

ASYMPTOTICS OF MARKOV KERNELS AND THE TAIL CHAIN

SIDNEY I. RESNICK * AND
DAVID ZEBER, ** *Cornell University*

Abstract

An asymptotic model for the extreme behavior of certain Markov chains is the ‘tail chain’. Generally taking the form of a multiplicative random walk, it is useful in deriving extremal characteristics, such as point process limits. We place this model in a more general context, formulated in terms of extreme value theory for transition kernels, and extend it by formalizing the distinction between extreme and nonextreme states. We make the link between the update function and transition kernel forms considered in previous work, and we show that the tail chain model leads to a multivariate regular variation property of the finite-dimensional distributions under assumptions on the marginal tails alone.

Keywords: Extreme values; multivariate regular variation; Markov chain; transition kernel; tail chain; heavy tail

2010 Mathematics Subject Classification: Primary 60G70; 60J05
Secondary 62P05

1. Introduction

A method of approximating the extremal behavior of discrete-time Markov chains is to use an asymptotic process called the *tail chain* under an asymptotic assumption on the transition kernel of the chain. Loosely speaking, if the distribution of the next state converges under some normalization as the current state becomes extreme, then the Markov chain behaves approximately as a multiplicative random walk upon leaving a large initial state. This approach leads to intuitive extremal models in such cases as autoregressive processes with random coefficients, which include a class of ARCH models. The focus on Markov kernels was introduced by Smith [23]. Perfekt [17], [18] extended the approach to higher dimensions, and Segers [22] rephrased the conditions in terms of update functions. Further extensions were discussed in [6], [7], [24], [25], and [26].

Though not restrictive in practice, the previous approach tends to mask aspects of the processes’ extremal behavior. Markov chains that admit the tail chain approximation fall into one of two categories. Starting from an extreme state, the chain either remains extreme over any finite time horizon, or will drop to a ‘nonextreme’ state of lower order after a finite amount of time. The latter case is problematic in that the tail chain model is not sensitive to possible subsequent jumps from a nonextreme state to an extreme one. Previous developments handle this by ruling out the class of processes exhibiting this behavior via a technical condition, which

Received 3 January 2012; revision received 2 May 2012.

* Postal address: School of Operations Research and Industrial Engineering, Cornell University, 284 Rhodes Hall, Ithaca, NY 14853, USA. Email address: sir1@cornell.edu

** Postal address: Department of Statistical Science, Cornell University, 301 Malott Hall, Ithaca, NY 14853, USA. Email address: dsz5@cornell.edu

we refer to as the *regularity condition*. Also, most previous work has assumed stationarity, since interest focused on computing the extremal index or deriving limits for the exceedance point processes, drawing on the theory established for stationary processes with mixing by Leadbetter *et al.* [16] and its extensions [10], [14]. However, stationarity is not fundamental in determining the extremal behavior of the finite-dimensional distributions (FDDs).

We place the tail chain approximation in the context of an extreme value theory for Markovian transition kernels, which *a priori* does not necessitate any such restrictions on the class of processes to which it may be applied. Although distributional convergence results are more naturally phrased in terms of transition kernels, we treat the equivalent update function forms as an integral component to interfacing with applications, and we phrase relevant assumptions in terms of both. While not making explicit a complete tail chain model for the class of chains excluded previously, we demonstrate the extent to which previous models may be viewed as a partial approximation within our framework. This is accomplished by formalizing the division between extreme and nonextreme states as a level we term the *extremal boundary*. We show that, in general, the tail chain approximates the *extremal component*, the portion of the original chain having yet to cross below this boundary. Phrased in these terms, the regularity condition requires that the distinction between the original chain and its extremal component disappears asymptotically.

After introducing our extreme value theory for transition kernels, along with a representation in terms of update functions, we derive limits of FDDs conditional on the initial state, as it becomes extreme. We then examine the effect of the regularity condition on these results. Finally, adding the assumption of marginal regularly varying tails leads to convergence results for the unconditional distributions akin to regular variation.

1.1. Notation and conventions

We review notation and relevant concepts. If not explicitly specified, assume that any space \mathbb{S} under discussion is a topological space paired with its Borel σ -field, $\mathcal{B}(\mathbb{S})$, generated by open sets. Denote by $\mathcal{K}(\mathbb{S})$ the collection of its compact sets; by $\mathcal{C}(\mathbb{S})$ the space of real-valued continuous, bounded functions on \mathbb{S} ; and by $\mathcal{C}_K^+(\mathbb{S})$ the space of nonnegative continuous functions with compact support. Weak convergence of probability measures is denoted by ‘ \Rightarrow ’.

For a space \mathbb{E} which is locally compact with countable base (for example, a subset of $[-\infty, \infty]^d$), $\mathbb{M}_+(\mathbb{E})$ is the space of nonnegative Radon measures on $\mathcal{B}(\mathbb{E})$; point measures consisting of single point masses at x will be written as $\varepsilon_x(\cdot)$. A sequence of measures $\{\mu_n\} \subset \mathbb{M}_+(\mathbb{E})$ converges vaguely to $\mu \in \mathbb{M}_+(\mathbb{E})$ (written $\mu_n \xrightarrow{v} \mu$) if $\int_{\mathbb{E}} f \, d\mu_n \rightarrow \int_{\mathbb{E}} f \, d\mu$ as $n \rightarrow \infty$ for any $f \in \mathcal{C}_K^+(\mathbb{E})$. The shorthand $\mu(f) = \int f \, d\mu$ is handy. That the distribution of a random vector \mathbf{X} is regularly varying on a cone $\mathbb{E} \subset [-\infty, \infty]^d \setminus \{\mathbf{0}\}$ means that $t P[\mathbf{X}/b(t) \in \cdot] \xrightarrow{v} \mu^*(\cdot)$ in $\mathbb{M}_+(\mathbb{E})$ as $t \rightarrow \infty$ for some nondegenerate limit measure $\mu^* \in \mathbb{M}_+(\mathbb{E})$ and scaling function $b(t) \rightarrow \infty$. The limit μ^* is necessarily homogeneous in the sense that $\mu^*(c \cdot) = c^{-\alpha} \mu^*(\cdot)$ for some $\alpha > 0$. For further details on vague convergence and regular variation, see [15] or [19].

If $\mathbf{X} = (X_0, X_1, X_2, \dots)$ is a (homogeneous) Markov chain and K is a Markov transition kernel, we write $\mathbf{X} \sim K$ to mean that the dependence structure of \mathbf{X} is specified by K , i.e.

$$P[X_{n+1} \in \cdot \mid X_n = x] = K(x, \cdot), \quad n = 0, 1, \dots$$

We adopt the standard shorthand $P_x[(X_1, \dots, X_m) \in \cdot] = P[(X_1, \dots, X_m) \in \cdot \mid X_0 = x]$. Some useful technical results are assembled in Section A.

2. Extremal theory for Markov kernels

We begin by focusing on the Markov transition kernels rather than the stochastic processes they determine, and introduce a class of kernels we term ‘tail kernels’, which we will view as scaling limits of certain kernels. Antecedents include Segers’ [22] definition of ‘back-and-forth tail chains’ that approximate certain Markov chains started from an extreme value.

For a Markov chain $X \sim K$ on $[0, \infty)$, it is reasonable to expect that the extremal behavior of X is determined by pairs (X_n, X_{n+1}) , and one way to control such pairs is to assume that (X_n, X_{n+1}) belongs to a bivariate domain of attraction (cf. [5] and [23]). In the context of regular variation, writing

$$t P \left[\frac{X_n}{b(t)} \in A_0, \frac{X_{n+1}}{b(t)} \in A_1 \right] = \int_{A_0} K(b(t)u, b(t)A_1) t P \left[\frac{X_n}{b(t)} \in du \right] \tag{2.1}$$

suggests combining the marginal regular variation of X_n with a scaling kernel limit to derive extremal properties of the FDDs [17], [18], [22], and this is the direction we take. We first discuss the kernel scaling operation.

For simplicity, we assume that the state space of the Markov chain is $[0, \infty)$, although with suitable modifications, it is relatively straightforward to extend the results to \mathbb{R}^d . Henceforth, G will denote a general probability distribution on $[0, \infty)$.

2.1. Tail kernels

Definition 2.1. The tail kernel associated with G is given by

$$K^*(y, A) = \begin{cases} G(y^{-1}A), & y > 0, \\ \varepsilon_0(A), & y = 0, \end{cases} \tag{2.2}$$

for measurable A .

Recall that ε_0 is the probability measure assigning unit mass to $\{0\}$. Thus, the class of tail kernels on $[0, \infty)$ is parameterized by probability distributions G . Such kernels are characterized by a scaling property.

Proposition 2.1. A Markov transition kernel K is a tail kernel associated with some distribution G if and only if it satisfies the relation

$$K(uy, A) = K(y, u^{-1}A) \tag{2.3}$$

with measurable A and $y \geq 0$ for any $u > 0$, in which case $G(\cdot) = K(1, \cdot)$.

Proof. Fix $u > 0$. If K is a tail kernel, (2.3) follows directly from the definition, using the fact that $\varepsilon_0(A) = \varepsilon_0(u^{-1}A)$ in the case $y = 0$. Conversely, assuming that (2.3) holds, for $y > 0$, we can write $K(y, A) = K(1, y^{-1}A)$, satisfying the first case of (2.2) with $G(\cdot) = K(1, \cdot)$. For the case $y = 0$, write $H(\cdot) = K(0, \cdot)$. We must show that $H(\cdot) = H(u^{-1}\cdot)$ implies that $H = \varepsilon_0$. Indeed, $H(0, \infty) = \lim_{n \rightarrow \infty} H(n^{-1}, \infty) = H(1, \infty)$, so $H(0, 1] = 0$. A similar argument shows that $H(1, \infty) = 0$ as well.

We call the Markov chain $T \sim K^*$ the tail chain associated with G . Such a chain can be represented as

$$T_n = \xi_n T_{n-1} = T_0 \xi_1 \cdots \xi_n, \quad n = 1, 2, \dots, \tag{2.4}$$

where the ξ_n are independent and identically distributed (i.i.d.) copies of $\xi \sim G$ and independent of T_0 . Thus, T is a multiplicative random walk with step distribution G and absorbing barrier at $\{0\}$.

2.2. Convergence to tail kernels

The tail chain approximates the behavior of a Markov chain $X \sim K$ in extreme states. Asymptotic results require that the normalized distribution of X_1 be well approximated by some distribution G when X_0 is large, and we interpret this requirement as a domain of attraction condition for kernels.

Definition 2.2. A Markov transition kernel $K : [0, \infty) \times \mathcal{B}[0, \infty) \rightarrow [0, 1]$ is in the *domain of attraction of G* , written $K \in D(G)$, if, as $t \rightarrow \infty$,

$$K(t, t \cdot) \Rightarrow G(\cdot) \quad \text{on } [0, \infty]. \tag{2.5}$$

Note that $D(G)$ contains at least the tail kernel associated with G . A simple scaling argument extends (2.5) to

$$K(tu, t \cdot) \Rightarrow G(u^{-1} \cdot) = K^*(u, \cdot), \quad u > 0, \tag{2.6}$$

where K^* is the tail kernel associated with G ; this is the form appearing in (2.1). Thus, tail kernels are scaling limits for kernels in a domain of attraction. In fact, tail kernels are the only possible limits.

Proposition 2.2. *Let K be a transition kernel on $[0, \infty)$. If, for each $u > 0$, there exists a distribution G_u such that $K(tu, t \cdot) \Rightarrow G_u(\cdot)$ as $t \rightarrow \infty$ then the function \hat{K} defined on $[0, \infty) \times \mathcal{B}[0, \infty)$ as*

$$\hat{K}(u, A) := \begin{cases} G_u(A), & u > 0, \\ \varepsilon_0(A), & u = 0, \end{cases}$$

is the tail kernel associated with G_1 .

Proof. It suffices to show that $G_u(\cdot) = G_1(u^{-1} \cdot)$ for any $u > 0$. But this follows directly from the uniqueness of weak limits, since (2.6) shows that $K(tu, t \cdot) \Rightarrow G_1(u^{-1} \cdot)$.

A version of (2.6) uniform in u is needed to derive the convergence of FDDs.

Proposition 2.3. *Suppose that $K \in D(G)$ and that K^* is the tail kernel associated with G . Then, for any $u > 0$ and any nonnegative function $u_t = u(t)$ such that $u_t \rightarrow u$ as $t \rightarrow \infty$, we have*

$$K(tu_t, t \cdot) \Rightarrow K^*(u, \cdot) \quad \text{as } t \rightarrow \infty. \tag{2.7}$$

Proof. Suppose that $u_t \rightarrow u > 0$. Observe that $K(tu_t, t \cdot) = K(tu_t, (tu_t)^{-1} \cdot)$, and put $h_t(x) = u_t x$, $h(x) = ux$. Writing $P_t(\cdot) = K(tu_t, tu_t \cdot)$, we have

$$K(tu_t, t \cdot) = P_t \circ h_t^{-1} \Rightarrow G \circ h^{-1} = G(u^{-1} \cdot) = K^*(u, \cdot)$$

by [2, Theorem 5.5, p. 34].

The measure G controls X upon leaving an extreme state via (2.5). However, (2.5) is uninformative once X has reached a ‘nonextreme’ state; indeed, (2.7) may fail if $u = 0$ —see Example 6.2 below. The requirement $K^*(0, \cdot) = \varepsilon_0$ reflects an assumption that a transition from nonextreme back to extreme is (asymptotically) impossible. The implications of such an assumption cannot be ignored if 0 is an accessible point of the state space, i.e. if $G(\{0\}) = K^*(y, \{0\}) > 0$.

One could conceivably extend this model to accommodate more general behavior upon leaving a nonextreme state by allowing $K^*(0, A) = H(A)$ in (2.2), where H is a general

probability distribution. The appropriate choice of H would then be determined via an asymptotic assumption analogous to (2.7); for example, $K(tu_t, t \cdot) \Rightarrow H(\cdot)$ whenever $u_t \rightarrow 0$. In Section 4 we show that the ‘regularity condition’ imposed by previous authors does in fact justify the choice $H = \varepsilon_0$ in this manner (see (4.3) below). We will not pursue this direction here. Instead, we proceed without assuming the regularity condition, viewing the prescription $H = \varepsilon_0$ as ‘uninformative’, in a sense to be made precise in Section 3.

Alternative formulations of (2.5) include replacing $K(t, t \cdot)$ with $K(t, a(t) \cdot)$ or else $K(t, a(t) \cdot + b(t))$ with appropriate functions $a(t) > 0$ and $b(t)$, in analogy with the usual domains of attraction conditions in extreme value theory. The second choice coincides with the presentation by Perfekt [17], and relates to the conditional extreme value model [8], [12], [13]. For clarity, and to maintain ties with regular variation, we retain the standard choice $a(t) = t, b(t) = 0$.

2.3. Representation

How do we characterize kernels belonging to $D(G)$? From (2.4), for chains transitioning according to a tail kernel, the next state is a random multiple of the previous one, provided the prior state is nonzero. We expect that chains transitioning according to $K \in D(G)$ behave approximately like this upon leaving a large state, and this is best expressed in terms of a function describing how a new state depends on the prior one.

Given a kernel K , we can always find a sample space \mathbb{E} , a measurable function $\psi: [0, \infty) \times \mathbb{E} \rightarrow [0, \infty)$, and an \mathbb{E} -valued random element V such that $\psi(y, V) \stackrel{D}{=} K(y, \cdot)$ for $y \geq 0$. Given a random variable X_0 , if we define the process $\mathbf{X} = (X_0, X_1, X_2, \dots)$ recursively as

$$X_{n+1} = \psi(X_n, V_{n+1}), \quad n \geq 0,$$

where $\{V_n\}$ is an i.i.d. sequence equal in distribution to V and independent of X_0 , then \mathbf{X} is a Markov chain with transition kernel K . Call the function ψ an *update function corresponding to K* . If, in addition, $K \in D(G)$, the domain of attraction condition (2.5) becomes

$$t^{-1}\psi(t, V) \Rightarrow \xi,$$

where $\xi \sim G$. Applying the probability integral transform or the Skorokhod representation theorems [3, Theorem 3.2, p. 6], [4, Theorem 6.7, p. 70], we get the following result.

Proposition 2.4. *If K is a transition kernel, $K \in D(G)$ if and only if there exists a measurable function $\psi^*: [0, \infty) \times [0, 1] \rightarrow [0, \infty)$ and a random variable $\xi^* \sim G$ on the uniform probability space $([0, 1], \mathcal{B}, \lambda)$ such that*

$$t^{-1}\psi^*(t, u) \rightarrow \xi^*(u) \quad \text{for all } u \in [0, 1] \tag{2.8}$$

as $t \rightarrow \infty$, and ψ^* is an update function corresponding to K in the sense that

$$\lambda[\psi^*(y, \cdot) \in A] = K(y, A)$$

for measurable sets A .

Think of the update function as $\psi^*(y, U)$, where $U(u) = u$ is a uniform random variable on $[0, 1]$.

Proof of Proposition 2.4. If there exist such ψ^* and ξ^* satisfying (2.8), clearly $K \in D(G)$. Conversely, assume that $K \in D(G)$, and let $\psi(\cdot, V)$ be an update function corresponding

to K . According to Skorokhod’s representation theorem (cf. [4, p. 70], with the necessary modifications to allow for an uncountable index set), there exists a random variable ξ^* and a stochastic process $\{Y_t^*; t \geq 0\}$ defined on the uniform probability space $([0, 1], \mathcal{B}, \lambda)$, taking values in $[0, \infty)$, such that

$$\xi^* \sim G, \quad Y_0^* \stackrel{D}{=} \psi(0, V), \quad Y_t^* \stackrel{D}{=} t^{-1}\psi(t, V) \quad \text{for } t > 0,$$

and $Y_t^*(u) \rightarrow \xi^*(u)$ as $t \rightarrow \infty$ for every $u \in [0, 1]$. Now, define $\psi^*: [0, \infty) \times [0, 1] \rightarrow [0, \infty)$ as

$$\psi^*(0, u) = Y_0^*(u) \quad \text{and} \quad \psi^*(t, u) = tY_t^*(u), \quad t > 0,$$

for all $u \in [0, 1]$. It is evident that $\lambda[\psi^*(y, \cdot) \in A] = P[\psi(y, V) \in A]$ for $y \in [0, \infty)$, so ψ^* is indeed an update function corresponding to K , and ψ^* satisfies (2.8) by construction.

Update functions corresponding to K are not unique, and some of them may fail to converge pointwise as in (2.8). However, (2.8) is convenient, and Proposition 2.4 shows that Segers’ [22] Condition 2.2 in terms of update functions is equivalent to our weak convergence formulation $K \in D(G)$.

Pointwise convergence in (2.8) gives an intuitive representation of kernels in a domain of attraction.

Corollary 2.1. *$K \in D(G)$ if and only if there exists a random variable $\xi \sim G$ defined on the uniform probability space, and a measurable function $\phi: [0, \infty) \times [0, 1] \mapsto (-\infty, \infty)$ satisfying $t^{-1}\phi(t, u) \rightarrow 0$ for all $u \in [0, 1]$ such that*

$$\psi(y, u) := \xi(u)y + \phi(y, u)$$

is an update function corresponding to K .

Proof. If such ξ and ϕ exist, then $t^{-1}\psi(t, u) = \xi(u) + t^{-1}\phi(t, u) \rightarrow \xi(u)$ for all u , so ψ satisfies (2.8). The converse follows from (2.8).

Many Markov chains such as ARCH, GARCH, and autoregressive processes are specified by structured recursions that allow quick recognition of update functions corresponding to kernels in a domain of attraction. A common example is the update function $\psi(y, (Z, W)) = Zy + W$, which behaves like $\psi'(y, Z) = Zy$ when y is large—compare ψ' to the form (2.4) discussed for tail kernels. In general, if K has an update function ψ of the form

$$\psi(y, (Z, W)) = Zy + \phi(y, W) \tag{2.9}$$

for a random variable $Z \geq 0$ and a random element W , where $t^{-1}\phi(t, w) \rightarrow 0$ whenever $w \in C$ for which $P[W \in C] = 1$, then $K \in D(G)$ with $G = P[Z \in \cdot]$. We will refer to update functions satisfying (2.9) as being in *canonical form*.

3. Finite-dimensional convergence and the extremal component

Given a Markov chain $X \sim K \in D(G)$, we show that the FDDs of X , started from an extreme state, converge to those of the tail chain T defined in (2.4).

We distinguish between two cases which represent substantially different types of behavior. If $G(\{0\}) = 0$, observe that $P[T \text{ eventually hits } \{0\}] = 0$. On the other hand, if $G(\{0\}) > 0$, T hits $\{0\}$ in finite time with probability 1. In this case, the tail chain model is appropriate only up until the first hitting time of $\{0\}$. For example, consider the trajectory of (X_1, \dots, X_m) , started

from $X_0 = t$, through the region $(t, \infty)^{m-2} \times [0, \delta] \times (t, \infty)$, where t is a high level. We would expect the tail chain to model this as a path through $(1, \infty)^{m-2} \times \{0\} \times (1, \infty)$. However, the probability of T taking paths through this set is 0, because $\{0\}$ is an absorbing state. Of course, if $G(\{0\}) = 0$, this restriction is moot.

This raises the question of how to interpret the first hitting time of $\{0\}$ for T in terms of the original Markov chain X . Such hitting times are important in the study of Markov chain point process models of exceedance clusters based on the tail chain. Intuitively, a transition to $\{0\}$ by T represents a transition from an extreme state to a nonextreme state by X . We make this notion precise in Section 3.2 by viewing such transitions as downcrossings of a certain level we term the ‘extremal boundary’.

Henceforth, X is a Markov chain on $[0, \infty)$ with transition kernel $K \in D(G)$, K^* is the tail kernel associated with G , and T is a Markov chain on $[0, \infty)$ with kernel K^* . Write $\mathbf{x}_m = (x_1, \dots, x_m)$, and similarly for X_m and T_m . The FDDs of X , conditional on $X_0 = y$, are given by

$$P_y[(X_1, \dots, X_m) \in d\mathbf{x}_m] = K(y, dx_1)K(x_1, dx_2) \cdots K(x_{m-1}, dx_m),$$

and analogously for T .

3.1. FDDs conditional on the initial state

Define the conditional distributions

$$\pi_m^{(t)}(u, \cdot) = P_{tu}\left[\frac{X_m}{t} \in \cdot\right] \quad \text{and} \quad \pi_m(u, \cdot) = P_u[T_m \in \cdot], \quad m \geq 1, \quad (3.1)$$

on $[0, \infty) \times \mathcal{B}[0, \infty]^m$. We consider when $\pi_m^{(t)} \Rightarrow \pi_m$ on $[0, \infty]^m$ pointwise in u . If $G(\{0\}) = 0$, this is a direct consequence of the domain of attraction condition (2.5), but, if $G(\{0\}) > 0$, more thought is required. We begin by restricting the convergence to the smaller space $\mathbb{E}'_m := (0, \infty]^{m-1} \times [0, \infty]$. Relatively compact sets in \mathbb{E}'_m are contained in rectangles $[a, \infty) \times [0, \infty]$, where $a \in (0, \infty)^{m-1}$.

Theorem 3.1. *Suppose that $X \sim K$ and $T \sim K^*$ are Markov chains, where $K \in D(G)$ and K^* is the tail kernel associated with G , and recall the conditional distributions $\pi_m^{(t)}$ and π_m defined in (3.1). Let $u_t = u(t)$ be a nonnegative function such that $u_t \rightarrow u > 0$ as $t \rightarrow \infty$.*

(a) *The restrictions to \mathbb{E}'_m ,*

$$\mu_m^{(t)}(u, \cdot) := \pi_m^{(t)}(u, \cdot \cap \mathbb{E}'_m) \quad \text{and} \quad \mu_m(u, \cdot) := \pi_m(u, \cdot \cap \mathbb{E}'_m), \quad (3.2)$$

satisfy

$$\mu_m^{(t)}(u_t, \cdot) \xrightarrow{v} \mu_m(u, \cdot) \quad \text{in } \mathbb{M}_+(\mathbb{E}'_m) \text{ as } t \rightarrow \infty. \quad (3.3)$$

(b) *If $G(\{0\}) = 0$, we have*

$$\pi_m^{(t)}(u_t, \cdot) \Rightarrow \pi_m(u, \cdot) \quad \text{on } [0, \infty]^m \text{ as } t \rightarrow \infty. \quad (3.4)$$

Proof. The Markov structure suggests an induction argument facilitated by Lemma A.2 in Appendix A. Consider (a) first. If $m = 1$ then (3.3) reduces to (2.7). Assume that $m \geq 2$, and let $f \in \mathcal{C}_K^+(\mathbb{E}'_m)$. Writing $\mathbb{E}'_m = (0, \infty] \times \mathbb{E}'_{m-1}$, we can find $a > 0$ and $B \in \mathcal{K}(\mathbb{E}'_{m-1})$ such

that f is supported on $[a, \infty] \times B$. Now, observe that

$$\begin{aligned} \mu_m^{(t)}(u_t, \cdot)(f) &= \int_{(0, \infty]} K(tu_t, t \, dx_1) \int_{\mathbb{E}'_{m-1}} K(tx_1, t \, dx_2) \cdots K(tx_{m-1}, t \, dx_m) f(\mathbf{x}_m) \\ &= \int_{(0, \infty]} K(tu_t, t \, dx_1) \int_{\mathbb{E}'_{m-1}} \mu_{m-1}^{(t)}(x_1, d(x_2, \dots, x_m)) f(\mathbf{x}_m). \end{aligned}$$

Defining

$$h_t(v) = \int_{\mathbb{E}'_{m-1}} \mu_{m-1}^{(t)}(v, d\mathbf{x}_{m-1}) f(v, \mathbf{x}_{m-1})$$

and

$$h(v) = \int_{\mathbb{E}'_{m-1}} \mu_{m-1}(v, d\mathbf{x}_{m-1}) f(v, \mathbf{x}_{m-1}),$$

the previous expression becomes

$$\mu_m^{(t)}(u_t, \cdot)(f) = \int_{(0, \infty]} K(tu_t, t \, dv) h_t(v).$$

Now, suppose that $v_t \rightarrow v > 0$: we verify that

$$h_t(v_t) \rightarrow h(v). \tag{3.5}$$

By continuity, we have $f(v_t, \mathbf{x}'_{m-1}) \rightarrow f(v, \mathbf{x}_{m-1})$ whenever $\mathbf{x}'_{m-1} \rightarrow \mathbf{x}_{m-1}$, and the induction hypothesis provides $\mu_{m-1}^{(t)}(v_t, \cdot) \xrightarrow{v} \mu_{m-1}(v, \cdot)$. Also, $f(x, \cdot)$ has compact support B (without loss of generality, $\mu_{m-1}(v, \partial B) = 0$). Combining these facts, (3.5) follows from Lemma A.2(b). Next, since the h_t and h have common compact support $[a, \infty]$, and recalling from Proposition 2.3 that $K(tu_t, t \cdot) \Rightarrow K^*(u, \cdot)$, Lemma A.2(a) yields

$$\mu_m^{(t)}(u_t, \cdot)(f) \rightarrow \int_{(0, \infty]} K^*(u, dv) h(v) = \mu_m(u, \cdot)(f).$$

Implication (b) follows from essentially the same argument. For $m \geq 2$, suppose that $f \in \mathcal{C}[0, \infty]^m$. Replacing μ by π and \mathbb{E}'_{m-1} by $[0, \infty]^{m-1}$ in the definitions of h_t and h , we have

$$\pi_m^{(t)}(u_t, \cdot)(f) = \int_{[0, \infty]} K(tu_t, t \, dv) h_t(v).$$

This time Lemma A.2(a) shows that $h_t(v_t) \rightarrow h(v)$ if $v_t \rightarrow v > 0$, and since $K^*(u, (0, \infty]) = 1$, resorting to Lemma A.2(a) once more yields

$$\pi_m^{(t)}(u_t, \cdot)(f) \rightarrow \int_{[0, \infty]} K^*(u, dv) h(v) = \pi_m(u, \cdot)(f).$$

This completes the proof.

If $G(\{0\}) > 0$ then $K^*(u, (0, \infty]) = 1 - G(\{0\}) < 1$, and, for (3.4) to hold, we would require knowing the behavior of $h_t(v_t)$ when $v_t \rightarrow 0$ as well. Previous work handled this using the regularity condition discussed in Section 4.

3.2. The extremal boundary

The normalization employed in the domain of attraction condition (2.5) suggests that, starting from a large state t , the extreme states are approximately scalar multiples of t . For example, we would consider a transition from t into $(t/3, 2t]$ to remain extreme. Thus, we think of states that can be made smaller than $t\delta$ for any δ , if t is large enough, as nonextreme. In this context, the set $[0, \sqrt{t}]$ would consist of nonextreme states. Asymptotically, a tail chain path through $(0, \infty)$ models the original chain X as it travels among extreme states, and all of the nonextreme states are compacted into the state $\{0\}$ in the state space of T . Since a transition of X from extreme to nonextreme is very unlikely if $G(\{0\}) = 0$, the tail chain captures all the relevant extremal behavior of X (Theorem 3.1(b)).

Drawing upon this interpretation, we develop a rigorous formulation of the distinction between extreme and nonextreme states, and we recast Theorem 3.1 as convergence on the unrestricted space $[0, \infty]^m$ of the conditional FDDs, given that X has not yet reached a nonextreme state.

Definition 3.1. Suppose that $K \in D(G)$. An *extremal boundary* for K is a nonnegative function $y(t)$ defined on $[0, \infty)$, satisfying $\lim_{t \rightarrow \infty} y(t) = 0$ and

$$K(t, t [0, y(t)]) \rightarrow G(\{0\}) \quad \text{as } t \rightarrow \infty. \tag{3.6}$$

Such a function is guaranteed to exist by Lemma A.5 in Appendix A.

If $G(\{0\}) = 0$ then $y(t) \equiv 0$ is a trivial choice. For any function $0 \leq y(t) \rightarrow 0$, we have $\limsup_{t \rightarrow \infty} K(t, t [0, y(t)]) \leq G(\{0\})$, so (3.6) is equivalent to

$$\liminf_{t \rightarrow \infty} K(t, t [0, y(t)]) \geq G(\{0\}).$$

If $y(t)$ is an extremal boundary, it follows that any function $0 \leq \tilde{y}(t) \rightarrow 0$ with $\tilde{y}(t) \geq y(t)$ for $t \geq t_0$ is also an extremal boundary for K . Taking $\tilde{y}(t) = \bigvee_{s \geq t} y(s)$ shows that, without loss of generality, we can assume $y(t)$ to be nonincreasing.

The extremal boundary has a natural formulation in terms of the update function. As in (2.9), let $\psi(y, (Z, W)) = Zy + \phi(y, W)$ be an update function in canonical form, where y is extreme. If $Z > 0$ then the next state is approximately Zy , another extreme state. Otherwise, if $Z = 0$, the next state is $\phi(y, W)$, and a transition from an extreme to a nonextreme state has taken place. This suggests choosing an extremal boundary whose order is between t and $\phi(t, w)$.

Proposition 3.1. Suppose that $\psi(y, (Z, W))$ is an update function in canonical form as in (2.9). If $\zeta(t) > 0$ is a function on $[0, \infty)$ such that

$$\phi(t, w)/\zeta(t) \rightarrow 0$$

as $t \rightarrow \infty$ whenever $w \in B$ for which $P[W \in B] = 1$, then $\liminf_{t \rightarrow \infty} K(t, [0, \zeta(t)]) \geq G(\{0\})$. Provided $\lim_{t \rightarrow \infty} \zeta(t)/t = 0$, an extremal boundary is given by $y(t) := \zeta(t)/t$.

Thus, if $\phi(t, w) = o(\zeta(t))$ and $\zeta(t) = o(t)$, then $\zeta(t)/t$ is an extremal boundary. For example, if $\psi(y, (Z, W)) = Zy + W$, so that $\phi(t, w) = w$, then choosing $\zeta(t)$ to be any function $\zeta(t) \rightarrow \infty$ such that $\zeta(t) = o(t)$ makes $\zeta(t)/t$ an extremal boundary. Choosing $\zeta(t) = \sqrt{t}$, we find that $y(t) = 1/\sqrt{t}$ is an extremal boundary.

Proof of Proposition 3.1. Since

$$\begin{aligned} P[\psi(t, (Z, W)) \leq \zeta(t), Z = 0] &= P[\phi(t, W) \leq \zeta(t), Z = 0] \\ &\geq P[|\phi(t, W)| \leq \zeta(t), Z = 0] \\ &\geq P[Z = 0] - P\left[\frac{|\phi(t, W)|}{\zeta(t)} > 1\right] \\ &\rightarrow P[Z = 0], \end{aligned}$$

we have

$$\liminf_{t \rightarrow \infty} K(t, [0, \zeta(t)]) = \liminf_{t \rightarrow \infty} P[\psi(t, (Z, W)) \leq \zeta(t)] \geq P[Z = 0].$$

We will need an extremal boundary for which (3.6) still holds upon replacing the initial state t with tu_t , where $u_t \rightarrow u > 0$. Compare the following extension with Proposition 2.3.

Proposition 3.2. *If $K \in D(G)$ then there exists an extremal boundary $y^*(t)$ such that*

$$K(tu_t, t[0, y^*(t)]) \rightarrow G(\{0\}) \quad \text{as } t \rightarrow \infty \tag{3.7}$$

for any nonnegative function $u_t = u(t) \rightarrow u > 0$.

We will refer to y^* as a *uniform extremal boundary*.

Proof of Proposition 3.2. Let $y(t)$ be an extremal boundary for K . As a first step, fix $u_0 > 1$, and suppose that $u_0^{-1} < u < u_0$. Define $\tilde{y}(t) = u_0 y(tu_0^{-1})$. Now, if $u_t \rightarrow u$ then $y_{\{u\}}(t) := u_t y(tu_t)$ satisfies (3.7), since

$$K(tu_t, t[0, y_{\{u\}}(t)]) = K(tu_t, tu_t[0, y(tu_t)]) \rightarrow G(\{0\}).$$

Here $y_{\{u\}}$ depends on the choice of function u_t . However, since we eventually have $u_0^{-1} < u_t < u_0$ for large enough t , it follows that $\tilde{y}(t) > y_{\{u\}}(t)$ for such t . Hence, $\tilde{y}(t)$ satisfies (3.7) for any $u_t \rightarrow u$ with $u_0^{-1} < u < u_0$.

Next, we remove the restriction in u_0 via a diagonalization argument. For $k = 2, 3, \dots$, let $y_k(t)$ be extremal boundaries such that $K(tu_t, t[0, y_k(t)]) \rightarrow G(\{0\})$ whenever $u_t \rightarrow u$ for $u \in (k^{-1}, k)$, and put $y_0 = y_1 = y$. Next, define the sequence $\{(s_k, x_k) : k = 0, 1, \dots\}$ inductively as follows. Setting $s_0 = 0$ and $x_0 = y_0(1)$, choose $s_k \geq s_{k-1} + 1$ such that $y_j(t) \leq k^{-1} \wedge x_{k-1}$ for all $j = 0, \dots, k$ whenever $t \geq s_k$, and put $x_k = \max\{y_j(s_k) : j = 0, \dots, k\}$. Note that $x_k \leq k^{-1} \wedge x_{k-1}$, so $x_k \downarrow 0$ and $s_k \uparrow \infty$. Finally, set

$$y^*(t) = \sum_{k=0}^{\infty} x_k \mathbf{1}_{[s_k, s_{k+1})}(t).$$

Observe that $0 \leq y^*(t) \downarrow 0$, and suppose that $u_t \rightarrow u > 0$. Then $u \in (k_0^{-1}, k_0)$ for some k_0 , so $K(tu_t, t[0, y_{k_0}(t)]) \rightarrow G(\{0\})$, and, for $k \geq k_0$, our construction ensures that, whenever $s_k \leq t < s_{k+1}$, we have $y_{k_0}(t) \leq y_{k_0}(s_k) \leq x_k = y^*(t)$. Therefore, $y^*(t) \geq y_{k_0}(t)$ for $t \geq s_{k_0}$, so y^* satisfies (3.7).

Henceforth, we assume that any $K \in D(G)$ is accompanied by a uniform extremal boundary denoted by $y(t)$, and we consider extreme states on the order of t to be $(ty(t), \infty]$. If $G(\{0\}) = 0$ then all positive states are extreme states. We now use the extremal boundary to reformulate the

convergence of Theorem 3.1 on the larger space $[0, \infty]^m$. Put $\mathbb{E}'_m(t) = (y(t), \infty]^{m-1} \times [0, \infty]$, so that $\mathbb{E}'_m(t) \uparrow \mathbb{E}'_m = (0, \infty]^{m-1} \times [0, \infty]$. Recall the notation $\mu_m^{(t)}$ and μ_m^* from (3.1) and (3.2).

Theorem 3.2. *Let $u_t = u(t)$ be a nonnegative function such that $u_t \rightarrow u > 0$ as $t \rightarrow \infty$. Taking*

$$\tilde{\mu}_m^{(t)}(u, \cdot) = \pi_m^{(t)}(u, \cdot \cap \mathbb{E}'_m(t)),$$

we have

$$\tilde{\mu}_m^{(t)}(u_t, \cdot) \xrightarrow{v} \mu_m(u, \cdot) \text{ in } \mathbb{M}_+[0, \infty]^m \text{ as } t \rightarrow \infty.$$

Proof. Note that we can just as well write $\tilde{\mu}_m^{(t)}(u, \cdot) = \mu_m^{(t)}(u, \cdot \cap \mathbb{E}'_m(t))$. Suppose that $m \geq 2$, and let $f \in \mathcal{C}_K^+[0, \infty]^m$. For $\delta > 0$, define $A_\delta = (\delta, \infty]^{m-1} \times [0, \infty]$, and choose δ such that $\mu_m(u, \partial A_\delta) = 0$. On the one hand, for large t , we have

$$\begin{aligned} \tilde{\mu}_m^{(t)}(u_t, \cdot)(f) &= \int_{[0, \infty]^m} f(\mathbf{x}) \mathbf{1}_{\mathbb{E}'_m(t)}(\mathbf{x}) \mu_m^{(t)}(u_t, d\mathbf{x}) \\ &\geq \int_{\mathbb{E}'_m} f(\mathbf{x}) \mathbf{1}_{A_\delta}(\mathbf{x}) \mu_m^{(t)}(u_t, d\mathbf{x}) \\ &\rightarrow \int_{\mathbb{E}'_m} f(\mathbf{x}) \mathbf{1}_{A_\delta}(\mathbf{x}) \mu_m(u, d\mathbf{x}) \end{aligned}$$

as $t \rightarrow \infty$ by Lemma A.3 in Appendix A. Letting $\delta \downarrow 0$ yields

$$\liminf_{t \rightarrow \infty} \tilde{\mu}_m^{(t)}(u_t, \cdot)(f) \geq \mu_m(u, \cdot)(f) \tag{3.8}$$

by monotone convergence. On the other hand, fixing δ , we can decompose the space according to the first downcrossing of δ , i.e.

$$\tilde{\mu}_m^{(t)}(u_t, \cdot)(f) = \int_{[0, \infty]^m} f(\mathbf{x}) \mathbf{1}_{A_\delta}(\mathbf{x}) \tilde{\mu}_m^{(t)}(u_t, d\mathbf{x}) + \sum_{k=1}^{m-1} \int_{[0, \infty]^m} f(\mathbf{x}) \mathbf{1}_{A_\delta^k}(\mathbf{x}) \tilde{\mu}_m^{(t)}(u_t, d\mathbf{x}), \tag{3.9}$$

where $A_\delta^k = (\delta, \infty]^{k-1} \times [0, \delta] \times [0, \infty]^{m-k}$. On the subsets A_δ^k we appeal to the bound on f , say M , to obtain

$$\int_{[0, \infty]^m} f(\mathbf{x}) \mathbf{1}_{A_\delta^k}(\mathbf{x}) \tilde{\mu}_m^{(t)}(u_t, d\mathbf{x}) \leq M \tilde{\mu}_m^{(t)}(u_t, A_\delta^k).$$

Now,

$$\begin{aligned} \tilde{\mu}_m^{(t)}(u_t, A_\delta^k) &\leq \mu_k^{(t)}(u_t, (\delta, \infty]^{k-1} \times (y(t), \delta]) \\ &= \mu_k^{(t)}(u_t, (\delta, \infty]^{k-1} \times [0, \delta]) - \mu_k^{(t)}(u_t, (\delta, \infty]^{k-1} \times [0, y(t)]). \end{aligned} \tag{3.10}$$

Considering the second term, we have

$$\begin{aligned} &\mu_k^{(t)}(u_t, (\delta, \infty]^{k-1} \times [0, y(t)]) \\ &= \int_{[0, \infty]} K(tu_t, t dx_1) \mathbf{1}_{(\delta, \infty]}(x_1) \times \cdots \\ &\quad \times \int_{[0, \infty]} K(tx_{k-2}, t dx_{k-1}) \mathbf{1}_{(\delta, \infty]}(x_{k-1}) K(tx_{k-1}, t[0, y(t)]) \\ &= \int_{\mathbb{E}'_{k-1}} \mu_{k-1}^{(t)}(u_t, d\mathbf{x}_{k-1}) h_t(\mathbf{x}_{k-1}), \end{aligned}$$

where

$$h_t(\mathbf{x}_{k-1}) = K(t\mathbf{x}_{k-1}, t[0, y(t)]) \mathbf{1}_{(\delta, \infty)^{k-1}}(\mathbf{x}_{k-1}).$$

Moreover, if $\mathbf{x}_{k-1}^t \rightarrow \mathbf{x}_{k-1} \in (\delta, \infty)^{k-1}$ then

$$h_t(\mathbf{x}_{k-1}^t) = K(t\mathbf{x}_{k-1}^t, t[0, y(t)]) \mathbf{1}_{(\delta, \infty)^{k-1}}(\mathbf{x}_{k-1}^t) \rightarrow G(\{0\}) \mathbf{1}_{(\delta, \infty)^{k-1}}(\mathbf{x}_{k-1}),$$

using the fact that $y(t)$ is a uniform extremal boundary. Since $\mu_{k-1}(u, \partial(\delta, \infty)^{k-1}) = 0$, without loss of generality by the choice of δ , we conclude that

$$\mu_k^{(t)}(u_t, (\delta, \infty)^{k-1} \times [0, y(t)]) \rightarrow G(\{0\})\mu_{k-1}(u, (\delta, \infty)^{k-1}) = \mu_k(u, (\delta, \infty)^{k-1} \times \{0\})$$

as $t \rightarrow \infty$. Now, let us return to (3.10). Given any $\varepsilon > 0$, by choosing small enough δ , we can make

$$\begin{aligned} &\mu_k^{(t)}(u_t, (\delta, \infty)^{k-1} \times (y(t), \delta]) \\ &\rightarrow \mu_k(u, (\delta, \infty)^{k-1} \times [0, \delta]) - \mu_k(u, (\delta, \infty)^{k-1} \times \{0\}) \\ &\leq \mu_k(u, (0, \infty)^{k-1} \times [0, \delta]) - \mu_k(u, (\delta, \infty)^{k-1} \times \{0\}) \\ &< \mu_k(u, (0, \infty)^{k-1} \times \{0\}) + \frac{\varepsilon}{2} - \left(\mu_k(u, (0, \infty)^{k-1} \times \{0\}) - \frac{\varepsilon}{2} \right) \\ &= \varepsilon, \end{aligned}$$

i.e.

$$\limsup_{t \rightarrow \infty} \tilde{\mu}_m^{(t)}(u_t, A_\delta^k) < \varepsilon$$

for $k = 1, \dots, m - 1$. Therefore, (3.9) implies that, given $\varepsilon' > 0$,

$$\begin{aligned} \limsup_{t \rightarrow \infty} \tilde{\mu}_m^{(t)}(u_t, \cdot)(f) &\leq \int_{[0, \infty]^m} f(\mathbf{x}) \mathbf{1}_{A_\delta}(\mathbf{x}) \mu_m(u, d\mathbf{x}) \\ &\quad + M \sum_{k=1}^{m-1} \limsup_{t \rightarrow \infty} \tilde{\mu}_m^{(t)}(u_t, A_\delta^k) \\ &< \mu_m(u, \cdot)(f) + \varepsilon' \end{aligned}$$

for small enough δ . Combining this with (3.8) yields the result.

3.3. The extremal component

Having thus formalized the distinction between extreme and nonextreme states, we return to the question of phrasing a general FDD limit result for X . The extremal boundary allows us to interpret the first hitting time of $\{0\}$ by the tail chain as approximating the time of the first transition from extreme down to nonextreme. In this terminology, Theorem 3.2 provides a result, given that such a transition has yet to occur.

Define

$$\tau(t) = \inf\{n \geq 0: X_n \leq ty(t)\}$$

to be the first hitting time of a nonextreme state. For a Markov chain started from tu_t , where $u_t \rightarrow u > 0$, we have $tu_t > ty(t)$ for large t , so $\tau(t)$ is the first downcrossing of the extremal boundary.

For the tail chain $T = \{T_0\xi_1 \cdots \xi_n : n = 0, 1, \dots\}$, put $\tau^* = \inf\{n \geq 0 : T_n = 0\}$. Given $T_0 > 0$, $\tau^* = \inf\{n \geq 1 : \xi_n = 0\}$, i.e. τ^* follows a geometric distribution with parameter $p = G(\{0\})$. Thus, $P[\tau^* = m] = p(1 - p)^{m-1}$ for $m \geq 1$ if $p > 0$, and $P[\tau^* = \infty] = 1$ if $p = 0$. Theorem 3.2 becomes

$$P_{tu_t}[t^{-1}X_m \in \cdot, \tau(t) \geq m] \xrightarrow{v} P_u[T_m \in \cdot, \tau^* \geq m], \tag{3.11}$$

implying that τ^* approximates $\tau(t)$:

$$P_{tu_t}[\tau(t) \in \cdot] \Rightarrow P[\tau^* \in \cdot] \quad \text{as } t \rightarrow \infty \text{ and } u_t \rightarrow u > 0. \tag{3.12}$$

So, if $G(\{0\}) > 0$, X takes an average of approximately $G(\{0\})^{-1}$ steps to return to a nonextreme state. However, if $G(\{0\}) = 0$, $P_{tu_t}[\tau_1 \leq m] \rightarrow 0$ for any $m \geq 1$; in other words, starting from a larger and larger initial state, it will take longer and longer for X to downcross to a nonextreme state.

We now restate (3.11) in terms of a process derived from X , called the *extremal component* of X , whose unrestricted FDDs converge weakly to those of T .

Definition 3.2. The *extremal component* of X relative to t is the process $X^{(t)}$ defined for $t > 0$ as

$$X_n^{(t)} = X_n \mathbf{1}_{(n < \tau(t))}, \quad n = 0, 1, \dots$$

Observe that $X^{(t)}$ is a Markov chain on $[0, \infty)$ with transition kernel

$$K^{(t)}(x, A) = \begin{cases} K(x, A \cap (ty(t), \infty]) + \varepsilon_0(A)K(x, [0, ty(t)]), & x > ty(t), \\ \varepsilon_0(A), & x \leq ty(t). \end{cases}$$

It follows that $K^{(t)}(t, \cdot) \Rightarrow G$ as $t \rightarrow \infty$, and, additionally, that $K^{(t)}(t, \{0\}) \rightarrow G(\{0\})$. The relation between $X^{(t)}$ and X is

$$P_{tu_t}[t^{-1}X_m^{(t)} \in \cdot \mid \tau(t) > m] = P_{tu_t}[t^{-1}X_m \in \cdot \mid \tau(t) > m].$$

Theorem 3.3. Let $u_t = u(t) \geq 0$ satisfy $u_t \rightarrow u > 0$ as $t \rightarrow \infty$. Then, on $[0, \infty)^m$,

$$\tilde{\pi}_m^{(t)}(u_t, \cdot) := P_{tu_t} \left[\left(\frac{X_1^{(t)}}{t}, \dots, \frac{X_m^{(t)}}{t} \right) \in \cdot \right] \Rightarrow P_u[(T_1, \dots, T_m) \in \cdot] \quad \text{as } t \rightarrow \infty.$$

Proof. Suppose that $m \geq 2$ and $f \in \mathcal{C}[0, \infty)^m$. Without loss of generality, $f \in \mathcal{C}_K^+[0, \infty)^m$ as well, since the space is compact. Recall the notation introduced in Theorem 3.2. Conditioning on $\tau(t)$, we can write

$$\begin{aligned} \tilde{\pi}_m^{(t)}(u_t, \cdot)(f) &= \int_{(0, \infty]^m} f(\mathbf{x}_m) \tilde{\pi}_m^{(t)}(u_t, d\mathbf{x}_m) + \sum_{k=1}^m \int_{(0, \infty]^{k-1} \times \{0\}^{m-k+1}} f(\mathbf{x}_m) \tilde{\pi}_m^{(t)}(u_t, d\mathbf{x}_m) \\ &= \int_{(0, \infty]^m} f(\mathbf{x}_m) \tilde{\pi}_m^{(t)}(u_t, d\mathbf{x}_m) \\ &\quad + \sum_{k=1}^m \int_{(0, \infty]^{k-1} \times \{0\}} f(\mathbf{x}_k, 0, \dots, 0) \tilde{\pi}_k^{(t)}(u_t, d\mathbf{x}_k) \end{aligned}$$

by the Markov property. Since

$$\begin{aligned} \tilde{\pi}_m^{(t)}(u_t, \cdot \cap (0, \infty]^m) &= P_{tu_t}[t^{-1} \mathbf{X}_m^{(t)} \in \cdot, \tau(t) > m] \\ &= P_{tu_t}[t^{-1} \mathbf{X}_m \in \cdot \cap (y(t), \infty]^m] \\ &= \tilde{\mu}_{m+1}^{(t)}(u_t, \cdot \times [0, \infty]), \end{aligned}$$

the first term becomes

$$\begin{aligned} \tilde{\mu}_{m+1}^{(t)}(u_t, \cdot \times [0, \infty])(f) &\rightarrow \mu_{m+1}(u, \cdot \times [0, \infty])(f) \\ &= \int_{(0, \infty]^m} f(\mathbf{x}_m) \pi_m(u, d\mathbf{x}_m) \\ &= \int_{(0, \infty]^m} f(\mathbf{x}_m) P_u[\mathbf{T}_m \in d\mathbf{x}_m] \end{aligned}$$

as $t \rightarrow \infty$. Next, for any measurable $A \subset [0, \infty]^k$, write $A_0 = \{\mathbf{x}_{k-1} : (\mathbf{x}_{k-1}, 0) \in A\} \subset [0, \infty]^{k-1}$, and observe that

$$\begin{aligned} \tilde{\pi}_k^{(t)}(u_t, A \cap (0, \infty]^{k-1} \times \{0\}) &= P_{tu_t}[t^{-1} \mathbf{X}_{k-1}^{(t)} \in A_0 \cap (0, \infty]^{k-1}, X_k^{(t)} = 0] \\ &= P_{tu_t}[t^{-1} \mathbf{X}_{k-1} \in A_0 \cap (y(t), \infty]^{k-1}, t^{-1} X_k \leq y(t)] \\ &= \tilde{\mu}_k^{(t)}(u_t, A_0 \times [0, \infty]) - \tilde{\mu}_{k+1}^{(t)}(u_t, A_0 \times [0, \infty]^2). \end{aligned}$$

Applying this reasoning to the terms in the summation yields

$$\begin{aligned} &\int_{[0, \infty]^k} f(\mathbf{x}_{k-1}, 0, \dots, 0) \tilde{\mu}_k^{(t)}(u_t, d\mathbf{x}_k) - \int_{[0, \infty]^{k+1}} f(\mathbf{x}_{k-1}, 0, \dots, 0) \tilde{\mu}_{k+1}^{(t)}(u_t, d\mathbf{x}_{k+1}) \\ &\rightarrow \int_{[0, \infty]^k} f(\mathbf{x}_{k-1}, 0, \dots, 0) \mu_k(u, d\mathbf{x}_k) \\ &\quad - \int_{[0, \infty]^{k+1}} f(\mathbf{x}_{k-1}, 0, \dots, 0) \mu_{k+1}(u, d\mathbf{x}_{k+1}) \\ &= \int_{(0, \infty]^{k-1} \times \{0\}} f(\mathbf{x}_k, 0, \dots, 0) \pi_k(u, d\mathbf{x}_k) \\ &= \int_{(0, \infty]^{k-1} \times \{0\}^{m-k+1}} f(\mathbf{x}_m) P_u[\mathbf{T}_m \in d\mathbf{x}_m]. \end{aligned}$$

Combining these limits shows that $E_{tu_t} f(t^{-1} \mathbf{X}_m^{(t)}) \rightarrow E_u f(\mathbf{T}_m)$ as $t \rightarrow \infty$.

4. The regularity condition

Previous work on the tail chain has addressed FDD convergence of \mathbf{X} to \mathbf{T} under a single assumption analogous to our domain of attraction condition (2.5). As observed in Section 3.1, when $G(\{0\}) = 0$, FDD convergence of $\{t^{-1} \mathbf{X}\}$ follows directly, but, when $G(\{0\}) > 0$, it was common to assume an additional technical condition that made (2.5) imply FDD convergence to \mathbf{T} as well. This condition, which we refer to as the ‘regularity condition’, controls \mathbf{X} upon leaving a nonextreme state. We consider equivalences between different forms appearing in the literature, in terms of both kernels and update functions, and show that, under the regularity condition, the extremal behavior of \mathbf{X} is asymptotically the same as that of its extremal component $\mathbf{X}^{(t)}$.

In the cases where $G(\{0\}) > 0$, Perfekt [17], [18] required that

$$\lim_{\delta \downarrow 0} \limsup_{t \rightarrow \infty} \sup_{u \in [0, \delta]} K(tu, (t, \infty]) = 0, \tag{4.1}$$

while Segers [22] stipulated that the chosen update function corresponding to K must be of at most linear order in the initial state:

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq y \leq t} t^{-1} \psi(y, v) < \infty, \quad v \in B_0, P[V \in B_0] = 1. \tag{4.2}$$

Smith [23] used a variant of (4.1). We deem a formulation in terms of distributional convergence to be instructive in our context.

Definition 4.1. A Markov transition kernel $K \in D(G)$ satisfies the *regularity condition* if

$$K(tu_t, t \cdot) \Rightarrow \varepsilon_0(\cdot) \tag{4.3}$$

on $[0, \infty]$ as $t \rightarrow \infty$ for any nonnegative function $u_t = u(t) \rightarrow 0$.

Thus, the regularity condition complements the domain of attraction condition expressed in the form (2.7), with the effect that the extremal behavior of K is completely described by the tail kernel K^* .

We now consider the relationships between (4.1), (4.2), and (4.3), and propose an intuitive equivalent for update functions in canonical form.

Proposition 4.1. *Suppose that $K \in D(G)$, and let $\psi(\cdot, V)$ be an update function corresponding to K such that*

$$t^{-1} \psi(t, v) \rightarrow \xi(v) \tag{4.4}$$

whenever $v \in B$ for which $P[V \in B] = 1$ and $\xi \circ V \sim G$. Then the following statements hold.

- (a) *Condition (4.1) is necessary and sufficient for K to satisfy the regularity condition (4.3).*
- (b) *Condition (4.2) is sufficient for K to satisfy the regularity condition (4.3).*
- (c) *If ψ is in canonical form, i.e.*

$$\psi(y, (Z, W)) = Zy + \phi(y, W),$$

then ψ satisfies (4.2) if and only if $\phi(\cdot, w)$ is bounded on any neighborhood of 0 for each $w \in C$, a set for which $P[W \in C] = 1$.

Proof. (a) Assume that (4.1) holds, and suppose that $u_t \rightarrow 0$. We show that $K(tu_t, t(x, \infty]) \rightarrow 0$ for any $x > 0$. Write

$$\omega(t, \delta) = \sup_{u \in [0, \delta]} K(tu, (t, \infty]).$$

Let $\varepsilon > 0$ be given, and choose δ small enough that $\limsup_{t \rightarrow \infty} \omega(t, \delta) < \varepsilon/2$. Then, for t large enough that both $u_t < \delta x$ and $\omega(tx, \delta) - \limsup_{t \rightarrow \infty} \omega(t, \delta) < \varepsilon/2$, we have

$$K(tu_t, t(x, \infty]) \leq \sup_{u \in [0, \delta x]} K(tu, t(x, \infty]) = \omega(tx, \delta) < \limsup_{t \rightarrow \infty} \omega(t, \delta) + \frac{\varepsilon}{2}.$$

Our choice of δ implies that $K(tu_t, t(x, \infty]) < \varepsilon$.

Conversely, assume that K satisfies (4.3) but that (4.1) fails. Choose $\varepsilon > 0$ and a sequence $\delta_n \downarrow 0$ such that $\limsup_{t \rightarrow \infty} \omega(t, \delta_n) \geq \varepsilon$ for $n = 1, 2, \dots$. Then, for each n , we can find a sequence $t_k^n \rightarrow \infty$ as $k \rightarrow \infty$ such that $\omega(t_k^n, \delta_n) \geq \varepsilon$ for each k . Diagonalize to find $k_1 < k_2 < \dots$ such that $s_n = t_{k_n}^n \rightarrow \infty$ and $\omega(s_n, \delta_n) \geq \varepsilon$ for all n . Finally, for $n = 1, 2, \dots$, choose $u_n \in [0, \delta_n]$ such that

$$K(s_n u_n, (s_n, \infty]) > \omega(s_n, \delta_n) - \frac{\varepsilon}{2},$$

and put $u(t) = \sum_n u_n \mathbf{1}_{[s_n, s_{n+1})}(t)$. Clearly, $u(t) \rightarrow 0$, but $K(s_n u(s_n), (s_n, \infty]) \geq \varepsilon/2$ for all n , contradicting (4.3).

(b) Write $M(v) = \limsup_{t \rightarrow \infty} \sup_{0 \leq y \leq t} t^{-1} \psi(y, v)$. Since

$$\sup_{0 \leq y \leq t} t^{-1} \psi(y, v) = \sup_{0 \leq y \leq \delta} \frac{\psi(t\delta^{-1}y, v)}{t\delta^{-1}} \delta^{-1}$$

for $\delta > 0$, we have

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq y \leq \delta} t^{-1} \psi(ty, v) = \delta M(v).$$

Now, suppose that $u_t \rightarrow 0$. Given any $\delta > 0$ we have

$$t^{-1} \psi(tu_t, v) \leq \sup_{0 \leq y \leq \delta} t^{-1} \psi(ty, v),$$

provided t is large enough, so $\limsup_{t \rightarrow \infty} t^{-1} \psi(tu_t, v) \leq \delta M(v)$. Consequently,

$$\limsup_{t \rightarrow \infty} t^{-1} \psi(tu_t, v) = 0$$

for every v such that $M(v) < \infty$. Under (4.2), this means that $\mathbb{P}[t^{-1} \psi(tu_t, V) \rightarrow 0] = 1$, implying (4.3).

(c) Suppose first that $\chi_w(a) = \sup_{0 \leq y \leq a} \phi(y, w) < \infty$ for all $a > 0$, whenever $w \in C$. Fixing $w \in C$ and $z \geq 0$, note that

$$\sup_{0 \leq y \leq t} t^{-1} \psi(y, (z, w)) \leq z + \sup_{0 \leq y \leq t} t^{-1} \phi(y, w),$$

and observe for any $a > 0$ that

$$\begin{aligned} \sup_{0 \leq y \leq t} t^{-1} \phi(y, w) &\leq \left(\sup_{0 \leq y \leq a} t^{-1} \phi(y, w) \right) \vee \left(\sup_{a \leq y \leq t} y^{-1} \phi(y, w) \right) \\ &\leq t^{-1} \chi_w(a) \vee \left(\sup_{a \leq y} y^{-1} \phi(y, w) \right). \end{aligned}$$

Choosing a large enough that $\sup_{a \leq y} y^{-1} \phi(y, w) \leq 1$, say, it follows that

$$\limsup_{t \rightarrow \infty} \sup_{0 \leq y \leq t} t^{-1} \psi(y, (z, w)) \leq z + 1,$$

so $v = (z, w) \in B_0$. Therefore, $\mathbb{P}[(Z, W) \in B_0] \geq \mathbb{P}[Z \geq 0, W \in C] = 1$.

Conversely, suppose that there is a set D with $\mathbb{P}[W \in D] > 0$ such that $w \in D$ implies that $\chi_w(a) = \infty$ for some $0 < a < \infty$. Since $\sup_{0 \leq y \leq t} t^{-1} \psi(y, (z, w)) \geq t^{-1} \chi_w(t)$, we have $[0, \infty) \times D \subset B_0^c$, contradicting (4.2).

The exclusion of necessity from part (b) results from the fact that a kernel K does not uniquely specify an update function ψ . Even when K satisfies the regularity condition (4.3), it may be possible to choose a nasty update function ψ which satisfies (4.4), but not (4.2). However, in such cases there may exist a different update function ψ' corresponding to K which does satisfy (4.2).

We now give an example of such a situation. We exhibit an update function ψ for which (i) (4.4) holds; (ii) (4.2) fails because condition (c) of Proposition 4.1 fails; but yet (iii) the corresponding kernel satisfies the regularity condition (4.3). Furthermore, we present a different choice of update function corresponding to the same kernel that satisfies (4.2). Define $\psi(y, V = (Z, W)) = Zy + \phi(y, W)$, where

$$\phi(y, w) = \sum_{k=1}^{\infty} k \mathbf{1}_{(yw=1/k)}$$

and $W \sim U(0, 1)$. (i) Since $\phi(t, w) = 0$ for $t > 1/w$, it is clear that ψ satisfies (4.4) with $\xi = Z$. (ii) Observe that, for any $w \in (0, 1)$, $\phi(\cdot, w)$ is unbounded on the interval $[0, 1]$. Therefore, by part (c) of Proposition 4.1, (4.2) cannot hold for ψ . (iii) However, the corresponding kernel does satisfy the regularity condition (4.3). Suppose that $u_t \rightarrow 0$ and $a > 0$ is arbitrarily large. Write

$$\begin{aligned} P[t^{-1}\psi(tu_t, (Z, W)) > x] &= P[Zu_t + t^{-1}\phi(tu_t, W) > x] \\ &\leq P[t^{-1}\phi(tu_t, W) > x'] + P[Z > a], \end{aligned}$$

choosing $0 < x' < x - au_t$. Since, for any t , $\{w : \phi(tu_t, w) > tx'\} \subset \{(tu_t k)^{-1} : k = 1, 2, \dots\}$, a set of measure 0 with respect to $P[W \in \cdot]$, (4.3) follows by letting $a \rightarrow \infty$. On the other hand, the update function $\psi'(y, Z) = Zy$ does satisfy (4.2), and, for any y ,

$$P[\psi'(y, Z) \neq \psi(y, (Z, W))] = P[W \in \{(yk)^{-1} : k = 1, 2, \dots\}] = 0,$$

so ψ' does indeed correspond to K .

The regularity condition (4.3) restricts attention to Markov chains for which the probability of returning to an extreme state in the next m steps after falling below the extremal boundary is asymptotically negligible. For such chains, as well as those for which $y(t) \equiv 0$ is an extremal boundary for K , X has the same asymptotic behavior as its extremal component, as described next.

Theorem 4.1. *Suppose that $X \sim K$ with $K \in D(G)$, and let ρ be a metric on \mathbb{R}^m . If $y(t) \equiv 0$ is an extremal boundary for K , or if K satisfies the regularity condition (4.3), then, for any $\varepsilon > 0$, we have*

$$P_{tu_t} \left[\rho \left(\frac{X_m^{(t)}}{t}, \frac{X_m}{t} \right) > \varepsilon \right] \rightarrow 0 \quad \text{as } t \rightarrow \infty \text{ and } u_t \rightarrow u > 0. \tag{4.5}$$

Consequently,

$$P_{tu_t} \left[\left(\frac{X_1}{t}, \dots, \frac{X_m}{t} \right) \in \cdot \right] \Rightarrow P_u [(T_1, \dots, T_m) \in \cdot] \quad \text{as } t \rightarrow \infty \text{ and } u_t \rightarrow u > 0. \tag{4.6}$$

First let us extend the regularity condition to higher-order transition kernels.

Lemma 4.1. *If K satisfies (4.3) then so do the m -step transition kernels K^m .*

Proof. This is established by induction. Let $u_t \rightarrow 0$ and $f \in \mathcal{C}[0, \infty]$. For $m \geq 2$, we have

$$K^m(tu_t, \cdot)(f) = \int_{[0, \infty]} K^{m-1}(tu_t, t \, dv) \int_{[0, \infty]} K(tv, t \, dx) f(x).$$

Assume that $K^{m-1}(tu_t, \cdot) \Rightarrow \varepsilon_0$; (4.3) implies that $\int K(tv, t \, dx) f(x) \rightarrow f(0)$ whenever $v_t \rightarrow 0$. Therefore, by Lemma A.2(a), we conclude that

$$K^m(tu_t, \cdot)(f) \rightarrow f(0) = \varepsilon_0(f).$$

Proof of Theorem 4.1. Suppose that $\varepsilon > 0$ and $u_t \rightarrow u > 0$. Write

$$P_{tu_t}[\rho(t^{-1} X_m^{(t)}, t^{-1} X_m) > \varepsilon] = \sum_{k=1}^m P_{tu_t}[\rho(t^{-1} X_m^{(t)}, t^{-1} X_m) > \varepsilon, \tau(t) = k].$$

Since $X_j = X_j^{(t)}$ while $j < \tau(t)$, for the k th summand to converge to 0, it is sufficient that

$$P_{tu_t} \left[\left| \frac{X_j^{(t)}}{t} - \frac{X_j}{t} \right| > \delta, \tau(t) = k \right] = P_{tu_t} \left[\frac{X_j}{t} > \delta, \tau(t) = k \right] \rightarrow 0$$

for $j = k, \dots, m$ and any $\delta > 0$. If $j = k$, we have

$$P_{tu_t} \left[\frac{X_j}{t} > \delta, \tau(t) = k \right] \leq P_{tu_t} \left[\frac{X_k}{t} > \delta, \frac{X_k}{t} \leq y(t) \right] = 0$$

for large t . For $j > k$, recalling the notation of Theorem 3.2,

$$\begin{aligned} P_{tu_t} \left[\frac{X_j}{t} > \delta, \tau(t) = k \right] &= \int_{\mathbb{E}_k^{(t)}} \mathbf{1}_{[0, y(t)]}(x_k) P_{tu_t} \left[\frac{X_j}{t} > \delta \mid \frac{X_k}{t} = x_k \right] P_{tu_t} \left[\frac{X_k}{t} \in dx_k \right] \\ &= \int_{[0, \infty]^k} P_{tx_k} [X_{j-k} > t\delta] \mathbf{1}_{[0, y(t)]}(x_k) \tilde{\mu}_k^{(t)}(u_t, dx_k) \end{aligned}$$

using the Markov property. We claim that this integral tends to 0 as $t \rightarrow \infty$. If $y(t) \equiv 0$, this follows directly. Otherwise, recall that $\tilde{\mu}_k^{(t)}(u_t, \cdot) \xrightarrow{v} \mu_k(u, \cdot)$, and consider $h_t(x_k) = P_{tx_k} [X_{j-k} > t\delta] \mathbf{1}_{[0, y(t)]}(x_k)$. Suppose that $x^{(t)} \rightarrow x \in [0, \infty]^k$. If $x_k > 0$ then $h_t(x^{(t)}) = 0$ for large t because $y(t) \rightarrow 0$. Otherwise, if $x_k = 0$, we have $h_t(x^{(t)}) \rightarrow 0$ since Lemma 4.1 implies that $P_{tx_k} [X_{j-k} > t\delta] \rightarrow 0$ as $t \rightarrow \infty$. Lemma A.2(b) establishes (4.5); (4.6) follows by Slutsky’s theorem applied to the result of Theorem 3.3.

Therefore, X converges to T in FDDs under (a) $G(\{0\}) = 0$, (b) $G(\{0\}) > 0$ combined with (4.3), or (c) $G(\{0\}) > 0$ combined with the extremal boundary $y(t) \equiv 0$. In any of these cases, we will be able to replace the extremal component $X^{(t)}$ with the complete chain X in the results given below in Sections 5.1 and 5.2. However, that $y(t) \equiv 0$ is an extremal boundary, and consequently that (4.6) holds, does not imply that the regularity condition holds, regardless of $G(\{0\})$; in particular, a kernel for which $G(\{0\}) = 0$ need not satisfy (4.3). This is illustrated in Example 6.3 below.

5. Convergence of the unconditional FDDs

5.1. Effect of a regularly varying initial distribution

So far our convergence results required that the initial state become large, and the only distributional assumption was that the transition kernel K determining X be attracted to some distribution G . To obtain a result for the unconditional distribution of (X_0, \dots, X_m) , we require an additional assumption about how likely the initial observation X_0 is to be large. Using Lemma A.4 in Appendix A, the results of the previous sections extend to multivariate regular variation on the cone $\mathbb{E}_m = (0, \infty] \times [0, \infty]^m$ when the distribution of X_0 has a regularly varying tail. This cone is smaller than the cone $[0, \infty]^{m+1} \setminus \{0\}$ traditionally employed in extreme value theory because the kernel domain of attraction condition (2.5) is uninformative when the initial state is not extreme. This is analogous to the setting of the conditional extreme value model considered in [8] and [12].

Proposition 5.1. *Assume that $X \sim K$ with $K \in D(G)$, and that $X_0 \sim F$, where F is a distribution on $[0, \infty)$ with a regularly varying tail (denoted by $1 - F \in \text{RV}_{-\alpha}$). This means that, as $t \rightarrow \infty$, for some scaling function $b(t) \rightarrow \infty$,*

$$tF(b(t)\cdot) \xrightarrow{v} \nu_\alpha(\cdot) \text{ in } \mathbb{M}_+(0, \infty],$$

where $\nu_\alpha(x, \infty] = x^{-\alpha}$ and $\alpha > 0$. Define the measure ν^* on $\mathbb{E}_m = (0, \infty] \times [0, \infty]^m$ by

$$\nu^*(dx_0, d\mathbf{x}_m) = \nu_\alpha(dx_0) P_{X_0}[(T_1, \dots, T_m) \in d\mathbf{x}_m].$$

Then, for $m = 1, 2, \dots$, the following convergences take place as $t \rightarrow \infty$.

(a) In $\mathbb{M}_+((0, \infty]^m \times [0, \infty])$,

$$tP[b(t)^{-1}(X_0, X_1, \dots, X_m) \in \cdot \cap (0, \infty]^m \times [0, \infty]] \xrightarrow{v} \nu^*(\cdot \cap (0, \infty]^m \times [0, \infty]).$$

(b) In $\mathbb{M}_+(\mathbb{E}_m)$,

$$tP[b(t)^{-1}(X_0^{(b(t))}, X_1^{(b(t))}, \dots, X_m^{(b(t))}) \in \cdot] \xrightarrow{v} \nu^*(\cdot).$$

(c) If either $G(\{0\}) = 0$, $y(t) \equiv 0$ is an extremal boundary, or K satisfies the regularity condition (4.3), then, in $\mathbb{M}_+(\mathbb{E}_m)$,

$$tP[b(t)^{-1}(X_0, X_1, \dots, X_m) \in \cdot] \xrightarrow{v} \nu^*(\cdot).$$

(d) In $\mathbb{M}_+(0, \infty]$,

$$tP\left[\frac{X_0}{b(t)} \in dx_0, \tau(b(t)) \geq m\right] \xrightarrow{v} (1 - G(\{0\}))^{m-1} \nu_\alpha(dx_0).$$

Remark. The convergence statements in Proposition 5.1 may be reformulated equivalently as, say,

$$P[b(t)^{-1}(X_0, X_1, \dots, X_m) \in \cdot \mid X_0 > b(t)] \Rightarrow P[(T_0, T_1, \dots, T_m) \in \cdot],$$

where $T_0 \sim \text{Pareto}(\alpha)$. This is the form considered by Segers [22].

Proof of Proposition 5.1. Apply Lemma A.4 in Appendix A to the results of Theorems 3.1, 3.3, and 4.1, and (3.12).

In the case $m = 1$, \mathbb{E}_1 is a rotated version of \mathbb{E}_Π used in the conditional extreme value model in [8] and [9], and the limit can be expressed as

$$v^*((x_0, \infty) \times [0, x_1]) = \int_{x_0}^\infty v_\alpha(du) P\left[\xi \leq \frac{x_1}{u}\right] = x_0^{-\alpha} P\left[\xi \leq \frac{x_1}{x_0}\right] - x_1^{-\alpha} E \xi^\alpha \mathbf{1}_{\{\xi \leq x_1/x_0\}}$$

for $(x_0, x_1) \in (0, \infty) \times [0, \infty]$, where $\xi \sim G$ (with $E \xi^\alpha \leq \infty$). Since

$$v^*((x_0, \infty) \times \{0\}) = x_0^{-\alpha} P[\xi = 0] \quad \text{and} \quad v^*((0, \infty) \times (x_1, \infty]) = x_1^{-\alpha} E \xi^\alpha,$$

sets on the x_0 -axis incur mass proportional to $G(\{0\})$, and sets bounded away from this axis are weighted according to $E \xi^\alpha$. A consequence of the second observation is that

$$\liminf_{t \rightarrow \infty} t P\left[\frac{X_1}{b(t)} > x\right] \geq E \xi^\alpha \cdot x^{-\alpha}.$$

Thus, knowledge concerning the tail behavior of X_1 imposes a restriction on the distributions G to which K can be attracted via the α th moment. For example, if $t P[X_1/b(t) \in \cdot] \xrightarrow{v} \nu_\alpha$ then we must have $E \xi^\alpha \leq 1$; this property will be examined further in the next section and appears in various forms in [1] and [22], in the stationary setting.

5.2. Joint tail convergence

What additional assumptions are necessary for convergences (b) and (c) of Proposition 5.1 to take place on the larger cone $\mathbb{E}_m^* = [0, \infty]^{m+1} \setminus \{\mathbf{0}\}$? This was considered by Basrak and Segers [1] and Segers [22] for stationary Markov chains. In (b), the dependence on the extremal threshold and, hence, on t , means that we are in the context of a triangular array and not, strictly speaking, in the setting of joint regular variation. However, the result is still useful; for example, to derive a point process convergence via the Poisson transform [20, p. 183].

As a first step, we characterize convergence on the larger cone by decomposing it into smaller, more familiar cones. This is similar to Theorem 6.1 of [22] and one of the implications of Theorem 2.1 of [1]. As a convention in what follows, set $[0, \infty]^0 \times A = A$. Also, recall the notation $\mathbb{E}_m = (0, \infty) \times [0, \infty]^m$.

Proposition 5.2. *Suppose that $Y_t = (Y_{t,0}, Y_{t,1}, \dots, Y_{t,m})$ is a random vector on $[0, \infty]^{m+1}$ for each $t > 0$. Then there exists a nonnull Radon measure μ^* on $\mathbb{E}_m^* = [0, \infty]^{m+1} \setminus \{\mathbf{0}\}$ such that*

$$t P[(Y_{t,0}, Y_{t,1}, \dots, Y_{t,m}) \in \cdot] \xrightarrow{v} \mu^*(\cdot) \quad \text{in } \mathbb{M}_+(\mathbb{E}_m^*) \text{ as } t \rightarrow \infty \tag{5.1}$$

if and only if, for $j = 0, \dots, m$, there exist Radon measures μ_j on $\mathbb{E}_j = (0, \infty) \times [0, \infty]^j$, not all null, such that

$$t P[(Y_{t,j}, \dots, Y_{t,m}) \in \cdot] \xrightarrow{v} \mu_{m-j}(\cdot) \quad \text{in } \mathbb{M}_+(\mathbb{E}_{m-j}). \tag{5.2}$$

The relation between the limit measures is

$$\mu_{m-j}(\cdot) = \mu^*([0, \infty]^j \times \cdot) \quad \text{on } \mathbb{E}_{m-j}$$

for $j = 0, \dots, m$, and

$$\mu^*([0, \mathbf{x}]^c) = \sum_{j=0}^m \mu_{m-j}((x_j, \infty) \times [0, x_{j+1}] \times \dots \times [0, x_m]) \quad \text{for } \mathbf{x} \in \mathbb{E}_m^*.$$

Furthermore, given $j \in \{0, \dots, m - 1\}$, if $A \subset [0, \infty]^{m-j} \setminus \{0\}^{m-j}$ is relatively compact then $\mu_{m-j}((0, \infty) \times A) < \infty$.

Proof. Assume first that (5.1) holds. Fixing $j \in \{0, \dots, m\}$, define

$$\mu_{m-j}(\cdot) := \mu^*([0, \infty]^j \times \cdot)$$

(i.e. $\mu_m = \mu^*$). Let $A \subset \mathbb{E}_{m-j}$ be relatively compact with $\mu_{m-j}(\partial A) = 0$. Then $A^* = [0, \infty]^j \times A$ is relatively compact in \mathbb{E}_m^* , and $\partial_{\mathbb{E}_m^*} A^* = [0, \infty]^j \times \partial_{\mathbb{E}_{m-j}} A$, so $\mu^*(\partial_{\mathbb{E}_m^*} A^*) = \mu_{m-j}(\partial A) = 0$. Therefore,

$$t \mathbb{P}[(Y_{t,j}, \dots, Y_{t,m}) \in A] = t \mathbb{P}[(Y_{t,0}, \dots, Y_{t,m}) \in A^*] \rightarrow \mu^*(A^*) = \mu_{m-j}(A),$$

establishing (5.2).

Conversely, suppose that (5.2) holds for $j = 0, \dots, m$. For $\mathbf{x} \in (0, \infty]^{m+1}$, define

$$h(\mathbf{x}) = \sum_{j=0}^m \mu_{m-j}((x_j, \infty] \times [0, x_{j+1}] \times \dots \times [0, x_m]).$$

Decompose $[\mathbf{0}, \mathbf{x}]^c$ as a disjoint union

$$[\mathbf{0}, \mathbf{x}]^c = \bigcup_{j=0}^m [0, \infty]^j \times (x_j, \infty] \times [0, x_{j+1}] \times \dots \times [0, x_m], \tag{5.3}$$

and observe that at points of continuity of the limit

$$t \mathbb{P}[\mathbf{Y}_t \in [\mathbf{0}, \mathbf{x}]^c] = \sum_{j=0}^m t \mathbb{P}[(Y_{t,j}, \dots, Y_{t,m}) \in (x_j, \infty] \times [0, x_{j+1}] \times \dots \times [0, x_m]] \rightarrow h(\mathbf{x}). \tag{5.4}$$

Hence, (5.1) holds with the limit measure μ^* defined by $\mu^*([\mathbf{0}, \mathbf{x}]^c) = h(\mathbf{x})$. Indeed, given $f \in \mathcal{C}_K^+(\mathbb{E}_m^*)$, we can find $\delta > 0$ such that $\mathbf{x}_\delta = (\delta, \dots, \delta)$ is a continuity point of h , and f is supported on $[\mathbf{0}, \mathbf{x}_\delta]^c$. Therefore,

$$t \mathbb{E} f(\mathbf{Y}_t) \leq \sup_{\mathbf{x} \in \mathbb{E}_m^*} f(\mathbf{x}) \cdot \sup_{t>0} t \mathbb{P}[\mathbf{Y}_t \in [\mathbf{0}, \mathbf{x}_\delta]^c] < \infty,$$

implying that the set $\{t \mathbb{P}[\mathbf{Y}_t \in \cdot]; t > 0\}$ is relatively compact in $\mathbb{M}_+(\mathbb{E}_m^*)$ (see [20, p. 51]). Furthermore, if $t_k \mathbb{P}[\mathbf{Y}_{t_k} \in \cdot] \rightarrow \mu$ and $s_k \mathbb{P}[\mathbf{Y}_{s_k} \in \cdot] \rightarrow \mu'$ as $k \rightarrow \infty$, then $\mu = \mu' = \mu^*$ on sets $[\mathbf{0}, \mathbf{x}]^c$ which are continuity sets of μ^* by (5.4). This extends to measurable rectangles in \mathbb{E}_m^* bounded away from $\mathbf{0}$ whose vertices are continuity points of h , leading us to the conclusion that $\mu = \mu' = \mu^*$ on \mathbb{E}_m^* .

Moreover, since we can decompose $[\mathbf{0}, \mathbf{x}]^c$ for any $\mathbf{x} \in \mathbb{E}_m^*$ as in (5.3), it is clear that μ^* is nonnull if and only if not all of the μ_j are null.

Finally, for $1 \leq j \leq m - 1$, if $A \subset [0, \infty]^{m-j} \setminus \{0\}^{m-j}$ is relatively compact then it is contained in $[(0, \dots, 0), (x_{j+1}, \dots, x_m)]^c$ for some $(x_{j+1}, \dots, x_m) \in (0, \infty]^{m-j}$. Applying (5.3) once again, we find that

$$\begin{aligned} \mu_{m-j}((0, \infty] \times A) &= \mu^*([0, \infty]^j \times (0, \infty] \times A) \\ &\leq \sum_{k=j+1}^m \mu^*([0, \infty]^{j+1} \times [0, \infty]^{k-j-1} \times (x_k, \infty] \times [0, x_{k+1}] \\ &\quad \times \dots \times [0, x_m]) \\ &= \sum_{k=j+1}^m \mu_{m-k}((x_k, \infty] \times [0, x_{k+1}] \times \dots \times [0, x_m]) \\ &< \infty. \end{aligned}$$

Consequently, the extension of the convergences in Proposition 5.1 to the larger cone \mathbb{E}_m^* follows from regular variation of the marginal tails.

Theorem 5.1. *Suppose that $X \sim K \in D(G)$, and let $b(t) \rightarrow \infty$ be a scaling function and $\alpha > 0$. Let $\xi \sim G$. Then*

$$t \mathbb{P}[b(t)^{-1}(X_0^{(b(t))}, X_1^{(b(t))}, \dots, X_m^{(b(t))}) \in \cdot] \xrightarrow{v} \mu^*(\cdot) \text{ in } \mathbb{M}_+(\mathbb{E}_m^*) \text{ as } t \rightarrow \infty, \tag{5.5}$$

where μ^* is a Radon measure on \mathbb{E}_m^* given by

$$\mu^*|_{\mathbb{E}_m}(\mathbf{d}x_0, \mathbf{d}x_m) = v_\alpha(\mathbf{d}x_0) \mathbb{P}_{x_0}[(T_1, \dots, T_m) \in \mathbf{d}x_m] \text{ and } \mu^*(\mathbb{E}_m^* \setminus \mathbb{E}_m) = 0,$$

if and only if $\mathbb{E} \xi^\alpha < \infty$ and

$$t \mathbb{P}\left[\frac{X_j^{(b(t))}}{b(t)} \in \cdot\right] \xrightarrow{v} (\mathbb{E} \xi^\alpha)^j v_\alpha(\cdot) \text{ in } \mathbb{M}_+(0, \infty], j = 0, \dots, m. \tag{5.6}$$

Proof. Assume first that (5.5) holds, and let $x > 0$. Then, for $j = 0$,

$$t \mathbb{P}[X_0^{(b(t))} > b(t)x] \rightarrow \mu^*((x, \infty] \times [0, \infty]^m) = x^{-\alpha}.$$

For $j \geq 1$, (5.5) implies that

$$\begin{aligned} t \mathbb{P}[X_j^{(b(t))} > b(t)x] &\rightarrow \mu^*([0, \infty]^j \times (x, \infty] \times [0, \infty]^{m-j}) \\ &= \int_{(0, \infty]} v_\alpha(\mathbf{d}u) \mathbb{P}[\xi_1 \cdots \xi_j > xu^{-1}] \\ &= x^{-\alpha} \mathbb{E}(\xi_1 \cdots \xi_j)^\alpha \\ &= x^{-\alpha} (\mathbb{E} \xi^\alpha)^j \\ &< \infty \end{aligned}$$

since the set $[0, \infty]^j \times (x, \infty] \times [0, \infty]^{m-j}$ is relatively compact in \mathbb{E}_m^* . In particular, we have $\mathbb{E} \xi^\alpha < \infty$.

Conversely, suppose that (5.6) holds. Lemma A.4 in Appendix A implies that, in $\mathbb{M}_+(\mathbb{E}_{m-j})$,

$$\begin{aligned} t \mathbb{P}[b(t)^{-1}(X_j^{(b(t))}, \dots, X_m^{(b(t))}) \in (\mathbf{d}x_0, \mathbf{d}x)] \\ \xrightarrow{v} (\mathbb{E} \xi^\alpha)^j v_\alpha(\mathbf{d}x_0) \mathbb{P}_{x_0}[(T_1, \dots, T_{m-j}) \in \mathbf{d}x] \\ =: \mu_{m-j}((\mathbf{d}x_0, \mathbf{d}x)) \end{aligned}$$

by the Markov property, and Proposition 5.2 yields (5.5) with $\mu^*|_{\mathbb{E}_m}(\cdot) = \mu_m(\cdot)$. It remains to verify that $\mu^*(\mathbb{E}_m^* \setminus \mathbb{E}_m) = 0$. Writing $A_j(\mathbf{x}_m) = [0, \infty]^{j-1} \times (x_j, \infty] \times [0, \infty]^{m-j}$, note that

$$\mu^*({0} \times [0, \mathbf{x}_m]^c) \leq \sum_{j=1}^m \mu^*({0} \times A_j(\mathbf{x}_m)),$$

and

$$\begin{aligned} \mu^*({0} \times A_j(\mathbf{x}_m)) &= \mu_{m-j}((x_j, \infty] \times [0, \infty]^{m-j}) - \mu_m((0, \infty] \times A_j(\mathbf{x}_m)) \\ &= (\mathbb{E} \xi^\alpha)^j x_j^{-\alpha} - \lim_{x_0 \downarrow 0} \int_{[x_0, \infty]} v_\alpha(\mathbf{d}u) \mathbb{P}[\xi_1 \cdots \xi_j > x_j u^{-1}] \\ &= (\mathbb{E} \xi^\alpha)^j x_j^{-\alpha} - (\mathbb{E} \xi^\alpha)^j x_j^{-\alpha} \\ &= 0. \end{aligned}$$

Since $\mathbb{E}_m^* \setminus \mathbb{E}_m = \{0\} \times \mathbb{E}_{m-1}^*$, it follows that $\mu^*(\mathbb{E}_m^* \setminus \mathbb{E}_m) = 0$.

At the end of Section 4, cases were outlined in which we could replace $X_j^{(b(t))}$ by X_j . Theorem 5.1 is most striking for these since it shows that, for a Markov chain whose kernel is in a domain of attraction, to obtain joint regular variation of the FDDs, it is enough to know that the marginal tails are regularly varying. In particular, if X has a regularly varying stationary distribution then the FDDs are jointly regularly varying. This result was presented by Segers [22], and Basrak and Segers [1] showed that, for a general stationary process, joint regular variation of FDDs is equivalent to the existence of a ‘tail process’ which reduces to the tail chain in the case of Markov chains. However, what Proposition 5.1 emphasizes is that it is the marginal tail behavior alone, rather than stationarity, which provides the link with joint regular variation.

Theorem 5.1 also extends the observation made in Section 5.1 that knowledge of the marginal tail behavior for a Markov chain whose kernel is in a domain of attraction constrains the class of possible limit distributions G via its moments. If a particular choice of regularly varying initial distribution leads to $t P[X_j > b(t) \cdot] \xrightarrow{v} a_j v_\alpha(\cdot)$ then we have $E \xi^\alpha \leq a_j^{1/j}$. In particular, if X admits a stationary distribution whose tail is $RV_{-\alpha}$ then $E \xi^\alpha \leq 1$.

6. Examples

Our first example illustrates the main results.

Example 6.1. Let $V = (Z, \eta)$ be any random vector on $[0, \infty) \times \mathbb{R}$. Consider the update function $\psi(y, V) = (Zy + \eta)_+$ and its canonical form (with $W = (Z, \eta)$)

$$\psi(y, V) = Zy + \phi(y, W) = Zy + (\eta \mathbf{1}_{\{\eta > -Zy\}} - Zy \mathbf{1}_{\{\eta \leq -Zy\}}).$$

For $y > 0$ and $x \geq 0$, the transition kernel has the form $K(y, (x, \infty)) = P[Zy + \eta > x]$. Since $t^{-1}\psi(t, V) = (Z + t^{-1}\eta)_+ \rightarrow Z$ almost surely (a.s.), we have $K \in D(G)$ with $G = P[Z \in \cdot]$. Furthermore, using Proposition 3.1, the function $\gamma(t) \equiv \sqrt{t}$ is of larger order than $\phi(t, w)$, so $y(t) = 1/\sqrt{t}$ is an extremal boundary. Since $\phi(\cdot, w)$ is bounded on neighborhoods of 0, Proposition 4.1(c) implies that K satisfies the regularity condition (4.3). Consequently, from Theorem 4.1, we obtain FDD convergence of $t^{-1}X$ to T as in (4.6).

If K does not satisfy the regularity condition (4.3), Theorem 4.1 may fail to hold, and starting from tu , $t^{-1}X$ may fail to converge to T started from u .

Example 6.2. Let $V = (Z, W, W')$ be any nondegenerate random vector on $[0, \infty)^3$, and consider the Markov chain determined by the update function

$$\psi(y, V) = Zy + Wy^{-1} \mathbf{1}_{\{y>0\}} + W' \mathbf{1}_{\{y=0\}}.$$

For $y > 0$ and $x \geq 0$, the transition kernel is $K(y, (x, \infty)) = P[Zy + Wy^{-1} > x]$, and, since $t^{-1}\psi(t, V) = Z + Wt^{-2} \rightarrow Z$ a.s., we have $K \in D(G)$ with $G = P[Z \in \cdot]$. Furthermore, using Proposition 3.1, the function $\gamma(t) \equiv 1$ is of larger order than $\phi(t, w)$, so $y(t) = 1/t$ is an extremal boundary.

However, note that $\phi(y, (W, W')) = Wy^{-1} \mathbf{1}_{\{y>0\}} + W' \mathbf{1}_{\{y=0\}}$ is unbounded near 0, implying that Segers’ boundedness condition (4.2) does not hold. In fact, our form of the regularity condition (4.3) fails for K . Indeed,

$$K(tu_t, t(x, \infty)) = P\left[Ztu_t + \frac{W}{tu_t} > tx \right] = P\left[Zu_t + \frac{W}{t^2u_t} > x \right].$$

Choosing $u_t = t^{-2}$ yields $K(tu_t, t(x, \infty)) \rightarrow P[W > x]$. For appropriate x , this shows that (4.3) fails.

Not only does (4.3) fail but so does Theorem 4.1, since the asymptotic behavior of X is not the same as that of $X^{(t)}$. We show directly that the conditional FDDs of $t^{-1}X$ fail to converge to those of T . The idea is that, if $X_k < y(t) = t^{-1}$, there is a positive probability that $X_{k+1} > t$. We illustrate this for $m = 2$. Take $f \in \mathcal{C}[0, \infty]^2$ and $u > 0$. Observe that, if $X_0 = tu > 0$, from the definition of ψ , $X_1 = Z_1tu + W_1/(tu)$ and $X_2 = Z_2X_1 + (W_2/X_1)\mathbf{1}_{\{X_1>0\}} + W'_2\mathbf{1}_{\{X_1=0\}}$. Furthermore, on $\{Z_1 > 0\}$, we have $X_1 > 0$ and $X_2 = Z_2X_1 + W_2/X_1$. On $\{Z_1 = 0, W_1 > 0\}$, $X_1 > 0$ and $X_2 = Z_2X_1 + W_2/X_1$. On $\{Z_1 = 0, W_1 = 0\}$, we have $X_1 = 0$ and $X_2 = W'_2$. Therefore,

$$\begin{aligned} E_{tu} f\left(\frac{X_1}{t}, \frac{X_2}{t}\right) &= E_{tu} f\left(\frac{X_1}{t}, \frac{X_2}{t}\right) \mathbf{1}_{\{Z_1>0\}} + E_{tu} f\left(\frac{X_1}{t}, \frac{X_2}{t}\right) \mathbf{1}_{\{Z_1=0, W_1>0\}} \\ &\quad + E_{tu} f\left(\frac{X_1}{t}, \frac{X_2}{t}\right) \mathbf{1}_{\{Z_1=0, W_1=0\}} \\ &= A + B + C. \end{aligned}$$

For A , as $t \rightarrow \infty$, we have

$$\begin{aligned} A &= E f\left(Z_1u + \frac{W_1}{t^2u}, Z_2\left[Z_1u + \frac{W_1}{t^2u}\right] + \frac{W_2}{Z_1t^2u + W_1u^{-1}}\right) \mathbf{1}_{\{Z_1>0\}} \\ &\rightarrow E f(Z_1u, Z_1Z_2u) \mathbf{1}_{\{Z_1>0\}}, \end{aligned}$$

while for B we obtain, for $t \rightarrow \infty$,

$$B = E f\left(\frac{W_1}{t^2u}, \frac{Z_2W_1}{t^2u} + \frac{W_2u}{W_1}\right) \mathbf{1}_{\{Z_1=0, W_1>0\}} \rightarrow E f\left(0, \frac{uW_2}{W_1}\right) \mathbf{1}_{\{Z_1=0, W_1>0\}}.$$

Finally, for C ,

$$\begin{aligned} C &= E f\left(0, \frac{W'_2}{t}\right) \mathbf{1}_{\{Z_1=0, W_1=0\}} \\ &= P[Z_1 = 0, W_1 = 0] E f\left(0, \frac{W'_2}{t}\right) \\ &\rightarrow P[Z_1 = 0, W_1 = 0] f(0, 0). \end{aligned}$$

Observe that $\lim_{t \rightarrow \infty} [A + B + C] \neq E_u f(T_1, T_2) = E f(uZ_1, uZ_1Z_2)$.

In the final example, the conditional distributions of $t^{-1}X$ converge to those of the tail chain T , even though the regularity condition does not hold. This includes cases for which $G(\{0\}) = 0$ and $G(\{0\}) > 0$ with extremal boundary $y(t) \equiv 0$.

Example 6.3. Let $\{(\xi_j, \eta_j), j \geq 1\}$ be i.i.d. copies of the nondegenerate random vector (ξ, η) on $[0, \infty)^2$. Taking $V = (\xi, \eta)$, consider a Markov chain that transitions according to the update function

$$\psi(y, V) = \xi(y + y^{-1}) \mathbf{1}_{\{y>0\}} + \eta \mathbf{1}_{\{y=0\}} = \xi y + (\xi y^{-1} \mathbf{1}_{\{y>0\}} + \eta \mathbf{1}_{\{y=0\}}),$$

where the last expression is the canonical form (with $Z = \xi$ and $W = (\xi, \eta)$). For $y > 0$ and $x \geq 0$, the transition kernel is

$$K(y, [0, x]) = P[\xi(y + y^{-1}) \leq x] = P\left[\xi \leq \frac{x}{y + y^{-1}}\right].$$

For $t > 0$, $t^{-1}\psi(t, V) = \xi(1 + t^{-2}) \rightarrow \xi$ a.s., so $K \in D(G)$ with $G = P[\xi \in \cdot]$. Note that $\phi(y, W) = \xi y^{-1} \mathbf{1}_{\{y>0\}} + \eta \mathbf{1}_{\{y=0\}}$ is unbounded near 0, implying that Segers' boundedness condition (4.2) does not hold. Also, our regularity condition (4.3) fails for K . To see this, write

$$K(tu_t, t(x, \infty)) = P\left[\xi > \frac{x}{u_t + (t^2u_t)^{-1}}\right].$$

Fix x so that $P[\xi > x] > 0$, and choose $u_t = t^{-2}$. This yields $u_t + (t^2u_t)^{-1} = 1 + t^{-2}$, implying that

$$K(tu_t, t(x, \infty)) = P\left[\xi > \frac{x}{1 + t^{-2}}\right] \geq P[\xi > x] > 0,$$

so (4.3) fails for K . However, since $K(t, \{0\}) = P[\xi = 0] = G(\{0\})$, the choice $y(t) \equiv 0$ satisfies the definition of an extremal boundary (3.6), even if $G(\{0\}) > 0$. This leads to FDD convergence of $P_{tu}[t^{-1}X \in \cdot]$ to $P_u[T \in \cdot]$, and, thus, we learn that conclusion (4.6) of Theorem 4.1 may hold without (4.3) being true.

7. Concluding remarks

We have thus placed the traditional tail chain model for the extremes of a Markov chain in a more general context through the introduction of the extremal boundary. A common application of the tail chain model is in deriving the weak limits of exceedance point processes for X [1], [17], [21]. We will shortly use our results to develop a detailed description of the clustering properties of extremes of Markov chains by means of such point processes. Furthermore, as we have not employed stationarity in our finite-dimensional results, we propose to substitute the inherent regenerative structure of a Harris recurrent Markov chain for the traditional assumption of stationarity.

Appendix A. Technical lemmas

This section collects lemmas needed to prove convergence of integrals of the form $\int f_n d\mu_n$, assuming that $f_n \rightarrow f$ and $\mu_n \rightarrow \mu$ in their respective spaces. An example is the *second continuous mapping theorem* [2, Theorem 5.5, p. 34].

Lemma A.1. *Assume that \mathbb{E} and \mathbb{E}' are complete separable (c.s.) metric spaces, and, for $n \geq 0$, $h_n: \mathbb{E} \rightarrow \mathbb{E}'$ are measurable. Put $A = \{x \in \mathbb{E}: h_n(x_n) \rightarrow h_0(x) \text{ whenever } x_n \rightarrow x\}$. If the P_n , $n \geq 0$, are probability measures on \mathbb{E} with $P_n \Rightarrow P_0$, and $h_n \rightarrow h_0$ almost uniformly in the sense that $P_0(A) = 1$, then $P_n \circ h_n^{-1} \Rightarrow P_0 \circ h_0^{-1}$ in \mathbb{E}' .*

The result provides a way to handle the convergence of a family of integrals.

Lemma A.2. *In addition to the assumptions of Lemma A.1, require that $\mathbb{E}' = \mathbb{R}$ and $\{h_n, n \geq 0\}$ is uniformly bounded, so $\sup_{n \geq 0} \sup_{x \in \mathbb{E}} |h_n(x)| < \infty$.*

(a) *We have*

$$\int_{\mathbb{E}} h_n dP_n \rightarrow \int_{\mathbb{E}} h_0 dP_0.$$

(b) *Suppose additionally that \mathbb{E} is locally compact with a countable base (l.c.c.b.), and that $\mu_n \xrightarrow{v} \mu_0$ is in $\mathbb{M}_+(\mathbb{E})$ with $\mu_0(A^c) = 0$. If there exists a compact set $B \in \mathcal{K}(\mathbb{E})$ with*

$\mu_0(\partial B) = 0$ such that $h_n(x) = 0, n \geq 0$, whenever $x \notin B$ (i.e. B is a common compact support of each h_n), then

$$\int_{\mathbb{E}} h_n \, d\mu_n \rightarrow \int_{\mathbb{E}} h_0 \, d\mu_0.$$

Proof. (a) If $X_n \sim P_n$ for $n \geq 0$ then $h_n(X_n) \Rightarrow h_0(X_0)$. The uniform boundedness of the h_n guarantees that $E h_n(X_n) \rightarrow E h_0(X_0)$.

(b) View B as a compact subspace of \mathbb{E} inheriting the relative topology. Then, assuming that $\mu_0(B) > 0$ to rule out a trivial case, define probabilities on B by $P_n(\cdot) = \mu_n(\cdot \cap B)/\mu_n(B), n \geq 0$. Since $\mu_n(\cdot \cap B) \xrightarrow{v} \mu_0(\cdot \cap B)$ by Proposition 3.3 of [11], and B is compact, we get $P_n \Rightarrow P_0$. Denote by $h'_n, n \geq 0$, the restriction of h_n to B . Observe that, for any $x \in A \cap B$, we have $h'_n(x_n) \rightarrow h'_0(x)$ whenever $x_n \rightarrow x$ in B , and $P_0(A^c \cap B) \leq \mu_0(A^c)/\mu_0(B) = 0$. Therefore, apply part (a) to obtain

$$\int_{\mathbb{E}} h_n \, d\mu_n = \int_{\mathbb{E}} h_n \mathbf{1}_B \, d\mu_n = \mu_n(B) \int_B h'_n \, dP_n \rightarrow \mu_0(B) \int_B h'_0 \, dP_0 = \int_{\mathbb{E}} h_0 \, d\mu_0.$$

A convenient specialization of Lemma A.2(b) is the following.

Lemma A.3. *Suppose that \mathbb{E} is l.c.c.b. and that $\mu_n \xrightarrow{v} \mu$ is in $\mathbb{M}_+(\mathbb{E})$. If $f: \mathbb{E} \rightarrow \mathbb{R}$ is continuous and bounded, and $B \in \mathbb{E}$ is relatively compact with $\mu(\partial B) = 0$, then*

$$\int_B f \, d\mu_n \rightarrow \int_B f \, d\mu.$$

Take $h_n = f \mathbf{1}_B$ for $n \geq 0$. Since $f \mathbf{1}_B$ is continuous except possibly on ∂B , we have $\mu(A^c) \leq \mu(\partial B) = 0$.

The next result is used to extend convergence of substochastic transition functions to multivariate regular variation on a larger space.

Lemma A.4. *Let $\mathbb{E} \subset [0, \infty]^m$ and $\mathbb{E}' \subset [0, \infty]^{m'}$ be two nice (l.c.c.b.) spaces. Suppose that for $t \geq 0, \{p^{(t)}(\cdot, \cdot)\}_{t \geq 0}$ are substochastic transition functions on $\mathbb{E} \times \mathcal{B}(\mathbb{E}')$. This means that $p^{(t)}(\cdot, B)$ is a measurable function for any fixed $B \in \mathcal{B}(\mathbb{E}')$, $p^{(t)}(x, \cdot)$ is a measure for any $x \in \mathbb{E}$, and $\sup_{t \geq 0} \sup_{u \in \mathbb{E}} p^{(t)}(u, \mathbb{E}') \leq 1$. Assume that there is a set $A \subset \mathbb{E}$ such that*

$$p^{(t)}(u_t, \cdot) \xrightarrow{v} p^{(0)}(u, \cdot) \text{ in } \mathbb{M}_+(\mathbb{E}') \text{ as } t \rightarrow \infty$$

whenever $u_t \rightarrow u$ in \mathbb{E} and $u \in A$. Suppose also that $\{v^{(t)}\}_{t \geq 0}$ are measures on \mathbb{E} such that $v^{(0)}(A^c) = 0$, and that $v^{(t)} \xrightarrow{v} v^{(0)}$ is in $\mathbb{M}_+(\mathbb{E})$. Then, defining measures $\mu^{(t)}$ for $t \geq 0$ on $\mathbb{E} \times \mathbb{E}'$ as

$$\mu^{(t)}(du, dx) = v^{(t)}(du) p^{(t)}(u, dx),$$

we have

$$\mu^{(t)} \xrightarrow{v} \mu^{(0)} \text{ in } \mathbb{M}_+(\mathbb{E} \times \mathbb{E}') \text{ as } t \rightarrow \infty.$$

Proof. Let $f \in C_K^+(\mathbb{E} \times \mathbb{E}')$; without loss of generality, assume that f is supported on $K \times K'$, where $K \in \mathcal{K}(\mathbb{E})$ and $K' \in \mathcal{K}(\mathbb{E}')$. We have

$$\int_{\mathbb{E} \times \mathbb{E}'} \mu^{(t)}(du, dx) f(u, x) = \int_{\mathbb{E}} v^{(t)}(du) \int_{\mathbb{E}'} p^{(t)}(u, dx) f(u, x).$$

For $t \geq 0$, write

$$\varphi_t(u) = \int_{\mathbb{E}'} p^{(t)}(u, dx) f(u, x),$$

and suppose that $u_t \rightarrow u_0$ with $u_0 \in A$; we verify that $\varphi_t(u_t) \rightarrow \varphi_0(u_0)$. Writing $g_t(x) = f(u_t, x)$, $t \geq 0$, we have $g_t(x_t) \rightarrow g_0(x_0)$ whenever $x_t \rightarrow x_0 \in \mathbb{E}'$ by the continuity of f . Also, the g_t are uniformly bounded by the bound on f and $g_t(x) = 0$ for all t whenever $x \notin K'$. Furthermore, without loss of generality, we can assume that $p^{(0)}(u_0, \partial K') = 0$. Now apply Lemma A.2(b) to obtain

$$\varphi_t(u_t) = \int_{\mathbb{E}'} p^{(t)}(u_t, dx) g_t(x) \rightarrow \int_{\mathbb{E}'} p^{(0)}(u_0, dx) g_0(x) = \varphi_0(u_0).$$

Since the $p^{(t)}$ are substochastic, and $\varphi_t(u) = 0$ for all t whenever $u \notin K$, the φ_t are uniformly bounded by the bound on f . Assume similarly that $v^{(0)}(\partial K) = 0$, and recall that $v^{(0)}(A^c) = 0$. Apply Lemma A.2(b) once more to conclude that, as $t \rightarrow \infty$,

$$\begin{aligned} \int_{\mathbb{E} \times \mathbb{E}'} \mu^{(t)}(du, dx) f(u, x) &= \int_{\mathbb{E}} v^{(t)}(du) \varphi_t(u) \\ &\rightarrow \int_{\mathbb{E}} v^{(0)}(du) \varphi_0(u) \\ &= \int_{\mathbb{E} \times \mathbb{E}'} \mu^{(0)}(du, x) f(u, x). \end{aligned}$$

We conclude this section with a result used to verify the existence of the extremal boundary.

Lemma A.5. *Suppose that the P_t , $t \geq 0$, are probability measures on a c.s. metric space \mathbb{E} such that $P_t \Rightarrow P_0$, and let $A \subset \mathbb{E}$ be measurable. Then there exists a sequence of sets $A_t \downarrow \bar{A}$ such that $P_t(A_t) \rightarrow P_0(\bar{A})$.*

Remark. Note that if $P_0(\partial A) = 0$ then we can take $A_t = \bar{A}$. In the case of distribution functions $F_t \Rightarrow F$ on \mathbb{R}^m , taking $A = (-\infty, \mathbf{x}]$ and metric $\rho = \rho_\infty$ shows that, for any $\mathbf{x} \in \mathbb{R}^m$, there exists $\mathbf{x}_t \downarrow \mathbf{x}$ such that $F_t(\mathbf{x}_t) \rightarrow F(\mathbf{x})$.

Proof of Lemma A.5. Let ρ be a metric on \mathbb{E} , and consider sets $A_\delta = \{x: \rho(x, A) \leq \delta\}$. Recall that $P_0(\partial A_\delta) = 0$ for all but a countable number of choices of δ , since $F(\delta) = P_0(A_\delta) - P_0(\bar{A})$ is a distribution function. First choose $\{\delta_k: k = 1, 2, \dots\}$ such that $0 < \delta_{k+1} \leq \delta_k \wedge 1/(k + 1)$ and $P_0(\partial A_{\delta_k}) = 0$ for all k . Next, let $s_0 = 0$ and take $s_k \geq s_{k-1} + 1$, $k = 1, 2, \dots$, such that $P_t(A_{\delta_k}) > P_0(\bar{A}) - 1/k$ whenever $t \geq s_k$; this is possible since $P_t(A_{\delta_k}) \rightarrow P_0(A_{\delta_k}) \geq P_0(\bar{A})$ for all k . Finally, for $t > 0$, set

$$A(t) = A_{\delta_1} \mathbf{1}_{(0, s_1)}(t) + \sum_{k=1}^{\infty} A_{\delta_k} \mathbf{1}_{[s_k, s_{k+1})}(t).$$

We claim that $A(t) \downarrow \bar{A}$ and that $P_t(A(t)) \rightarrow P_0(\bar{A})$ as $t \rightarrow \infty$. It is clear that $A(t) \supset A(t')$ for $t \leq t'$, and $\bigcap_t A(t) = \bigcap_k A_{\delta_k} = \bar{A}$. On the one hand, for large t , we have $A(t) \subset A_{\delta_k}$ for any k , so

$$\limsup_{t \rightarrow \infty} P_t(A(t)) \leq \limsup_{t \rightarrow \infty} P_t(A_{\delta_k}) \leq P_0(A_{\delta_k}).$$

Letting $k \rightarrow \infty$ shows that $\limsup_{t \rightarrow \infty} P_t(A(t)) \leq P_0(\bar{A})$. On the other hand, if $k(t)$ denotes the value of k for which $s_k \leq t < s_{k+1}$, then

$$P_t(A(t)) = P_t(A_{\delta_{k(t)}}) > P_0(\bar{A}) - \frac{1}{k(t)},$$

so $\liminf_t P_t(A(t)) \geq P_0(\bar{A})$. Combining these two inequalities shows that $P_t(A(t)) \rightarrow P_0(\bar{A})$.

Acknowledgements

S. I. Resnick and D. Zeber were partially supported by ARO contract W911NF-10-1-0289 and NSA grant H98230-11-1-0193 at Cornell University. The authors wish to thank an anonymous referee who provided many helpful comments and suggestions.

References

- [1] BASRAK, B. AND SEGERS, J. (2009). Regularly varying multivariate time series. *Stoch. Process. Appl.* **119**, 1055–1080.
- [2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.
- [3] BILLINGSLEY, P. (1971). *Weak Convergence of Measures: Applications in Probability*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [4] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
- [5] BORTOT, P. AND COLES, S. (2000). A sufficiency property arising from the characterization of extremes of Markov chains. *Bernoulli* **6**, 183–190.
- [6] BORTOT, P. AND COLES, S. (2003). Extremes of Markov chains with tail switching potential. *J. R. Statist. Soc. B* **65**, 851–867.
- [7] BORTOT, P. AND TAWN, J. A. (1998). Models for the extremes of Markov chains. *Biometrika* **85**, 851–867.
- [8] DAS, B. AND RESNICK, S. I. (2011). Conditioning on an extreme component: model consistency with regular variation on cones. *Bernoulli* **17**, 226–252.
- [9] DAS, B. AND RESNICK, S. I. (2011). Detecting a conditional extreme value model. *Extremes* **14**, 29–61.
- [10] DAVIS, R. A. AND HSING, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Prob.* **23**, 879–917.
- [11] FEIGIN, P. D., KRATZ, M. F. AND RESNICK, S. I. (1996). Parameter estimation for moving averages with positive innovations. *Ann. Appl. Prob.* **6**, 1157–1190.
- [12] HEFFERNAN, J. E. AND RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Prob.* **17**, 537–571.
- [13] HEFFERNAN, J. E. AND TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B* **66**, 497–546.
- [14] HSING, T. (1989). Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.* **29**, 274–291.
- [15] KALLENBERG, O. (1997). *Foundations of Modern Probability*. Springer, New York.
- [16] LEADBETTER, M. R., LINDGREN, G. AND ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.
- [17] PERFEKT, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Prob.* **4**, 529–548.
- [18] PERFEKT, R. (1997). Extreme value theory for a class of Markov chains with values in \mathbb{R}^d . *Adv. Appl. Prob.* **29**, 138–164.
- [19] RESNICK, S. I. (2007). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.
- [20] RESNICK, S. I. (2007). *Heavy-Tail Phenomena*. Springer, New York.
- [21] ROOTZÉN, H. (1988). Maxima and exceedances of stationary Markov chains. *Adv. Appl. Prob.* **20**, 371–390.
- [22] SEGERS, J. (2007). Multivariate regular variation of heavy-tailed Markov chains. Preprint. Available at <http://arxiv.org/abs/math/0701411v1>.
- [23] SMITH, R. L. (1992). The extremal index for a Markov chain. *J. Appl. Prob.* **29**, 37–45.
- [24] SMITH, R. L., TAWN, J. A. AND COLES, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika* **84**, 249–268.
- [25] YUN, S. (1998). The extremal index of a higher-order stationary Markov chain. *Ann. Appl. Prob.* **8**, 408–437.
- [26] YUN, S. (2000). The distributions of cluster functionals of extreme events in a d th-order Markov chain. *J. Appl. Prob.* **37**, 29–44.