| \quad (13) \\
&\qquad + \left|\mathbb{E}[\mathbb{1}(\Theta_k < \delta) f(\Theta_{1:(d-1)}, 0)] - \mathbb{E}[\mathbb{1}(\Theta_k < \delta) f(\Theta_{1:d})]\right|. \quad (14)
\end{aligned}$$

Let $L > 0$ be such that $|f(y) - f(z)| \leqslant L \sum_{j=1}^{d} |y_j - z_j|$ for all $y, z \in \mathbb{R}^d$. Furthermore, recall that $0 \leqslant f \leqslant 1$, so that also $|f(y) - f(z)| \leqslant 1$ for all $y, z \in \mathbb{R}^d$.

Step 4.*b.i: The term* (12). — The term (12) is bounded by

$$\mathbb{E}\left[\mathbb{1}\left(\frac{X_k}{t} < \delta\right) |f\left(\frac{X_{1:d}}{t}\right) - f\left(\frac{X_{1:(d-1)}}{t}, 0\right)| \,\middle|\, X_0 = t\right] \leqslant \mathbb{E}\left[\mathbb{1}\left(\frac{X_k}{t} < \delta\right) \min\left(1, L\frac{X_d}{t}\right) \mid X_0 = t\right]. \quad (15)$$

The node k separates the nodes 0 and d. By the global Markov property, the expectation on the right-hand side of (15) is therefore equal to

$$\int_{[0,\delta)} \mathbb{E}\left[\min\left(1, L\frac{X_d}{t}\right) \mid X_k = \varepsilon t\right] \mathbb{P}\left[\frac{X_k}{t} \in \mathrm{d}\varepsilon \mid X_0 = t\right]. \quad (16)$$

Let $\eta \in (0, 2/L)$. The conditional expectation in the integrand in (16) satisfies

$$\begin{aligned}
\mathbb{E}[\min(1, LX_d/t) \mid X_k = \varepsilon t] &= \mathbb{E}[\min(1, LX_d/t)\, \mathbb{1}(X_d/t \leqslant \eta) \mid X_k = \varepsilon t] \\
&\quad + \mathbb{E}[\min(1, LX_d/t)\, \mathbb{1}(X_d/t > \eta) \mid X_k = \varepsilon t] \\
&\leqslant L\eta + \mathbb{P}(X_d/t > \eta \mid X_k = \varepsilon t).
\end{aligned}$$

Therefore, the integral in (16) is bounded by

$$L\eta + \sup_{\varepsilon \in [0,\delta)} \mathbb{P}(X_d > \eta t \mid X_k = \varepsilon t).$$

By Condition 1(ii), we can first take the limit superior as $t \to \infty$ and then the limit superior as $\delta \downarrow 0$ to find that

$$\limsup_{\delta \downarrow 0} \limsup_{t \to \infty} \mathbb{E}[\mathbb{1}(X_k/t < \delta) \min(1, LX_d/t) \mid X_0 = t] \leqslant L\eta.$$

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

Since η can be chosen arbitrarily close to zero, we find that the double limit superior above is equal to zero.

Step 4.*b.ii: The term* (13). — By the induction hypothesis, the term (13) converges to zero as $t \to \infty$.

Step 4.*b.iii: The term* (14). — Since $\Theta_d = \Theta_k M_{k,d}$, the term (14) is bounded by

$$\mathbb{E}[\mathbb{1}(\Theta_k < \delta)\, |f(\Theta_{1:(d-1)}, 0) - f(\Theta_d)|] \leqslant \mathbb{E}[\mathbb{1}(\Theta_k < \delta) \min(1, L\Theta_k M_{k,d})].$$

By the dominated convergence theorem, the expectation on the right-hand side converges to zero as $\delta \downarrow 0$.

Step 5. — The terms (9) and (10) were analyzed in Steps 3 and 4, respectively. In Step 3, it was shown that the term (9) converges to zero as $t \to \infty$, for any $\delta > 0$ such that $\mathbb{P}(\Theta_{0,k} = \delta) = 0$. In Step 4, it was shown that the limit superior as $t \to \infty$ of the term in (10) is bounded by a quantity depending on δ which converges to zero as $\delta \downarrow 0$. Since the expression in (8) does not depend on δ, its limit as $t \to \infty$ must thus be zero.

This completes the proof of the induction step and thus of the theorem.

Corollary 1. *In the setting of Theorem 2, we also have* $\mathcal{L}(X/X_u \mid X_u > t) \xrightarrow{d} \Theta_u$ *as* $t \to \infty$.

Proof. For a bounded and continuous function $f : [0, \infty)^V \to \mathbb{R}$, we have

$$\begin{aligned}\big|\mathbb{E}[f(X/X_u) \mid X_u > t] - \mathbb{E}[f(\Theta_u)]\big| \\ \leqslant \frac{1}{\mathbb{P}(X_u > t)} \int_{(t,\infty)} \big|\mathbb{E}[f(X/X_u) \mid X_u = s] - \mathbb{E}[f(\Theta_u)]\big|\, \mathbb{P}(X_u \in \mathrm{d}s).\end{aligned}$$

Given $\varepsilon > 0$, Theorem 1 allows us to find $t(\varepsilon)$ sufficiently high so that the absolute value inside the integral is bounded by ε for all $s \geqslant t(\varepsilon)$. But then the left-hand side in the previous display is bounded by ε too, for all $t \geqslant t(\varepsilon)$. Since $\varepsilon > 0$ was arbitrary, the stated convergence in distribution follows.

3. One- versus multi-component regular variation

Let $X = (X_1, \ldots, X_d)$ be a random vector of nonnegative variables. Upon an obvious change in notation, Corollary 1 concerned weak convergence of $\mathcal{L}(X/X_i \mid X_i > t)$ as $t \to \infty$ for some $i \in \{1, \ldots, d\}$. This convergence plus regular variation of the marginal distribution of X_i is a special case of what is called one-component regular variation in [17]. The weak limit, $\Theta_i = (\Theta_{i,j})_{j=1}^d$, depends on the choice of i. There may be good reasons to consider these limits for several indices i. Let $I \subset \{1, \ldots, d\}$ be the set of all indices i for which such a limit Θ_i exists. How are these random vectors Θ_i related?

In this section, several such one-component statements are combined into a single one which could be called multi-component regular variation. If $I = \{1, \ldots, d\}$, this is just ordinary multivariate regular variation. As discussed already in Section 1.2, the connections between the limits Θ_i generalize the time change formula for stationary regularly varying time series and can be deduced from their connections to a limiting tail measure.

Let $\mathbb{S} = [0, \infty)^d$ for some positive integer d and let $I \subset \{1, \ldots, d\}$ be non-empty. For $x \in \mathbb{S}$, put $x_I = (x_i)_{i \in I}$. Define $\mathbb{S}_{0,I} = \{x \in \mathbb{S} : \max(x_I) > 0\}$. Let $\mathcal{M}_{0,I}$ denote the collection of Borel measures ν on $\mathbb{S}_{0,I}$ with the property that $\nu(B)$ is finite for every Borel set B of $\mathbb{S}_{0,I}$ that is contained in a set of the form $\{x \in \mathbb{S} : \max(x_I) \geqslant \varepsilon\}$ for some $\varepsilon > 0$. Let $\mathcal{C}_{0,I}$ denote the collection of bounded, continuous functions $f : \mathbb{S}_{0,I} \to \mathbb{R}$ for which there exists $\varepsilon > 0$ such that $f(x) = 0$ as soon as $\max(x_I) \leqslant \varepsilon$. Let $\mathcal{M}_{0,I}$ be equipped with the smallest topology that makes

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

the evaluation mappings $\nu \mapsto \nu(f) = \int f \, \mathrm{d}\nu$ continuous, where f ranges over $\mathcal{C}_{0,I}$. This is the notion of $\mathcal{M}_{\mathbb{O}}$ convergence in [25], with, in their notation, $\mathbb{C} = \{x \in [0,\infty)^d : \forall i \in I, x_i = 0\}$ and $\mathbb{O} = \mathbb{S} \setminus \mathbb{C} = \mathbb{S}_{0,I}$. The topology just defined is metrizable, turning $\mathcal{M}_{0,I}$ into a complete, separable metric space, with convenient characterizations of relative compactness, a portmanteau theorem, and a mapping theorem, all very much in the spirit of the notion of vague convergence of Borel measures on locally compact second countable Hausdorff spaces. Convergence of measures with respect to this topology is denoted by the arrow $\xrightarrow{0}$. If I is just a singleton, $\{i\}$ say, then notation is simplifed from $\mathbb{S}_{0,\{i\}}$ to $\mathbb{S}_{0,i}$ and so on. For $\alpha > 0$, let $\mathrm{Pa}(\alpha)$ denote the Pareto distribution on $[1,\infty)$ with shape parameter α, that is, the distribution of a random variable Z such that $\mathbb{P}(Z > z) = z^{-\alpha}$ for $z \geqslant 1$. Product measure is denoted by $\otimes$.

Theorem 2. *Let $X = (X_1, \ldots, X_d)$ be a random vector in $\mathbb{S} = [0,\infty)^d$ and let $I \subset \{1, \ldots, d\}$ be non-empty. Let $F_i(x) = \mathbb{P}(X_i \leqslant x)$ and $\overline{F}_i = 1 - F_i$ for $i \in I$ and $x \in [0,\infty)$. Assume there exists a function b, regularly varying at infinity with index $\alpha > 0$, such that $\lim_{t\to\infty} b(t)\overline{F}_i(t) = c_i \in (0,\infty)$ for $i \in I$. The following statements are equivalent:*

(a) *For every $i \in I$ we have $\mathcal{L}(X/X_i \mid X_i > t) \xrightarrow{d} \mathcal{L}(\Theta_i)$ as $t \to \infty$ for some random vector $\Theta_i = (\Theta_{i,j})_{j=1}^d$ on $\mathbb{S}$.*

(b) *For every $i \in I$ we have $\mathcal{L}(X_i/t, X/X_i \mid X_i > t) \xrightarrow{d} \mathrm{Pa}(\alpha) \otimes \mathcal{L}(\Theta_i)$ as $t \to \infty$ for some random vector Θ_i on $\mathbb{S}$.*

(c) *For every $i \in I$ we have $\mathcal{L}(X/t \mid X_i > t) \xrightarrow{d} \mathcal{L}(Y_i)$ as $t \to \infty$ for some random vector $Y_i = (Y_{i,j})_{j=1}^d$ on $\mathbb{S}$.*

(d) *For every $i \in I$ there exists $\nu_i \in \mathcal{M}_{0,i}$ such that $b(t)\mathbb{P}(X/t \in \cdot\,) \xrightarrow{0} \nu_i$ as $t \to \infty$ in $\mathcal{M}_{0,i}$.*

(e) *There exists $\nu \in \mathcal{M}_{0,I}$ such that $b(t)\mathbb{P}(X/t \in \cdot\,) \xrightarrow{0} \nu$ as $t \to \infty$ in $\mathcal{M}_{0,I}$.*

When these statements hold, the limiting objects are connected in the following ways: for all $i \in I$,

(i) *Y_i is equal in distribution to $Y_{i,i}\Theta_i$, where $\mathcal{L}(Y_{i,i}) = \mathrm{Pa}(\alpha)$ and $Y_{i,i}$ and Θ_i are independent;*

(ii) *ν_i is equal to the restriction of ν to $\mathbb{S}_{0,i}$;*

(iii) *$\mathbb{P}(Y_i \in \cdot\,) = c_i^{-1}\nu(\cdot \cap \{x : x_i > 1\})$;*

(iv) *for every Borel measurable $f : \mathbb{S}_{0,I} \to [0,\infty]$, we have*

$$\int_{\mathbb{S}_{0,I}} f(x)\, \mathbb{1}\{x_i > 0\} \, \mathrm{d}\nu(x) = c_i \mathbb{E}\left[\int_0^\infty f(z\Theta_i)\, \alpha z^{-\alpha-1} \, \mathrm{d}z\right]. \tag{17}$$

The proof of Theorem 2, together with the proofs of the other theorems in this section, is given in Appendix A.

To highlight the connection with the theory of one-component regular variation in [17], note that the random vector (X, Y) taking values in $[1,\infty) \times \mathbb{R}^{d-1}$ in [17, Theorem 1.4] plays the same role as the random vector $(X_i, X/X_i)$ in Theorem 2 above. The equivalence between (ii) and (iv) in the cited theorem is then the same as the equivalence between (a) and (b) in Theorem 2.

Further, note that in Theorem 2(a), for $j \in \{1, \ldots, d\}$ such that $\Theta_{i,j}$ is not degenerate at 0, we necessarily have $\liminf_{t\to\infty} \mathbb{P}(X_j > t)/\mathbb{P}(X_i > t) > 0$; i.e., the tail of X_j is at least as heavy

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

as that of X_i. If also $j \in I$, we can reverse the roles of i and j to find that the condition that the tails of all variables X_i with $i \in I$ are balanced is almost forced.

Apart from the characterizations (a)–(e) in Theorem 2, other equivalent ones are possible, for instance, involving sequences rather than functions, with a scaling function inside the probability rather than outside, or with respect to radial and 'angular' coordinates $(\rho(X)/t, X/\rho(X))$ for some appropriate functional ρ. See for instance [29, Theorem 6.1] and [25, Theorem 3.1]. The tail measures ν and ν_i are homogeneous with index $-\alpha$ [25, Theorem 3.1] and, upon a coordinate transformation, can be written as product measures. Since the focus here is on the weak limits Θ_i, these properties are not further elaborated upon. The statement (e) in Theorem 3 implies that the vector $X_I = (X_i)_{i \in I}$ is multivariate regularly varying with limit measure $\nu_I(\,\cdot\,) = \nu(\{x \in [0,\infty)^d : x_I \in \,\cdot\,\})$ on $[0,\infty)^I \setminus \{0\}$, which in turn implies, among other things, that it is in the domain of attraction of a multivariate max-stable distribution with Fréchet margins and exponent measure ν_I; see for instance [28] and [29].

A noteworthy special case of (17) is when f is the indicator function of the orthant $\{x \in \mathbb{S}_{0,I} : \forall j \in J,\ x_j > y_j\}$, where $J \subset \{1,\ldots,d\}$ has a non-empty intersection with I and where $y_j > 0$ for all $j \in J$. If $i \in I \cap J$, then $f(x)\mathbb{1}\{x_i > 0\} = f(x)$, and thus

$$\nu(\{x \in \mathbb{S}_{0,I} : \forall j \in J,\ x_j > y_j\}) = c_i\, \mathbb{E}[\,\min\{y_j^{-\alpha}\Theta_{i,j}^{\alpha} : j \in J\}]. \tag{18}$$

A remarkable consequence is that the right-hand side does not depend on the choice of $i \in I \cap J$. This invariance property is a special case of a more general mutual consistency property of the limit distributions $\mathcal{L}(\Theta_i)$ for $i \in I$ that is formulated in Corollary 2 below.

Corollary 2. (Model consistency.) *If the equivalent conditions of Theorem 2 are fulfilled, then, for Borel measurable $f : \mathbb{S}_{0,I} \to [0,\infty]$ and for $i, j \in I$, we have*

$$c_j \mathbb{E}\left[\mathbb{1}\{\Theta_{j,i} > 0\} \int_0^\infty f(z\Theta_j)\, \alpha z^{-\alpha-1}\, \mathrm{d}z\right] = c_i \mathbb{E}\left[\mathbb{1}\{\Theta_{i,j} > 0\} \int_0^\infty f(z\Theta_i)\, \alpha z^{-\alpha-1}\, \mathrm{d}z\right]. \tag{19}$$

Proof. By (17), both sides in (19) are equal to $\int_{\mathbb{S}_{0,I}} f(x)\, \mathbb{1}\{x_i > 0, x_j > 0\}\, \mathrm{d}\nu(x)$.

Corollary 3. (Root-change formula.) *If the equivalent conditions of Theorem 2 are fulfilled, then, for all Borel measurable $g : \mathbb{S}_{0,I} \to [0,\infty)$ and for all $i, j \in I$, we have*

$$c_j\, \mathbb{E}[g(\Theta_j)\, \mathbb{1}\{\Theta_{j,i} > 0\}] = c_i\, \mathbb{E}[g(\Theta_i/\Theta_{i,j})\, \Theta_{i,j}^{\alpha}]. \tag{20}$$

In particular, $\mathbb{P}[\Theta_{j,i} > 0] = 1$ if and only if $\mathbb{E}[\Theta_{i,j}^{\alpha}] = c_j/c_i$, and then, for all g as above,

$$\mathbb{E}[g(\Theta_j)] = \mathbb{E}[g(\Theta_i/\Theta_{i,j})\, \Theta_{i,j}^{\alpha}]/\mathbb{E}[\Theta_{i,j}^{\alpha}]. \tag{21}$$

Proof. In Corollary 2, take $f(x) = g(x/x_j)\, \mathbb{1}\{x_j > 1\}$. As $\Theta_{j,j} = 1$, the left-hand side of (19) is $c_j \mathbb{E}[\mathbb{1}\{\Theta_{j,i} > 0\} g(\Theta_j)]$. The right-hand side of (19) becomes $c_i \mathbb{E}[\mathbb{1}\{\Theta_{i,j} > 0\} g(\Theta_i/\Theta_{i,j}) \Theta_{i,j}^{\alpha}]$, in which the indicator function is redundant.

The special case $g \equiv 1$ in (20) yields $c_j \mathbb{P}(\Theta_{j,i} > 0) = c_i\, \mathbb{E}[\Theta_{i,j}^{\alpha}]$. If $\mathbb{E}[\Theta_{i,j}^{\alpha}] = c_j/c_i$, we have $\mathbb{P}(\Theta_{j,i} > 0) = 1$, and the indicator function on the left-hand side of (20) can be omitted.

If the limit measure ν does not assign any mass to the coordinate hyperplane $\{x : x_i = 0\}$, the indicator in (17) is redundant and ν can be expressed entirely in terms of $\mathcal{L}(\Theta_i)$. Moreover, whether this occurs or not can be read off from the αth moments of the components of Θ_i.

Corollary 4. *In Theorem 2, we have, for $i \in I$,*

$$\nu(\{x \in \mathbb{S}_{0,I} : x_i = 0\}) = \begin{cases} 0 & \text{if } \mathbb{E}[\Theta_{i,j}^{\alpha}] = c_j/c_i \text{ for all } j \in I, \\ \infty & \text{otherwise.} \end{cases} \tag{22}$$

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

If $\nu(\{x : x_i = 0\}) = 0$ *for some* $i \in I$*, then, for all Borel measurable* $f : \mathbb{S}_{0,I} \to [0, \infty]$*,*

$$\nu(f) = c_i \mathbb{E}\left[\int_0^\infty f(z\Theta_i)\, \alpha z^{-\alpha-1}\, \mathrm{d}z\right]. \tag{23}$$

Moreover, all tail trees Θ_j *for* $j \in I$ *are determined by* Θ_i *via (21).*

Proof. Choose $i, j \in I$. In (17), let f be the indicator function of the set $\{x : x_i = 0\}$ and let the index i in (17) be equal to the index j chosen here. It follows that $\nu(\{x \in \mathbb{S}_{0,I} : x_i = 0,\, x_j > 0\})$ is zero if $\mathbb{P}(\Theta_{j,i} = 0) = 0$ and infinity otherwise. By (20) with $g \equiv 1$, we have $\mathbb{P}(\Theta_{j,i} = 0) = 0$ if and only if $\mathbb{P}(\Theta_{j,i} > 0) = 1$ if and only if $\mathbb{E}[\Theta_{i,j}^\alpha] = c_j/c_i$. Equation (22) follows. If $\nu(\{x : x_i = 0\}) = 0$, we can omit the indicator $\mathbb{1}\{x_i > 0\}$ on the left-hand side of (17), yielding (23).

Let f be the indicator function of the set $\{x : \exists j \in J,\ x_j > y_j\}$, where $J \subset \{1, \ldots, d\}$ is non-empty and where $y_j > 0$ for all $j \in J$. If $\nu(\{x : x_i = 0\}) = 0$, then, by (23),

$$\nu(\{x \in \mathbb{S}_{0,I} : \exists j \in J,\ x_j > y_j\}) = c_i\, \mathbb{E}[\max\{y_j^{-\alpha} \Theta_{i,j}^\alpha : j \in J\}]. \tag{24}$$

In contrast to Equation (18), Equation (24) is true only when $\nu(\{x : x_i = 0\}) = 0$, a prerequisite for which Corollary 4 gives a necessary and sufficient condition.

By Corollary 4, the case where $\mathbb{E}[\Theta_{i,j}^\alpha] = c_j/c_i$ for all $j \in I$ leads to considerable simplifications. In fact, the special case $K = \{i\}$ in the next theorem implies that even the weak convergence of $\mathcal{L}(X/X_j \mid X_j > t)$ for $j \in I$ can then be deduced from the weak convergence of $\mathcal{L}(X/X_i \mid X_i > t)$ alone.

Theorem 3. *In Theorem 2, a sufficient condition for (a)–(e) to hold is that there exists a non-empty set* $K \subset I$ *with the following two properties:*

(i) *For every* $i \in K$ *we have* $\mathcal{L}(X/X_i \mid X_i > t) \xrightarrow{d} \mathcal{L}(\Theta_i)$ *as* $t \to \infty$ *for some random vector* Θ_i *on* $\mathbb{S}$*.*

(ii) *For every* $j \in I \setminus K$*, there exists* $i = i(j) \in K$ *such that* $\mathbb{E}[\Theta_{i,j}^\alpha] = c_j/c_i$*.*

In that case, we also have $\mathcal{L}(X/X_j \mid X_j > t) \xrightarrow{d} \mathcal{L}(\Theta_j)$ as $t \to \infty$ for $j \in I \setminus K$, where the law of Θ_j is given in terms of that of Θ_i with $i = i(j)$ via (21).

The focus so far has been on weak limits of conditional distributions involving a high-threshold exceedance by a specific component. In the spirit of the multivariate peaks-over-thresholds methodology [10], [22], the following result covers, among other possibilities, the case where the conditioning event involves a high-threshold exceedance in at least one of a number of components.

Theorem 4. *Suppose the conditions of Theorem 2 are fulfilled. Let* $\rho : \mathbb{S} \to [0, \infty)$ *be continuous and homogeneous of order one; that is,* $\rho(\lambda x) = \lambda \rho(x)$ *for* $\lambda \in [0, \infty)$ *and* $x \in \mathbb{S}$*. If* $\mathbb{S}_\rho := \{x \in \mathbb{S}_{0,I} : \rho(x) > 1\}$ *is contained in a set of the form* $\{x : \max(x_I) > \varepsilon\}$ *for some* $\varepsilon > 0$ *and if* $\nu(\mathbb{S}_\rho) > 0$*, then* $b(t)\mathbb{P}[\rho(X) > t] \to \nu(\mathbb{S}_\rho) \in (0, \infty)$ *as* $t \to \infty$ *and* $\mathcal{L}(X/t \mid \rho(X) > t) \xrightarrow{d} \nu(\cdot \cap \mathbb{S}_\rho)/\nu(\mathbb{S}_\rho)$ *as* $t \to \infty$*. The conclusions of Theorem 2 thus apply to the random vector* $(X, \rho(X))$ *in the space* $[0, \infty)^{d+1}$ *and relative to the index set* $I \cup \{d+1\}$*.*

Examples of ρ in Theorem 4 are $\rho(x) = \max_{i \in I} a_i x_i$ and $\rho(x) = \sum_{i \in I} a_i x_i$, where $a \in [0, \infty)^I \setminus \{0\}$. The special case $I = \{1, \ldots, d\}$ and $\rho(x) = \max(x_1, \ldots, x_d)$ produces multivariate Pareto distributions as in [29, Section 6.3] and other references mentioned in Section 1.2. Also covered by Theorem 4 is $\rho(x) = \min_{i \in J} a_i x_i$ for non-empty $J \subset I$ and $a_j \in (0, \infty)$ for all $j \in J$ provided $\mathbb{E}[\min\{a_j^\alpha \Theta_{i,j}^\alpha : j \in J\}] > 0$ for some (and hence all) $i \in I$. If, however, $\nu(\mathbb{S}_\rho) = 0$,

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

then $\mathbb{P}[\rho(X) > t]$ decays more rapidly than $b(t)$, and more refined models are needed, opening up a whole new world of possibilities.

Example 1. *Let $X = (X_1, \ldots, X_d)$ follow the max-linear model*

$$\forall i \in \{1, \ldots, d\}, \qquad X_i = \max_{r=1,\ldots,s} a_{i,r} Z_r, \tag{25}$$

where $a_{i,r} \in [0, \infty)$ are scalars such that $\max_r a_{i,r} > 0$ for all i, and where $Z_1, \ldots, Z_s$ are independent and identically distributed nonnegative random variables whose common distribution function F has a regularly varying tail function $\overline{F} = 1 - F$ with index $-\alpha < 0$. The marginal tails satisfy $\overline{F}_i(t)/\overline{F}(t) \to \sum_r a_{i,r}^\alpha =: c_i$ as $t \to \infty$. If X_i exceeds a large threshold $t \to \infty$, the probability that this was due to Z_r is proportional to $a_{i,r}^\alpha$, and then the other factors $Z_{\bar{r}}$ for $\bar{r} \neq r$ are of smaller order than Z_r. It follows that (a) in Theorem 2 holds where the law of Θ_i is discrete with at most s atoms and is given by

$$\mathcal{L}(\Theta_i) = \frac{1}{c_i} \sum_{r=1}^{s} a_{i,r}^\alpha \, \epsilon_{(a_{j,r}/a_{i,r})_{j=1}^d}, \tag{26}$$

with $\epsilon_x(\cdot)$ denoting a unit point mass at x. From (26), we find

$$\mathbb{E}[\Theta_{i,j}^\alpha] = \frac{1}{c_i} \sum_{r=1}^{s} a_{j,r}^\alpha \mathbb{1}\{a_{i,r} > 0\}.$$

It follows that $\mathbb{E}[\Theta_{i,j}^\alpha] = c_j/c_i$ as soon as $a_{i,r} > 0$ for every r such that $a_{j,r} > 0$. Indeed, in that case, if X_j is large, then some variable Z_r with $r \in \{1, \ldots, s\}$ such that $a_{j,r}$ is positive was large, which in turn implies that X_i is large as well, so that $\mathbb{P}(\Theta_{j,i} > 0) = 1$.

Example 2. *Recursive max-linear models on directed acyclic graphs were introduced in [13]. Borrowing some of the notation of that paper, consider a directed acyclic graph $\mathcal{D} = (V, E)$ with nodes $V = \{1, \ldots, d\}$ and edges $E = \{(k, i) : i \in V, k \in \mathrm{pa}(i)\}$, where $\mathrm{pa}(i) \subset V$ denotes the possibly empty set of parents of i. Consider a random vector $X = (X_1, \ldots, X_d)$ given by the structural equation model*

$$\forall i \in \{1, \ldots, d\}, \qquad X_i = \max\left[\max\{\gamma_{ki} X_k : k \in \mathrm{pa}(i)\}, \gamma_{ii} Z_i\right], \tag{27}$$

where the random variables $Z_1, \ldots, Z_d$ are as in Example 1 with $s = d$, and where all coefficients γ_{ki} and γ_{ii} are (strictly) positive; the maximum over the empty set is zero by convention. Then by [14, Theorem 2.2], the random vector X admits the max-linear representation

$$\forall i \in \{1, \ldots, d\}, \qquad X_i = \max_{j=1,\ldots,d} b_{ji} Z_j,$$

with coefficients b_{ji}, for $i, j \in \{1, \ldots, d\}$, defined as follows: $b_{ii} = \gamma_{ii}$ and $b_{ji} = 0$ if $j \in V \setminus (\mathrm{an}(i) \cup \{i\})$, while

$$\forall i \in \{1, \ldots, d\}, \forall j \in \mathrm{an}(i), \qquad b_{ji} = \max_{p \in P_{ji}} \left\{ \gamma_{jj} \prod_{e \in p} \gamma_e \right\}, \tag{28}$$

where P_{ji} is the collection of paths $p = \{e_1, \ldots, e_n\}$ from j to i in $\mathcal{D}$, and where $\mathrm{an}(i)$ denotes the set of ancestors of i; recall the definition of a path in the beginning of Section 2. The representation in (28) is of the form (25) with $s = d$ and with $a_{i,r} = b_{ri}$ for $i, r \in \{1, \ldots, d\}$. It follows that $a_{i,r} = 0$ unless $r = i$ or $r \in \mathrm{an}(i)$.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

The condition that $a_{i,r} > 0$ whenever $a_{j,r} > 0$ is satisfied as soon as $j \in \mathrm{an}(i)$: indeed, in that case, we have $\mathrm{an}(j) \cup \{j\} \subset \mathrm{an}(i)$, so that $b_{rj} > 0$ implies $r \in \mathrm{an}(j) \cup \{j\}$ and thus $r \in \mathrm{an}(i)$, implying $b_{ri} > 0$. Through the structural equation model (27), a large value appearing at a node j will also be felt at any of its descendants i. We get $\mathbb{P}(\Theta_{j,i} > 0) = 1$ for $j \in \mathrm{an}(i)$, which, by Corollary 3, means that the root-change formula (21) applies, by which the law of Θ_j can be recovered from that of Θ_i. Moreover, Theorem 3 applies with K equal to the set of leaf nodes, i.e., the nodes without descendants.

In the special case that the directed acyclic graph is also a directed, rooted tree, every node i either has exactly one parent or is equal to the root node, say u. In that case, the collection of paths P_{ji} between $j \in \mathrm{an}(i)$ and $i \in \{1, \ldots, d\} \setminus \{u\}$ is a singleton, $p = [j \rightsquigarrow i]$, and the formula for b_{ji} in (28) simplifies to $b_{ji} = \gamma_{jj} \prod_{e \in [j \rightsquigarrow i]} \gamma_e$. Furthermore, the tail tree Θ_u in (26) starting from the root node u simplifies to the degenerate distribution at the point $\theta_u = (\theta_{u,1}, \ldots, \theta_{u,d})$ with coordinates $\theta_{u,j} = \prod_{e \in [u \rightsquigarrow j]} \gamma_e \in (0, \infty)$ for $j \in \{1, \ldots, d\}$. This is of the form in (7) with degenerate increments $M_e = \gamma_e$ for all $e \in E$.

4. Regularly varying Markov trees

As in Section 2, let $(X, \mathcal{T})$ be a nonnegative Markov tree on the undirected tree $\mathcal{T} = (V, E)$. The general theory in Section 3 sheds light on the relation between two tail trees emanating from different roots. For two different nodes u and $\bar{u}$ in V, the sets of directed edges E_u and $E_{\bar{u}}$ are the same except for the edges connecting nodes on the path between u and $\bar{u}$, which are directed in opposite ways in the two edge sets: for every $(a, b) \in [u \rightsquigarrow \bar{u}] = E_u \setminus E_{\bar{u}}$, we have $(b, a) \in [\bar{u} \rightsquigarrow u] = E_{\bar{u}} \setminus E_u$, and vice versa.

Condition 1 was formulated relative to a single root $u \in V$. The next condition covers all nodes $u \in U$ in a non-empty subset U of V as possible roots. For such U, let $E_U = \bigcup_{u \in U} E_u$ denote the set of directed edges that appear in at least one of the directed trees E_u.

Condition 2. There exists a non-empty $U \subset V$ with the following two properties:

(i) For every $e = (a, b) \in E_U$, there exists a version of the conditional distribution of X_b given X_a and a probability measure μ_e on $[0, \infty)$ such that (4) holds.

(ii) For every edge $e = (a, b) \in E_U$ for which there exists $u \in U$ such that $e \in E_u$ and an edge $\bar{e} \in [u \rightsquigarrow a]$ such that $\mu_{\bar{e}}(\{0\}) > 0$, we have (5).

If $u, v \in U$ in Condition 2, then every node $w \in V$ that is on the path between u and v can be added to U and Condition 2 remains true. Indeed, for such u, v, w, we have $E_w \subset E_u \cup E_v$, which takes care of (i), and $[w \rightsquigarrow a] \subset [u \rightsquigarrow a] \cup [v \rightsquigarrow a]$ for every node $a \in V$, which takes care of (ii). The author is grateful to an anonymous reviewer for having pointed this out.

Corollary 5. *Let $(X, \mathcal{T})$ be a nonnegative Markov tree on the undirected tree $\mathcal{T} = (V, E)$. Let $U \subset V$ be non-empty. Let $F_u(x) = \mathbb{P}(X_u \leqslant x)$ and $\overline{F}_u = 1 - F_u$ for all $u \in U$. Assume that there exists a positive function b, regularly varying at infinity with index $\alpha > 0$, such that $b(t)\overline{F}_u(t) \to c_u \in (0, \infty)$ as $t \to \infty$ for every $u \in U$. Assume that Condition 2 holds. Let $(M_e)_{e \in E_U}$ be a vector of independent random variables such that M_e has law μ_e for each $e \in E_U$. Then all conclusions of Theorem 2 hold with $I = U$ and with Θ_u the tail tree in (7) for $u \in U$.*

Proof. Condition 2 and Corollary 1 imply that the statement (a) in Theorem 2 is satisfied for $I = U$ and with Θ_u the tail tree in (7), for every $u \in U$. All equivalence relations and other properties are then as stated in Theorem 2.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

Corollary 6. *In Corollary 5, if $a, b \in V$ are neighbours in E and if they both belong to U, then the distributions of $M_{a,b}$ and $M_{b,a}$ mutually determine each other by*

$$c_b \mathbb{E}[g(M_{b,a})\, \mathbb{1}\{M_{b,a} > 0\}] = c_a \mathbb{E}[g(1/M_{a,b})\, M_{a,b}^{\alpha}] \tag{29}$$

for all Borel measurable $g : (0, \infty) \to [0, \infty]$.

Proof. To derive (29), apply (20) to the case $d = 2$ and the random vector (X_a, X_b). The two limit random vectors Θ_u in Theorem 2(a) are $(1, M_{a,b})$ and $(M_{b,a}, 1)$ when conditioning on $X_a > t$ and on $X_b > t$, respectively. Equation (29) implies $\mathbb{P}(M_{b,a} > z) = (c_a/c_b)\mathbb{E}[\mathbb{1}\{zM_{a,b} < 1\}\, M_{a,b}^{\alpha}]$ for all $z \in [0, \infty)$, so that the distribution of $M_{b,a}$ can be recovered from that of $M_{a,b}$.

For different roots $u, \bar{u} \in U$, the tail trees Θ_u and $\Theta_{\bar{u}}$ have the same multiplicative structure. The differences between their distributions lie in the starting nodes of the paths and in the distributions of the multiplicative increments for edges on the paths $[u \rightsquigarrow \bar{u}]$ and $[\bar{u} \rightsquigarrow u]$, since these edges change direction. For such edges of which the nodes belong to U as well, the increment distributions are related by (29). See Figure 2 for an illustration.

For $u, \bar{u} \in U$, the equality $\mathbb{E}[\Theta_{u,\bar{u}}^{\alpha}] = c_{\bar{u}}/c_u$ has interesting ramifications; see Corollaries 3 and 4 and Theorem 3. If all nodes on the path between u and $\bar{u}$ belong to U as well, then, since

$$\mathbb{E}[\Theta_{u,\bar{u}}^{\alpha}] = \prod_{e \in [u \rightsquigarrow \bar{u}]} \mathbb{E}[M_e^{\alpha}], \tag{30}$$

we have $\mathbb{E}[\Theta_{u,\bar{u}}^{\alpha}] = c_{\bar{u}}/c_u$ as soon as $\mathbb{E}[M_{a,b}^{\alpha}] = c_b/c_a$ for every $e = (a, b) \in [u \rightsquigarrow \bar{u}]$.

Given the tree structure, the distribution of a Markov tree X on $\mathcal{T} = (V, E)$ is entirely determined by the bivariate distributions (X_a, X_b) for $e = (a, b) \in E$. Markov chains of which all pairs (X_i, X_{i+1}) are max-stable were proposed in [5, Section 4.6] and [36]. When extended to trees, this construction method provides models meeting Condition 2.

Example 3. *Let the distribution of the random pair (X, Y) on $(0, \infty)^2$ be bivariate max-stable with cumulative distribution function*

$$F(x, y) = \exp\{-(x^{-1} + y^{-1})\, A(x/(x + y))\}, \qquad (x, y) \in (0, \infty)^2,$$

where $A : [0, 1] \to [1/2, 1]$ is a Pickands dependence function, that is, a convex function such that $\max(w, 1 - w) \leqslant A(w) \leqslant 1$ for all $w \in [0, 1]$; see [15] and the references therein. Both marginal distributions are unit-Fréchet: $F(z, \infty) = F(\infty, z) = \exp(-1/z)$ for $z \in (0, \infty)$. In particular, the marginal tail functions are regularly varying at infinity with index $-\alpha = -1$.

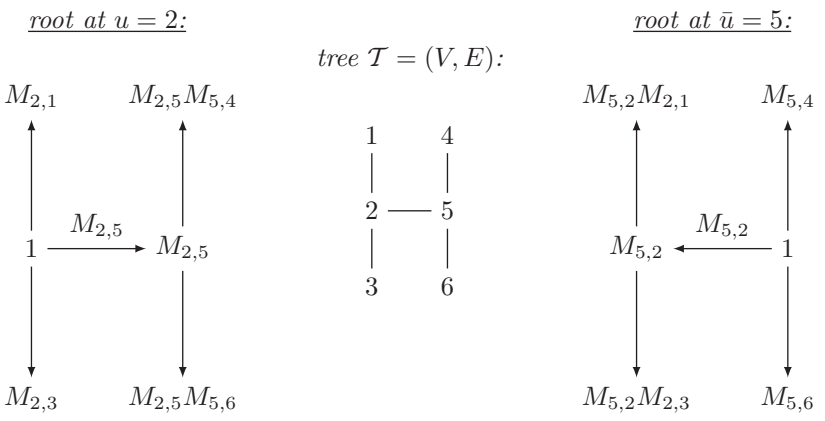
Let A' be the left-hand derivative of A, which exists everywhere on $(0, 1]$, takes values between -1 and 1, and is non-decreasing and continuous from the left; define $A'(0)$ as the right-hand limit. Since A is convex, it is absolutely continuous, and the set of points in $(0, 1)$ where it is not continuously differentiable is at most countable. For $x, y \in (0, \infty)$ such that A is differentiable at $w = x/(x + y)$, we have

$$\mathbb{P}(Y \leqslant y \mid X = x) = \frac{\partial F(x, y)/\partial x}{\partial F(x, \infty)/\partial x} = \exp\{-x^{-1}((1 - w)^{-1}A(w) - 1)\}\, \{A(w) - w\, A'(w)\}.$$

It follows that $\mathcal{L}(Y/x \mid X = x) \xrightarrow{d} M$ as $x \to \infty$, where

$$\mathbb{P}(M \leqslant z) = A(w) - w\, A'(w), \qquad z \in [0, \infty), \quad w = 1/(1 + z). \tag{31}$$

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press



$$\Theta_{2,1} = M_{2,1} \qquad \Theta_{5,1} = M_{5,2}M_{2,1}$$
$$\Theta_{2,2} = 1 \qquad \Theta_{5,2} = M_{5,2}$$
$$\Theta_{2,3} = M_{2,3} \qquad \Theta_{5,3} = M_{5,2}M_{2,3}$$
$$\Theta_{2,4} = M_{2,5}M_{5,4} \qquad \Theta_{5,4} = M_{5,4}$$
$$\Theta_{2,5} = M_{2,5} \qquad \Theta_{5,5} = 1$$
$$\Theta_{2,6} = M_{2,5}M_{5,6} \qquad \Theta_{5,6} = M_{5,6}$$

FIGURE 3: The tail tree with root at node $u = 2$ (left) versus the tail tree with root at node $\bar{u} = 5$ (right) of the same Markov tree X on the tree $\mathcal{T}$ depicted in the middle. The path $[u \rightsquigarrow \bar{u}]$ is replaced by the path $[\bar{u} \rightsquigarrow u]$, and the direction of any edge on the path between the two roots (here, just the edge (2, 5)) is reversed, while the other edges remain unaffected. For any edge $e = (a, b)$ that changes direction, the distribution of the increment $M_{a,b}$ changes to that of the increment $M_{b,a}$. If both a and b belong to U, the increment distributions mutually determine each other through (29).

This is Part (i) of Condition 1. Further, Equation (5) in Part (ii) of Condition 1 follows from the monotone regression dependence property of bivariate max-stable distributions established in [12], by which the supremum over ε in (5) is attained in $\varepsilon = \delta$ and the limit superior as $x \to \infty$ is bounded by $\mathbb{P}(M \geqslant \eta/\delta)$, which tends to 0 as $\delta \downarrow 0$ for every fixed $\eta > 0$.

This construction using bivariate max-stable distributions is in some sense generic. Given a random variable M on $[0, \infty)$ with expectation $\mathbb{E}(M) \leqslant 1$, one can define a Pickands dependence function A by $A(w) = 1 - \mathbb{E}[\min(1 - w, wM)]$ for $w \in [0, 1]$, and then (31) holds. The extension to general exponents α and tail constants c_u is straightforward.

5. Absolutely continuous case

If the joint distribution of the Markov tree $X = (X_v)_{v \in V}$ on $\mathcal{T} = (V, E)$ is absolutely continuous with respect to the Lebesgue measure on $[0, \infty)^V$, the formulations of the conditions and results simplify considerably. Let f denote the joint probability density function of X and let f_v, for $v \in V$, denote the marginal density of X_v.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

By the Hammersley–Clifford theorem [23, Theorem 3.9], X is a Markov tree as soon as the joint density factorizes as

$$\forall x \in \mathbb{R}^d, \qquad f(x) = \prod_{j=1}^{v} f_v(x_v) \prod_{\substack{\{a,b\} \subset V: \\ a \text{ and } b \text{ are neighbours}}} \frac{f_{a,b}(x_a, x_b)}{f_a(x_a) f_b(x_b)}.$$

The second product is over all unordered pairs of neighbours and $f_{a,b}$ denotes the bivariate density function of (X_a, X_b).

For $t \in (0, \infty)$ such that $f_a(t) \in (0, \infty)$, the density of $\mathcal{L}(X_b/t \mid X_a = t)$ is $t f_{a,b}(t, ty)/f_a(t)$ for $y \in (0, \infty)$. The following condition replaces Condition 2.

Condition 3. For every $e = (a, b) \in E$, there exists a probability density function $q_{a,b}$ on $(0, \infty)$ such that

$$\forall y \in (0, \infty), \qquad \lim_{t \to \infty} \frac{t f_{a,b}(t, ty)}{f_a(t)} = q_{a,b}(y).$$

Theorem 5. *Let the random vector X on $[0, \infty)^V$ be a Markov tree on the undirected tree $\mathcal{T} = (V, E)$ with joint density function f. Assume there exists a positive function g, regularly varying at infinity with index $-\alpha - 1 < -1$, such that $f_v(t)/g(t) \to c_v \in (0, \infty)$ as $t \to \infty$ for every $v \in (0, \infty)$. If Condition 3 holds, then the conditions of Corollary 5 are satisfied with $U = V$, the same constants c_u, and auxiliary function $b(t) = \alpha/\{t\, g(t)\}$. For all pairs of neighbours $a, b \in V$, the density of $M_{a,b}$ is $q_{a,b}$ and for almost every $y \in (0, \infty)$, we have*

$$c_b\, y^{\alpha}\, q_{b,a}(y) = c_a\, y^{-2}\, q_{a,b}(y^{-1}). \tag{32}$$

Moreover, $\mathbb{E}[\Theta_{u,v}^{\alpha}] = c_v/c_u$ for all $u, v \in V$, so that $b(t)\mathbb{P}(X/t) \xrightarrow{0} \nu$ as $t \to \infty$, where $\nu \in \mathcal{M}_0$ satisfies

$$\nu(f) = c_u \mathbb{E}\left[\int_0^{\infty} f(z\Theta_u)\, \alpha z^{-\alpha-1}\, \mathrm{d}z\right]$$

for every $u \in V$ and for every Borel measurable $f : [0, \infty)^V \setminus \{0\} \to [0, \infty]$, with Θ_u the tail tree in (7). Moreover, all tail trees are connected through (21).

Proof. The function f_v is regularly varying at infinity with index $-\alpha - 1$ too. By Karamata's theorem [3, Proposition 1.5.10], we have $t f_v(t)/\overline{F}_v(t) \to \alpha$ and thus $\overline{F}_v(t)/\{t g(t)\} \to c_v/\alpha$ as $t \to \infty$.

Condition 2 with $U = V$ follows from Condition 3 and Scheffé's theorem. Part (ii) of Condition 2 is void, since $\mu_e(\{0\}) = 0$ for every $e \in E$.

If $a, b \in V$ are neighbours, we can apply (29) to $g(y) = \mathbb{1}_{(0,z)}(y)$, where $z \in (0, \infty)$, to find

$$c_b \int_0^z q_{b,a}(y)\, \mathrm{d}y = c_a \int_{1/z}^{\infty} y^{\alpha}\, q_{a,b}(y)\, \mathrm{d}y = c_a \int_0^z y^{-\alpha-2}\, q_{a,b}(y^{-1})\, \mathrm{d}y.$$

Since this is true for every $z \in (0, \infty)$, we must have $c_b\, q_{b,a}(y) = c_a\, y^{-\alpha-2}\, q_{a,b}(y^{-1})$ for almost every $y \in (0, \infty)$, whence (32).

Since $\mu_{(b,a)}$ does not have an atom at 0, the identity (29) with $g = \mathbb{1}_{(0,\infty)}$ implies that $\mathbb{E}[M_{(a,b)}^{\alpha}] = c_b/c_a$. Apply (30) and the observation on the line just below that equation to see that $\mathbb{E}[\Theta_{u,v}^{\alpha}] = c_v/c_u$ for all $u, v \in V$. By Corollary 3, all tail trees are then connected via (21).

Finally, $\mathcal{M}_0$-convergence to ν with the stated expression follows from Theorem 2 and Corollary 4.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

Example 4. *In Example 3, assume that A is twice continuously differentiable on (0, 1) and that $A'(0) = -1$ and $A'(1) = 1$. The distribution of (X, Y) is then absolutely continuous and the conditional density of Y/x given that $X = x$ converges as $x \to \infty$ to the function*

$$q(z) = w^3 A''(w), \qquad z \in (0, \infty), \ \ w = 1/(1+z).$$

The conditions on A imply that $\int_0^\infty q(z)\,\mathrm{d}z = \int_0^1 w\,A''(w)\,\mathrm{d}w = 1$ and $\int_0^\infty z\,q(z)\,\mathrm{d}z = \int_0^1 (1-w)\,A''(w)\,\mathrm{d}w = 1$, so that q is a probability density function with first moment equal to 1. Moreover, replacing the function A by the Pickands dependence function $w \mapsto A(1-w)$ amounts to changing q by the function $z \mapsto z^{-3}q(z^{-1})$, in line with (32) with $c_a = c_b = 1$ and $\alpha = 1$.

An interesting example in this respect is the bivariate Hüsler–Reiss distribution [19] with Pickands dependence function

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

for $0 < w < 1$. Here $\lambda \in (0, \infty)$ is a parameter and $\Phi(z) = \int_{-\infty}^{z} (2\pi)^{-1/2} \exp(-z^2/2)\,\mathrm{d}z$ is the standard normal cumulative distribution function. After tedious calculations, we find that q is given by the density of the lognormal random variable $M = \exp\{2\lambda(Z-\lambda)\}$, where Z is a standard normal random variable. Note indeed that $\mathbb{E}[M] = 1$. Moreover, the density function satisfies $z\,q(z) = z^{-2}q(z^{-1})$ for all $z \in (0, \infty)$, which is (32) with $q_{a,b} = q_{b,a} = q$, $c_a = c_b = 1$, and $\alpha = 1$. This also follows from the symmetry of the Hüsler–Reiss Pickands dependence function; i.e., $A(w) = A(1-w)$ for all $w \in [0, 1]$, so that the pair (X, Y) is exchangeable.

If all neighbouring pairs (X_a, X_b) for $(a, b) \in E$ of the Markov tree follow such Hüsler–Reiss max-stable distributions, then the joint distribution of the tail tree is multivariate log-normal, since $\log \Theta_{u,v} = \sum_{e \in [u \rightsquigarrow v]} \log M_e$ for all $u, v \in V$, where the random variables $\log M_e$ are independent and normally distributed with expectation $-2\lambda_e^2$ and variance $4\lambda_e^2$, with dependence parameter $\lambda_e \in (0, \infty)$ for all $e \in E$.

Appendix A. Proofs for Section 3

Proof of Theorem 2. (a) and (b) are equivalent. — Clearly, (b) implies (a). Conversely, assume (a); let us show (b). Let $z \in [1, \infty)$ and let $\theta \in \mathbb{S}$ be such that $\mathbb{P}(\Theta_j = \theta_j) = 0$ for all $j \in \{1, \ldots, d\}$. We have

$$\begin{aligned}\mathbb{P}(X_i/t > z,\, X/X_i \leqslant \theta \mid X_i > t) &= \frac{b(zt)\overline{F}_i(zt)}{b(t)\overline{F}_i(t)}\,\frac{b(t)}{b(zt)}\,\mathbb{P}(X/X_i \leqslant \theta \mid X_i/t > z) \\ &\to z^{-\alpha}\mathbb{P}(\Theta_i \leqslant \theta), \qquad t \to \infty.\end{aligned}$$

It follows that $\mathbb{P}(X_i/t \leqslant z,\, X/X_i \leqslant \theta \mid X_i > t) \to \mathbb{P}(Z \leqslant z)\mathbb{P}(\Theta_i \leqslant \theta)$ as $t \to \infty$, where Z is a $\mathrm{Pa}(\alpha)$ random variable.

(b) implies (c) and (i). — Since $(X/t) = (X_i/t)(X/X_i)$, the statement (b) and the continuous mapping theorem [37, Theorem 2.3] imply that $\mathcal{L}(X/t \mid X_i > t)$ converges weakly to $Y_i = Z\Theta_i$, where Z is a $\mathrm{Pa}(\alpha)$ random variable independent of Θ_i. Since $\Theta_{i,i} = 1$ almost surely, we have $Y_{i,i} = Z$.

(c) implies (a). — Since $X/X_i = (X/t)/(X_i/t)$, the statement (c) and the continuous mapping theorem imply the statement (a) with $\Theta_i = Y_i/Y_{i,i}$.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

(b) implies (d). — Define a Borel measure ν_i on $\mathbb{S}_{0,i}$ by

$$\nu_i(\,\cdot\,) = c_i \int_0^\infty \mathbb{P}(z\Theta_i \in \,\cdot\,)\, \alpha z^{-\alpha-1}\, \mathrm{d}z.$$

If B is a Borel subset of $\mathbb{S}_{0,i}$ contained in $\{x \in \mathbb{S}_{0,i} : x_i \geqslant \varepsilon\}$ for some $\varepsilon > 0$, then $\mathbb{P}(z\Theta_i \in B) = 0$ as soon as $z < \varepsilon$, since $\Theta_{i,i} = 1$ almost surely. As a consequence, $\nu_i(B) \leqslant c_i \varepsilon^{-\alpha}$ for such B. It follows that $\nu_i \in \mathcal{M}_{0,i}$.

By linearity of the integral and by monotone convergence, we find that

$$\nu_i(f) = c_i \int_0^\infty \mathbb{E}[f(z\Theta_i)]\, \alpha z^{-\alpha-1}\, \mathrm{d}z \tag{33}$$

for every nonnegative Borel measurable function f on $\mathbb{S}_{0,i}$. The same expression is then true for real-valued Borel measurable functions f on $\mathbb{S}_{0,i}$ for which at least one of the two integrals with f replaced by $|f|$ is finite. This includes bounded, Borel measurable functions that vanish on a set of the form $\{x \in \mathbb{S}_{0,i} : x_i \leqslant \varepsilon\}$ for some $\varepsilon > 0$.

Let $f \in \mathcal{C}_{0,i}$ and let $\varepsilon > 0$ be such that $f(x) = 0$ as soon as $x_i \leqslant \varepsilon$. By (b), we have

$$\begin{aligned} b(t)\mathbb{E}[f(X/t)] &= b(t)\mathbb{E}[f((X_i/t)(X/X_i))\, \mathbb{1}(X_i/t > \varepsilon)] \\ &= b(\varepsilon t)\overline{F}_i(\varepsilon t)\frac{b(t)}{b(\varepsilon t)}\mathbb{E}[f(\varepsilon(X_i/(\varepsilon t))(X/X_i)) \mid X_i > \varepsilon t] \\ &\to c_i \varepsilon^{-\alpha}\mathbb{E}[f(\varepsilon Z\Theta_i)], \qquad t \to \infty, \end{aligned}$$

where Z is a $\mathrm{Pa}(\alpha)$ random variable, independent of Θ_i. The limit is equal to

$$c_i\varepsilon^{-\alpha}\int_1^\infty \mathbb{E}[f(\varepsilon z\Theta_i)]\, \alpha z^{-\alpha-1}\, \mathrm{d}z = c_i \int_\varepsilon^\infty \mathbb{E}[f(z\Theta_i)]\, \alpha z^{-\alpha-1}\, \mathrm{d}z = \nu_i(f),$$

since $f(z\Theta_i) = 0$ almost surely whenever $z \leqslant \varepsilon$, as $\Theta_{i,i} = 1$ almost surely.

(d) implies (c). — For $z \in (0, \infty)$, we have $b(t)\mathbb{P}(X_i/t > z) \to c_i z^{-\alpha}$ as $t \to \infty$, and thus $\nu_i(\{x : x_i > z\}) = c_i z^{-\alpha}$ by (d). For open $G \subset \mathbb{R}^d$, the portmanteau theorem [25, Theorem 2.1(iii)] yields

$$\begin{aligned} \liminf_{t\to\infty} \mathbb{P}(X/t \in G \mid X_i > t) &= \liminf_{t\to\infty} \frac{1}{b(t)\overline{F}_i(t)} b(t)\mathbb{P}(X/t \in G \cap \{x : x_i > 1\}) \\ &\geqslant c_i^{-1}\nu_i(G \cap \{x : x_i > 1\}). \end{aligned}$$

By the portmanteau lemma for weak convergence [37, Lemma 2.2] we obtain (c) where the law of Y_i is $\mathbb{P}(Y_i \in \,\cdot\,) = c_i^{-1}\nu_i(\,\cdot\, \cap \{x : x_i > 1\})$.

(d) implies (e). — For every $z > 0$, we have

$$\begin{aligned} b(t)\mathbb{P}[\max(X_I)/t > z] &\leqslant b(t)\sum_{i\in I}\mathbb{P}(X_i/t > z) \\ &= \frac{b(t)}{b(zt)}\sum_{i\in I} b(zt)\overline{F}_i(zt) \to z^{-\alpha}\sum_{i\in I} c_i, \qquad t \to \infty. \end{aligned}$$

Since the limit is finite for every $z > 0$ and since it converges to zero as $z \to \infty$, it follows by the relative compactness criterion in [25, Theorem 2.5] that for every sequence $(t_n)_n$ tending to

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

infinity, there exists a subsequence along which $b(t_n)\mathbb{P}(X/t_n \in \cdot\,)$ converges in $\mathcal{M}_{0,I}$. To show (e), we then need to show that these subsequence limits must coincide. To do so, we show that for every $f \in \mathcal{C}_{0,I}$, the limit of $b(t)\mathbb{E}[f(X/t)]$ exists as $t \to \infty$. This fixes the value of the integral of such f with respect to all subsequence limits, which then must be the same.

For $\varepsilon > 0$, let $h_\varepsilon : [0, \infty) \to [0, 1]$ be the piecewise linear function

$$h_\varepsilon(t) = \min\{\max(2t/\varepsilon - 1, 0), 1\} = \begin{cases} 0 & \text{if } t \in [0, \varepsilon/2], \\ 2t/\varepsilon - 1 & \text{if } t \in [\varepsilon/2, \varepsilon], \\ 1 & \text{if } t \in [\varepsilon, \infty). \end{cases}$$

Put $\hbar_\varepsilon = 1 - h_\varepsilon$. Write $I = \{i_1, \dots, i_k\}$. Then

$$\begin{aligned} 1 &= h_\varepsilon(x_{i_1}) + \hbar_\varepsilon(x_{i_1}) \\ &= h_\varepsilon(x_{i_1}) + \hbar_\varepsilon(x_{i_1})h_\varepsilon(x_{i_2}) + \hbar_\varepsilon(x_{i_1})\hbar_\varepsilon(x_{i_2}) \\ &= \dots \\ &= \sum_{\ell=1}^{k} \left(\prod_{m=1}^{\ell-1} \hbar_\varepsilon(x_{i_m}) \right) h_\varepsilon(x_{i_\ell}) + \prod_{\ell=1}^{k} \hbar_\varepsilon(x_{i_\ell}). \end{aligned}$$

For $f \in \mathcal{C}_{0,I}$ we can find $\varepsilon > 0$ such that $f(x) = 0$ if $\max(x_{i_1}, \dots, x_{i_k}) \leqslant \varepsilon$. Then $f(x) \prod_{\ell=1}^{k} \hbar_\varepsilon(x_{i_\ell}) = 0$ for all x, and thus $f = \sum_{i \in I} f_i$ where, for $\ell \in \{1, \dots, k\}$, we have

$$f_{i_\ell}(x) = f(x) \left(\prod_{m=1}^{\ell-1} \hbar_\varepsilon(x_{i_m}) \right) h_\varepsilon(x_{i_\ell}).$$

Each function f_i belongs to $\mathcal{C}_{0,I}$ too but has moreover the property that $f_i(x) = 0$ as soon as $x_i \leqslant \varepsilon/2$. The restriction of f_i to $\mathbb{S}_{0,i}$ thus belongs to $\mathcal{C}_{0,i}$. By (d),

$$b(t)\mathbb{E}[f(X/t)] = \sum_{i \in I} b(t)\mathbb{E}[f_i(X/t)] \to \sum_{i \in I} \nu_i(f_i), \qquad t \to \infty.$$

The existence of a limit has thus been shown, and convergence in $\mathcal{M}_{0,I}$ to some measure ν as stated in (e) follows.

(e) implies (d), (ii), (iii), and (iv). — A function f in $\mathcal{C}_{0,i}$ can be extended to a function in $\mathcal{C}_{0,I}$ denoted by the same symbol by putting $f(x) = 0$ for $x \in \mathbb{S}_{0,I} \setminus \mathbb{S}_{0,i}$. Hence, (e) implies (d), with ν_i as described in (ii).

The statement (iii) follows from (ii) and the description of the law of Y_i in terms of ν_i in the proof above of the fact that (d) implies (c).

Similarly, (iv) follows from (ii), Equation (33), and Fubini's theorem.

Proof of Theorem 3. It is sufficient to show the statement (a) in Theorem 2. By the property (i) in Theorem 2, the weak convergence in Theorem 3(a) already holds for all $i \in K$, and we need to show that it also holds for all $j \in I \setminus K$. Choose $j \in I \setminus K$ and let $i = i(j) \in K$ be as in the property (ii) of Theorem 3.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

We will show that $\mathcal{L}(X/X_j \mid X_j > t)$ converges weakly as $t \to \infty$ to Θ_j whose law is defined in (21). Let $G \subset \mathbb{S}$ be open and let $\delta > 0$. We have

$$\begin{aligned}\mathbb{P}(X/X_j \in G \mid X_j > t) &\geqslant \mathbb{P}(X/X_j \in G, X_i > \delta t \mid X_j > t) \\ &= \frac{b(\delta t)\overline{F}_i(\delta t)}{b(t)\overline{F}_j(t)} \frac{b(t)}{b(\delta t)} \mathbb{P}\left[\frac{X/X_i}{X_j/X_i} \in G,\ \delta \frac{X_i}{\delta t} \frac{X_j}{X_i} > 1 \,\middle|\, X_i > \delta t \right].\end{aligned}$$

By Theorem 2 applied to K, we have $\mathcal{L}(X_i/s, X/X_i \mid X_i > s) \xrightarrow{d} \mathrm{Pa}(\alpha) \otimes \mathcal{L}(\Theta_i)$ as $s \to \infty$. Let Z be a $\mathrm{Pa}(\alpha)$ random variable, independent of Θ_i. By the portmanteau lemma for weak convergence, we have

$$\begin{aligned}\liminf_{t\to\infty} \mathbb{P}(X/X_j \in G \mid X_j > t) &\geqslant \frac{c_i}{c_j} \delta^{-\alpha} \mathbb{P}[\Theta_i/\Theta_{i,j} \in G,\ \delta Z \Theta_{i,j} > 1] \\ &= \mathbb{E}[\mathbb{1}\{\Theta_i/\Theta_{i,j} \in G\} \min(\Theta_{i,j}^{\alpha}, \delta^{-\alpha})]/\mathbb{E}[\Theta_{i,j}^{\alpha}].\end{aligned}$$

The equality on the second line follows from (ii) and the fact that $Z^{-\alpha}$ is uniformly distributed on $(0, 1)$ and independent of Θ_i. Since $\delta > 0$ was arbitrary, the monotone convergence theorem yields

$$\liminf_{t\to\infty} \mathbb{P}(X/X_j \in G \mid X_j > t) \geqslant \mathbb{E}[\mathbb{1}\{\Theta_i/\Theta_{i,j} \in G\}\, \Theta_{i,j}^{\alpha}]/\mathbb{E}[\Theta_{i,j}^{\alpha}].$$

Apply the portmanteau lemma for weak convergence once more to obtain the stated weak convergence.

Proof of Theorem 4. The properties of ρ imply that $\mathbb{S}_\rho$ is open and non-empty and that $0 < \nu(\mathbb{S}_\rho) < \infty$. The boundary of $\mathbb{S}_\rho$ is $\{x \in \mathbb{S}_{0,I} : \rho(x_I) = 1\}$, which is a ν-null set, since its ν-measure is bounded by the sum over $i \in I$ of $\nu(\{x \in \mathbb{S}_{0,I} : \rho(x_I) = 1, x_i > 0\})$, which is zero by (17). Since $\rho(X_I) > t$ if and only if $X_I/t \in \mathbb{S}_\rho$, the portmanteau theorem [25, Theorem 2.1(iv)] implies $b(t)\mathbb{P}[\rho(X_I) > t] \to \nu(\mathbb{S}_\rho)$ as $t \to \infty$.

Let $G \subset \mathbb{R}^d$ be open. By (iii) of the same portmanteau theorem,

$$\liminf_{t\to\infty} \mathbb{P}(X/t \in G \mid \rho(X_I) > t) = \liminf_{t\to\infty} \frac{b(t)\mathbb{P}(X/t \in G \cap \mathbb{S}_\rho)}{b(t)\mathbb{P}(X \in \mathbb{S}_\rho)} \geqslant \nu(G \cap \mathbb{S}_\rho)/\nu(\mathbb{S}_\rho).$$

The portmanteau theorem for weak convergence implies the stated weak convergence of $\mathcal{L}(X/t \mid \rho(X) > t)$ as $t \to \infty$. This proves the statement (c) in Theorem 2 for the enlarged random vector $(X, \rho(X))$.

Acknowledgements

The author is grateful to two anonymous reviewers whose suggestions have led to various improvements throughout the text. The author also wishes to thank Stefka Asenova and Gildas Mazo for inspiring discussions.

References

[1] Asadi, P., Davison, A. C. and Engelke, S. (2015). Extremes on river networks. *Ann. Appl. Statist.* **9**, 2023–2050.

[2] Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stoch. Proc. Appl.* **119**, 1055–1080.

[3] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

[4] Bortot, P. and Coles, S. (2000). A sufficiency property arising from the characterization of extremes of a Markov chain. *Bernoulli* **6**, 183–190.
[5] Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Statist. Soc. B [Statist. Methodology]* **53**, 377–392.
[6] Das, B. and Resnick, S. I. (2011). Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli* **17**, 226–252.
[7] Dombry, C., Hashorva, E. and Soulier, P. (2018). Tail measure and spectral tail process of regularly varying time series. *Ann. Appl. Prob.* **28**, 3884–3921.
[8] Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statist. Interface* **8**, 9–17.
[9] Einmahl, J. H. J., Kiriliouk, A. and Segers, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes* **21**, 205–233.
[10] Engelke, S. and Hitz, A. S. (2019). Graphical models for extremes. To appear in *J. R. Statist. Soc. B [Statist. Methodology]*. Preprint available at https://arxiv.org/abs/1812.01734.
[11] Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20**, 1717–1737.
[12] Garralda Guillem, A. I. (2000). Structure de dépendance des lois de valeurs extrêmes bivariées. *C. R. Acad. Sci. Paris* **330**, 593–596.
[13] Gissibl, N. and Klüppelberg, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli* **24**, 2693–2720.
[14] Gissibl, N., Klüppelberg, C. and Otto, M. (2018). Tail dependence of recursive max-linear models with regularly varying noise variables. *Econometrics Statist.* **6**, 149–167.
[15] Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications*, eds. P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, Springer, Berlin, 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] Hitz, S. A. and Evans, J. R. (2016). One-component regular variation and graphical modeling of extremes. **53**, 733–746.
[18] Hult, H. and Lindskog, F. (2006). Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N. S.)* **80**, 121–140.
[19] Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Prob. Lett.* **7**, 283–286.
[20] Janssen, A. (2018). Spectral tail processes and max-stable approximations of multivariate regularly varying time series. *Stoch. Proc. Appl.* **129**, 1993–2009.
[21] Janssen, A. and Segers, J. (2014). Markov tail chains. *J. Appl. Prob.* **51**, 1133–1153.
[22] Kiriliouk, A., Rootzén, H., Segers, J. and Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics* **61**, 123–135.
[23] Lauritzen, S. L. (1996). *Graphical Models*. Oxford Science Publications, Oxford.
[24] Lee, D. and Joe, H. (2018). Multivariate extreme value copulas with factor and tree dependence structures. *Extremes* **21**, 147–176.
[25] Lindskog, F., Resnick, S. I. and Roy, J. (2014). Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Prob. Surveys* **11**, 270–314.
[26] Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Prob.* **4**, 529–548.
[27] Planinić, H. and Soulier, P. (2018). The tail process revisited. *Extremes* **21**, 551–579.
[28] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York.
[29] Resnick, S. I. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer-Verlag, New York.
[30] Resnick, S. I. and Zeber, D. (2014). Transition kernels and the conditional extreme value model. *Extremes* **17**, 263–287.
[31] Rootzén, H., Segers, J. and Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *J. Multivariate Anal.* **165**, 117–131.
[32] Samorodnitsky, G. and Owada, T. (2012) Tail measures of stochastic processes or random fields with regularly varying tails. Preprint.
[33] Segers, J. (2007). Multivariate regular variation of heavy-tailed Markov chains. Tech. rep. DP0703, Institut de statistique, Université catholique de Louvain. Available at https://arxiv.org/abs/0701411.

https://doi.org/10.1017/apr.2020.22 Published online by Cambridge University Press

[34] SEGERS, J., ZHAO, Y. AND MEINGUET, T. (2017). Polar decomposition of regularly varying time series in star-shaped metric spaces. *Extremes* **20**, 539–566.
[35] SMITH, R. L. (1992). The extremal index for a Markov chain. *J. Appl. Prob.* **29**, 37–45.
[36] SMITH, R. L., TAWN, J. A. AND COLES, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84**, 249–268.
[37] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
[38] 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/apr.2020.22 Published online by Cambridge University Press