

## Trickle-down processes and their boundaries\*

Steven N. Evans<sup>†</sup>    Rudolf Grübel<sup>‡</sup>    Anton Wakolbinger<sup>§</sup>

### Abstract

It is possible to represent each of a number of Markov chains as an evolving sequence of connected subsets of a directed acyclic graph that grow in the following way: initially, all vertices of the graph are unoccupied, particles are fed in one-by-one at a distinguished source vertex, successive particles proceed along directed edges according to an appropriate stochastic mechanism, and each particle comes to rest once it encounters an unoccupied vertex. Examples include the binary and digital search tree processes, the random recursive tree process and generalizations of it arising from nested instances of Pitman’s two-parameter Chinese restaurant process, tree-growth models associated with Mallows’  $\phi$  model of random permutations and with Schützenberger’s non-commutative  $q$ -binomial theorem, and a construction due to Luczak and Winkler that grows uniform random binary trees in a Markovian manner. We introduce a framework that encompasses such Markov chains, and we characterize their asymptotic behavior by analyzing in detail their Doob-Martin compactifications, Poisson boundaries and tail  $\sigma$ -fields.

**Keywords:** harmonic function;  $h$ -transform; tail  $\sigma$ -field; Poisson boundary; internal diffusion limited aggregation; binary search tree; digital search tree; Dirichlet random measure; random recursive tree; Chinese restaurant process; random partition; Ewens sampling formula; Griffiths–Engen–McCloskey distribution; Mallows model;  $q$ -binomial theorem; Catalan number; composition; quincunx.

**AMS MSC 2010:** Primary 60J50, Secondary 60J10; 68W40.

Submitted to EJP on October 3, 2010, final version accepted on November 21, 2011.

Supersedes arXiv:1010.0453v2.

## 1 Introduction

Several stochastic processes appearing in applied probability may be viewed as growing connected subsets of a directed acyclic graph that evolve according to the following dynamics: initially, all vertices of the graph are unoccupied, particles are fed in one-by-one at a distinguished source vertex, successive particles proceed along

\*SNE supported in part by NSF grants DMS-0405778 and DMS-0907630.

<sup>†</sup>Department of Statistics, University of California, 367 Evans Hall #3860, Berkeley, CA 94720-3860, USA. E-mail: evans@stat.berkeley.edu

<sup>‡</sup>Institut für Mathematische Stochastik, Leibniz Universität Hannover, Postfach 6009, 30060 Hannover, Germany. E-mail: rgrubel@stochastik.uni-hannover.de

<sup>§</sup>Institut für Mathematik, Goethe-Universität, 60054 Frankfurt am Main, Germany. E-mail: wakolbinger@math.uni-frankfurt.de

directed edges according to an appropriate stochastic mechanism, and each particle comes to rest once it encounters an unoccupied vertex. If we picture the source vertex as being at the “top” of the graph, then successive particles “trickle down” the graph until they find a vacant vertex that they can occupy.

We are interested in the question: “What is the asymptotic behavior of such a (highly transient) set-valued Markov chain?” For several of the models we consider, any finite neighborhood of the source vertex will, with probability one, be eventually occupied by a particle and so a rather unilluminating answer to our question is to say in such cases that the sequence of sets converges to the entire vertex set  $V$ . Implicit in the use of the term “converges” in this statement is a particular topology on the collection of subsets of  $V$ ; we are embedding the space of finite subsets of  $V$  into the Cartesian product  $\{0, 1\}^V$  and equipping the product space with the usual product topology. A quest for more informative answers can therefore be thought of as a search for an embedding of the state space of the chain into a topological space with a richer class of possible limits.

An ideal embedding would be one such that the chain converged almost surely to a limit and the  $\sigma$ -field generated by the limit coincided with the tail  $\sigma$ -field of the chain up to null events. For trickle-down processes, the Doob-Martin compactification provides such an embedding, and so our aim is to develop a body of theory that enables us to identify the compactification for at least some interesting examples. Moreover, a knowledge of the Doob-Martin compactification allows us to determine, via the Doob  $h$ -transform construction, all the ways in which it is possible, loosely speaking, to condition the Markov chain to behave for large times. This allows us to construct interesting new processes from existing ones or recognize that two familiar processes are related by such a conditioning.

A prime example of a Markov chain that fits into the trickle-down framework is the *binary search tree (BST) process*, and so we spend some time describing the BST process in order to give the reader some concrete motivation for the definitions we introduce later. The BST process and the related *digital search tree (DST) processes* that we consider in Section 5 arise from considering the behavior of tree-based searching and sorting algorithms. The trickle-down mechanism is at the heart of both algorithms: the vertices of the complete rooted binary tree are regarded as potential locations for the storage of data values  $x_1, x_2, \dots$  that arrive sequentially in time. We interpret these values as labels of particles. The particles are fed in at the root vertex, which receives  $x_1$ , and they are routed through the tree until a free vertex is found. How we travel onwards from an occupied vertex depends on the algorithm: in the BST case we assume that the input stream consists of real numbers and we compare the value  $x$  to be inserted with the content  $y$  of the occupied vertex, moving to the left or right depending on whether  $x < y$  or  $x > y$ , whereas in the DST case the inputs  $x_i$  are taken to be infinite 0-1 sequences, and we move from an occupied vertex of depth  $k$  to its left or right child if the  $k^{\text{th}}$  component of  $x_i$  is 0 or 1 respectively. If the input is random and we ignore the labeling of the vertices by elements of the input data sequence, then we obtain a sequence of subtrees of the complete binary tree; the  $n$ -th element of the sequence is the subtree consisting of the vertices occupied by the first  $n$  particles.

Binary trees in general and their role in the theory and practice of computer science are discussed in [32]. Several tree-based sorting and searching algorithms are described in [33]. In particular, a class of trees (generalizing binary search trees as well as digital search trees) with a construction similar to our trickle-down process is introduced in [10]. An introduction to the literature on tree-valued stochastic processes arising in this connection is [38]. Historically, real valued functionals such as the path length or the insertion depth of the next item were investigated first, with an emphasis

on the expected value for random input as a function of the amount of stored data (that is, of the number of vertices in the tree). In recent years, several infinite-dimensional random quantities related to the shape of the trees such as the node depth profile [6, 15], the subtree size profile [12, 17] and the silhouette [27] have been studied.

In the present paper we develop a framework for trickle-down processes that contains the BST and DST processes as special cases. As a consequence, we obtain limit results for the sequence of random trees themselves, using a topology on the space of finite binary trees that is dictated by the underlying stochastic mechanism. We also establish distributional relationships; for example, we show that the Markov chains generated by the BST and the DST algorithms are related via  $h$ -transforms – see Theorem 5.1.

In order to motivate our later formal definition of trickle-down processes, we now reconsider the BST process from a slightly different point of view by moving away somewhat from the search tree application and starting with a bijection from classical enumerative combinatorics (see, for example, [51]) between permutations of the finite set  $[n] := \{1, 2, \dots, n\}$  and certain trees with  $n$  vertices labeled by  $[n]$ .

Denote by  $\{0, 1\}^* := \bigsqcup_{k=0}^{\infty} \{0, 1\}^k$  the set of finite tuples or *words* drawn from the alphabet  $\{0, 1\}$  (with the empty word  $\emptyset$  allowed) – the symbol  $\bigsqcup$  emphasizes that this is a disjoint union. Write an  $\ell$ -tuple  $(v_1, \dots, v_\ell) \in \{0, 1\}^*$  more simply as  $v_1 \dots v_\ell$ . Define a directed graph with vertex set  $\{0, 1\}^*$  by declaring that if  $u = u_1 \dots u_k$  and  $v = v_1 \dots v_\ell$  are two words, then  $(u, v)$  is a directed edge (that is,  $u \rightarrow v$ ) if and only if  $\ell = k + 1$  and  $u_i = v_i$  for  $i = 1, \dots, k$ . Call this directed graph the *complete rooted binary tree*. Say that  $u < v$  for two words  $u = u_1 \dots u_k$  and  $v = v_1 \dots v_\ell$  if  $k < \ell$  and  $u_1 \dots u_k = v_1 \dots v_k$ ; that is,  $u < v$  if there exist words  $w_0, w_1, \dots, w_{\ell-k}$  with  $u = w_0 \rightarrow w_1 \rightarrow \dots \rightarrow w_{\ell-k} = v$ .

A *finite rooted binary tree* is a non-empty subset  $\mathbf{t}$  of  $\{0, 1\}^*$  with the property that if  $v \in \mathbf{t}$  and  $u \in \{0, 1\}^*$  is such that  $u \rightarrow v$ , then  $u \in \mathbf{t}$ . The vertex  $\emptyset$  (that is, the empty word) belongs to any such tree  $\mathbf{t}$  and is the *root* of  $\mathbf{t}$ . See Figure 1.

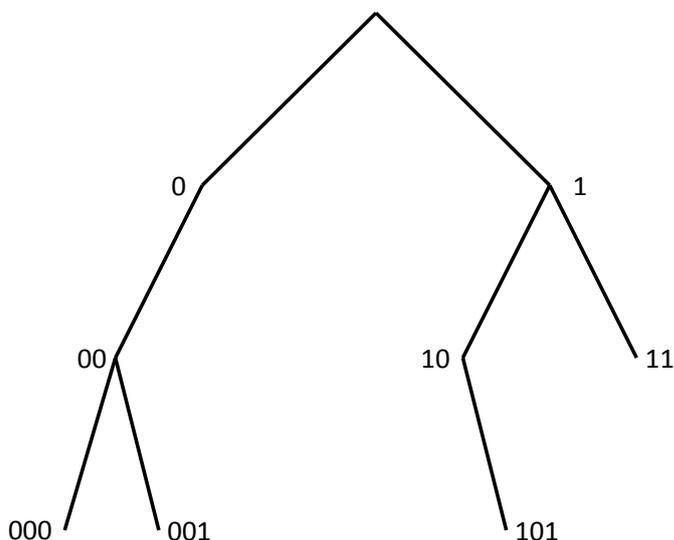


Figure 1: A finite rooted binary tree.

If  $\#\mathbf{t} = n$ , then a *labeling* of  $\mathbf{t}$  by  $[n]$  is a bijective map  $\phi : \mathbf{t} \rightarrow [n]$ .

Suppose that  $r(1), \dots, r(n)$  is an ordered listing of  $[n]$ . Define a permutation  $\pi$  of  $[n]$  by  $\pi^{-1}(k) = r(k)$ ,  $k \in [n]$ . There is a unique pair  $(\mathbf{t}, \phi)$ , where  $\mathbf{t}$  is a finite rooted binary tree with  $\#\mathbf{t} = n$  and  $\phi$  is a *labeling* of  $\mathbf{t}$  by  $[n]$ , such that

- $\phi(\emptyset) = 1$ ,
- if  $u, v \in \mathbf{t}$  and  $u < v$ , then  $\phi(u) < \phi(v)$ ,
- if  $u, v \in \mathbf{t}$ ,  $u0 \leq v$ , then  $\pi \circ \phi(u) > \pi \circ \phi(v)$ .
- if  $u, v \in \mathbf{t}$ ,  $u1 \leq v$ , then  $\pi \circ \phi(u) < \pi \circ \phi(v)$ .

The labeling may be constructed inductively as follows. If  $n = 1$ , then we just have the tree consisting of the root  $\emptyset$  labeled with 1. For  $n > 1$  we first remove  $n$  from the list  $r(1), \dots, r(n)$  and build the labeled tree  $(\mathbf{s}, \psi)$  for the resulting listing of  $[n - 1]$ . The labeled tree for  $r(1), \dots, r(n)$  is of the form  $(\mathbf{t}, \phi)$ , where  $\mathbf{t} = \mathbf{s} \cup \{u\}$  for  $u \notin \mathbf{s}$ ,  $\phi(u) = n$ ,  $\phi$  restricted to  $\mathbf{s}$  is  $\psi$ , and, setting  $u = u_1 \dots u_k$ ,

$$u_\ell = \begin{cases} 0, & \text{if } \pi \circ \psi(u_1 \dots u_{\ell-1}) < \pi(n), \\ 1, & \text{if } \pi \circ \psi(u_1 \dots u_{\ell-1}) > \pi(n). \end{cases}$$

To illustrate this construction, take  $n = 9$  and consider the ordered listing  $r(1), \dots, r(9)$  of the set  $[9]$  to be 8, 7, 9, 4, 1, 3, 5, 2, 6. See Table 1 for the resulting permutation, written in the usual two line format.

$k$	1	2	3	4	5	6	7	8	9
$\pi(k)$	5	8	6	4	7	9	2	1	3

Table 1: Permutation of  $[9]$  with 8, 7, 9, 4, 1, 3, 5, 2, 6 as the corresponding ordered listing  $r(1), \dots, r(9)$ .

The successive ordered listings of  $[1], \dots, [9]$  implicit in the recursive construction are

- 1
- 1, 2
- 1, 3, 2
- ...
- 8, 7, 4, 1, 3, 5, 2, 6
- 8, 7, 9, 4, 1, 3, 5, 2, 6.

As illustrated in Figure 2, the label 1 is inserted at the root, the label 2 trickles down to the vertex 1, the label 3 trickles down to the vertex 10, the label 4 trickles down to the vertex 0, and so on until the label 9 trickles down to the vertex 001.

Now let  $(U_n)_{n \in \mathbb{N}}$  be a sequence of independent identically distributed random variables that each have the uniform distribution on the interval  $[0, 1]$ . For each positive integer  $n$  define a uniformly distributed random permutation  $\Pi_n$  of  $[n]$  by requiring that  $\Pi_n(i) < \Pi_n(j)$  if and only if  $U_i < U_j$  for  $1 \leq i, j \leq n$ . That is,

$$\Pi_n(k) = \#\{1 \leq \ell \leq n : U_\ell \leq U_k\}$$

and the corresponding ordered list

$$R_n(k) := \Pi_n^{-1}(k), \quad 1 \leq k \leq n,$$

is such that  $U_{R_n(1)} < U_{R_n(2)} < \dots < U_{R_n(n)}$ . The corresponding ordered list for  $\Pi_{n+1}$  is thus obtained by inserting  $n + 1$  into one of the  $n - 1$  “slots” between the successive

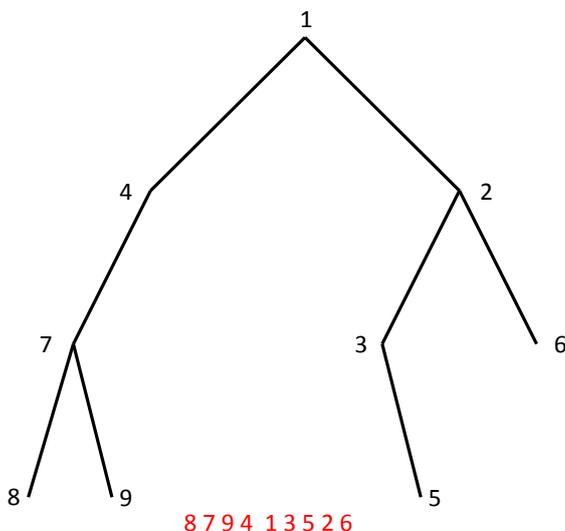


Figure 2: The labeled binary tree corresponding to the permutation of [9] with  $r(1), \dots, r(9) = 8, 7, 9, 4, 1, 3, 5, 2, 6$ . For the sake of clarity, the coding (see Figure 1) of the vertices as elements of  $\{0, 1\}^*$  is not shown. The correspondence between the labeling by the set [9] and the vertices as elements of  $\{0, 1\}^*$  is  $1 \leftrightarrow \emptyset, 2 \leftrightarrow 1, 3 \leftrightarrow 10, 4 \leftrightarrow 0, 5 \leftrightarrow 101, 6 \leftrightarrow 11, 7 \leftrightarrow 00, 8 \leftrightarrow 000, 9 \leftrightarrow 001$ .

elements of the existing list or into one of the two “slots” at the beginning and end of the list, with all  $n + 1$  possibilities being equally likely.

Applying the procedure above for building labeled rooted binary trees to the successive permutations  $\Pi_1, \Pi_2, \dots$  produces a sequence of labeled trees  $(L_n)_{n \in \mathbb{N}}$ , where  $L_n$  has  $n$  vertices labeled by  $[n]$ . This sequence is a Markov chain that evolves as follows. Given  $L_n$ , there are  $n + 1$  words of the form  $v = v_1 \dots v_\ell$  such that  $v$  is not a vertex of the tree  $L_n$  but the word  $v_1 \dots v_{\ell-1}$  is. Pick such a word uniformly at random and adjoin it (with the label  $n + 1$  attached) to produce the labeled tree  $L_{n+1}$ .

If we remove the labels from each tree  $L_n$ , then the resulting random sequence of unlabeled trees is also a Markov chain that has the same distribution as the sequence of trees generated by the BST algorithm when the input stream consists of independent random variables that all have the same continuous distribution function. In essence, at step  $n + 1$  of the BST algorithm there are  $n + 1$  vertices that can be added to the existing tree and the rank of the input value  $x_{n+1}$  within  $x_1, \dots, x_n, x_{n+1}$  determines the choice of this “external vertex”: for i.i.d. continuously distributed random input, this rank is uniformly distributed on  $\{1, \dots, n + 1\}$ , resulting in a uniform pick from the external vertices (see also the discussion following (4.2)). See Figure 3 for an example showing the external vertices of the finite rooted binary tree of Figures 1 and 2.

From now on we will refer to any Markov chain on the space of finite rooted binary trees with this transition mechanism as “the” BST process and denote it by  $(T_n)_{n \in \mathbb{N}}$ .

We note in passing that the labeled permutation trees  $L_1, \dots, L_{n-1}$  can be reconstructed from  $L_n$ , but a similar reconstruction of the history of the process from its current value is not possible if we consider the sequence of labeled trees obtained by labeling the vertices of the tree in the binary search tree algorithm with the input values  $x_1, \dots, x_n$  that created the tree.

Write  $G_n$  (respectively,  $D_n$ ) for the number of vertices in  $T_n$  of the form  $0v_2 \dots v_\ell$

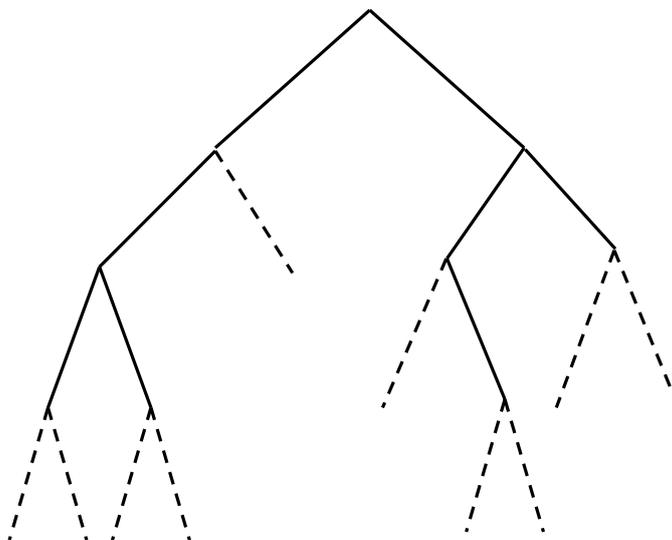


Figure 3: A finite rooted binary tree, the tree with 9 vertices connected by the solid edges, and its 10 external vertices, the vertices connected to the tree by dashed edges. For simplicity, the coding of the vertices as elements of  $\{0, 1\}^*$  is not shown.

(resp.  $1w_2 \dots w_m$ ). That is,  $G_n$  and  $D_n$  are the sizes of the “left” and “right” subtrees in  $T_n$  below the root  $\emptyset$ . Then,  $G_n + 1$  and  $D_n + 1$  are, respectively, the number of “slots” to the left and to the right of 1 in the collection of  $n + 1$  slots between successive elements or at either end of the ordered list  $\Pi_n^{-1}(1), \dots, \Pi_n^{-1}(n)$ . It follows that the sequence of pairs  $(G_n + 1, D_n + 1)$ ,  $n \in \mathbb{N}$ , is itself a Markov chain that evolves as the numbers of black and white balls in a classical Pólya urn (that is, as the process describing the successive compositions of an urn that initially contains one black and one white ball and at each stage a ball is drawn uniformly at random and replaced along with a new ball of the same color). More precisely, conditional on the past up to time  $n$ , if  $(G_n + 1, D_n + 1) = (b, w)$ , then  $(G_{n+1} + 1, D_{n+1} + 1)$  takes the values  $(b + 1, w)$  and  $(b, w + 1)$  with respective conditional probabilities  $\frac{b}{b+w}$  and  $\frac{w}{b+w}$ .

More generally, suppose for a fixed vertex  $u = u_1 \dots u_k \in \{0, 1\}^*$  that we write  $G_n^u$  (respectively,  $D_n^u$ ) for the number of vertices in  $T_n$  of the form  $u_1 \dots u_k 0v_2 \dots v_\ell$  (resp.  $u_1 \dots u_k 1w_2 \dots w_m$ ). That is,  $G_n^u$  and  $D_n^u$  are the sizes of the “left” and “right” subtrees in  $T_n$  below the vertex  $u$ . Put  $C_n^u := \#\{v \in T_n : u \leq v\}$  and  $S_r^u = \inf\{s \in \mathbb{N} : C_s^u = r\}$  for  $r \in \mathbb{N}$ ; that is,  $S_r^u$  is the first time that the subtree of  $T_n$  rooted at  $u$  has  $r$  vertices. Then, the sequence  $(G_{S_r^u}, D_{S_r^u})$ ,  $r \in \mathbb{N}$ , obtained by time-changing the sequence  $(G_n^u, D_n^u)$ ,  $n \in \mathbb{N}$ , so that we only observe it when it changes state is a Markov chain with the same distribution as  $(G_n, D_n)$ ,  $n \in \mathbb{N}$ .

It follows from this observation that we may construct the tree-valued stochastic process  $(T_n)_{n \in \mathbb{N}}$  from an infinite collection of independent, identically distributed Pólya urns, with one urn for each vertex of the complete binary tree  $\{0, 1\}^*$ , by running the urn for each vertex according to a clock that depends on the evolution of the urns associated with vertices that are on the path from the root to the vertex.

More specifically, we first equip each vertex  $u \in \{0, 1\}^*$  with an associated independent  $\mathbb{N}_0 \times \mathbb{N}_0$ -valued *routing instruction* process  $(Y_n^u)_{n \in \mathbb{N}_0}$  such that  $(Y_n^u + (1, 1))_{n \in \mathbb{N}_0}$  evolves like the pair of counts in a Pólya urn with an initial composition of one black and one white ball. Then, at each point in time we feed in a new particle at the root

$\emptyset$ . At time 0 the particle simply comes to rest at  $\emptyset$ . At time 1 the root is occupied and so the particle must be routed to either the vertex 0 or the vertex 1, where it comes to rest, depending on whether the value of  $Y_1^\emptyset$  is  $(1, 0)$  or  $(0, 1)$ . We then continue on in this way: at time  $n \geq 2$  we feed a particle in at the root  $\emptyset$ , it is routed to the vertex 0 or the vertex 1 depending on whether the value of  $Y_n^\emptyset - Y_{n-1}^\emptyset$  is  $(1, 0)$  or  $(0, 1)$ , the particle then trickles down through the tree until it reaches an unoccupied vertex. At each stage of the trickle-down, if the particle is routed to a vertex  $u$  that is already occupied, then it moves on to the vertex  $u0$  or the vertex  $u1$  depending on whether the value of  $Y_{A_n^u}^u - Y_{A_{n-1}^u}^u$  is  $(1, 0)$  or  $(0, 1)$ , where  $A_n^u$  is the number of particles that have passed through vertex  $u$  and been routed onwards by time  $n$ . The resulting sequence of trees is indexed by  $\mathbb{N}_0$  rather than  $\mathbb{N}$ , and if we shift the indices by one we obtain a sequence indexed by  $\mathbb{N}$  that has the same distribution as  $(T_n)_{n \in \mathbb{N}}$ .

It is well-known (see [4]) that the Doob-Martin compactification of the state space  $\mathbb{N}^2$  of the classical Pólya urn results in a Doob-Martin boundary that is homeomorphic to the unit interval  $[0, 1]$ : a sequence of pairs  $((b_n, w_n))_{n \in \mathbb{N}}$  from  $\mathbb{N}^2$  converges to a point in the boundary if and only if  $b_n + w_n \rightarrow \infty$  and  $\frac{w_n}{b_n + w_n} \rightarrow z$  for some  $z \in [0, 1]$ . We can, of course, identify  $[0, 1]$  with the space of probability measures on a set with two points, say  $\{0, 1\}$ , by identifying  $z \in [0, 1]$  with the probability measure that assigns mass  $z$  to the point 1.

It is a consequence of results we prove in Section 4 that this result “lifts” to the binary search tree process: the Doob-Martin boundary is homeomorphic to the space of probability measures on  $\{0, 1\}^\infty$  equipped with the weak topology corresponding to the product topology on  $\{0, 1\}^\infty$  and a sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}}$  of finite rooted binary trees converges to the boundary point identified with the probability measure  $\mu$  if and only if  $\#\mathbf{t}_n \rightarrow \infty$  and for each  $u \in \{0, 1\}^*$

$$\frac{\#\{v \in \mathbf{t}_n : u \leq v\}}{\#\mathbf{t}_n} \rightarrow \mu\{v \in \{0, 1\}^\infty : u \leq v\},$$

where we extend the partial order  $\leq$  on  $\{0, 1\}^*$  to  $\{0, 1\}^* \sqcup \{0, 1\}^\infty$  by declaring that two distinct elements of  $\{0, 1\}^\infty$  are not comparable and  $u \in \{0, 1\}^*$  is dominated by  $v \in \{0, 1\}^\infty$  if  $u$  is a prefix of  $v$ .

An outline of the remainder of the paper is the following. In Section 2 we give a general version of the trickle-down construction in which the complete rooted binary tree  $\{0, 1\}^*$  is expanded to a broad class of directed acyclic graphs with a unique “root” vertex and the independent Pólya urns at each vertex are replaced by independent Markov chains that keep a running total of how many particles have been routed onwards to each of the immediate successors of the vertex. For example, we could take the graph to be  $\mathbb{N}_0^2$  with directed edges of the form  $((i, j), (i + 1, j))$  and  $((i, j), (i, j + 1))$  (so that the root is  $(0, 0)$ ) and take the Markov chain at vertex  $(i, j)$  to correspond to successive particles being routed independently with equal probability to either  $((i, j), (i + 1, j))$  or  $((i, j), (i, j + 1))$ . This gives a process somewhat reminiscent of Sir Francis Galton’s *quincunx* – a device used to illustrate the binomial distribution and central limit theorem in which successive balls are dropped onto a vertical board with interleaved rows of horizontal pins that send a ball striking them downwards to the left or right “at random”. We illustrate the first few steps in the evolution of the set of occupied vertices in Figure 4.

We give a brief overview of the theory of Doob-Martin compactifications in Section 3. We present our main result, a generalization of the facts about the Doob-Martin boundary of the binary search tree process we have stated above, in Section 4. It says for a large class of trickle-down processes that if the convergence of a sequence to a point in the Doob-Martin boundary for each of the component Markov chains is deter-

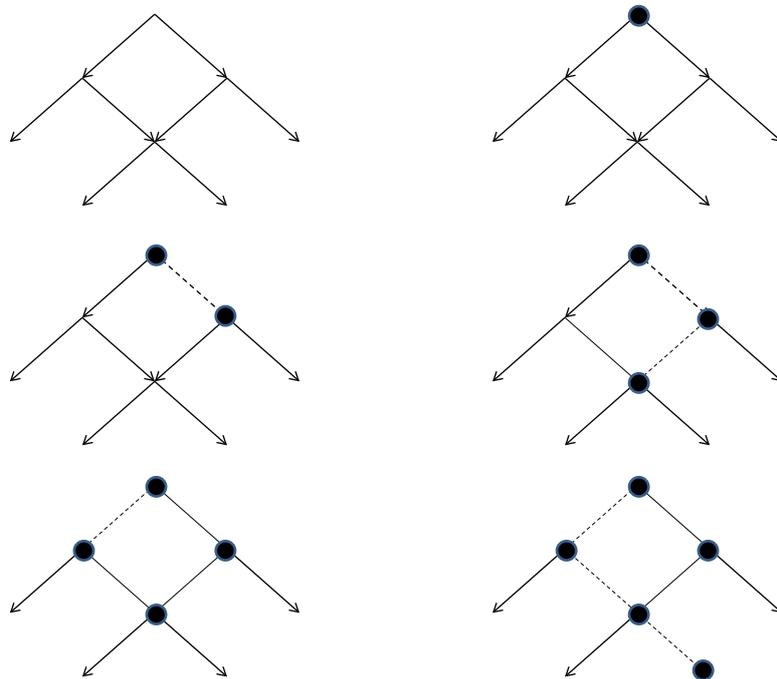


Figure 4: The first five steps in the trickle-down process for the directed acyclic graph  $\mathbb{N}_0^2$  with directed edges of the form  $((i, j), (i + 1, j))$  and  $((i, j), (i, j + 1))$ . The root  $(0, 0)$  is drawn at the top. Dashed lines show that paths taken by successive particles as they pass through occupied vertices until they come to rest at the first unoccupied vertex they encounter.

mined by the convergence of the proportions of points that are routed to each of the immediate successors, then the Doob-Martin boundary of the trickle-down process is homeomorphic to a space of probability measures on a set of directed paths from the root that either have infinite length or are “killed” at some finite time. We then consider special cases of this general result in Section 5, where we investigate the binary and digital search tree processes, and in Section 6, where we study random recursive tree processes that are related to a hierarchy of Chinese restaurant processes.

More specifically, we show in Section 5 that, as we already noted above, the Doob-Martin boundary of the BST process may be identified with the space of probability measures on  $\{0, 1\}^\infty$  equipped with the weak topology corresponding to the product topology on  $\{0, 1\}^\infty$ , that every boundary point is extremal, that the *digital search tree* process is a Doob  $h$ -transform of the BST process with respect to the extremal harmonic function corresponding to the fair coin-tossing measure on  $\{0, 1\}^\infty$ , and that an arbitrary Doob  $h$ -transform may be constructed from a suitable “trickle-up” procedure in which particles come in successively from the “leaves at infinity” of the complete rooted binary tree  $\{0, 1\}^*$  (that is from  $\{0, 1\}^\infty$ ) and work their way up the tree until they can move no further because their path is blocked by an earlier particle.

We observe in Section 6 that the *random recursive tree (RRT) process* – see [50] for a review – can be built from the above sequence  $(\Pi_n)_{n \in \mathbb{N}}$  of uniform permutations in a manner analogous to the construction of the BST process by using a different bijection between permutations and trees. The RRT process is also a trickle-down process similar to the BST process, with the tree  $\{0, 1\}^*$  replaced by the tree  $\mathbb{N}^*$  and the Pólya urn routing instructions replaced by the Markov chain that gives the block sizes in the simplest *Chinese restaurant process* model of growing random partitions. We extend this construction to incorporate Pitman’s two-parameter family of Chinese restaurant processes and then investigate the associated Doob-Martin compactification. We identify the Doob-Martin boundary as a suitable space of probability measures, show that all boundary points are extremal, demonstrate that  $h$ -transform processes may be constructed via a “trickle-up” procedure similar to that described above for the BST process, and relate the limit distribution to the *Griffiths–Engen–McCloskey (GEM) distributions*. Similar nested hierarchies of Chinese restaurant processes appear in [13, 44] and in [52, 3] in the statistical context of mixture models, hierarchical models, and nonparametric Bayesian inference.

A commonly used probability distribution on the set of permutations of a finite set is the Mallows  $\phi$  model – see [39, 8, 18, 14, 7, 40] – for which the uniform distribution is a limiting case. This distribution extends naturally to the set of permutations of  $\mathbb{N}$ , and applying the obvious generalization of the above bijection between finite permutations and labeled finite rooted subtrees of the complete rooted binary tree  $\{0, 1\}^*$  leads to an interesting probability distribution on infinite rooted subtrees of  $\{0, 1\}^*$ . In Section 7 we relate this distribution to yet another model for growing random finite trees that we call the *Mallows tree process*. We show that the Doob-Martin boundary of this Markov chain is a suitable space of infinite rooted subtrees of  $\{0, 1\}^*$ . We outline a parallel analysis in Section 8 for a somewhat similar process that is related to Schützenberger’s non-commutative  $q$ -binomial theorem and its connection to weighted enumerations of “north-east” lattice paths.

The routing instruction processes that appear in the trickle-down construction of the Mallows tree process have the feature that if we know the state of the chain at some time, then we know the whole path of the process up to that time. We observe in Section 9 that such processes may be thought of as Markov chains on a rooted tree with transitions that always go to states that are one step further from the root. As one might expect, the Doob-Martin compactification in this case is homeomorphic to

the usual end compactification of the tree. We use this observation to describe the Doob-Martin compactification of a certain Markov chain that takes values in the set of compositions of the integers and whose value at time  $n$  is uniformly distributed over the compositions of  $n$ .

As we have already remarked, our principal reason for studying the Doob-Martin compactification of a trickle-down chain is to determine the chain's tail  $\sigma$ -field. The Doob-Martin compactification gives even more information about the asymptotic behavior of the chain, but it is not always easy to compute. We describe another approach to determining the tail  $\sigma$ -field of certain trickle-down chains in Section 10. That result applies to the Mallows tree process and the model related to the non-commutative  $q$ -binomial theorem. We also apply it in Section 11 to yet another Markov chain model of growing random trees from [37]. The latter model, which turns out to be of the trickle-down type, has as its state space the set of finite rooted binary trees and is such that if it is started at time 0 in the trivial tree  $\{\emptyset\}$ , then the value of the process at time  $n$  is equally likely to be any of the  $C_n$  rooted binary trees with  $n$  vertices, where  $C_n := \frac{1}{n+1} \binom{2n}{n}$  is the  $n^{\text{th}}$  Catalan number. Even though we cannot determine the Doob-Martin compactification of this chain, we are able to show that its tail  $\sigma$ -field is generated by the random infinite rooted subtree of the complete binary tree that is the (increasing) union of the successive values of the chain. Also, knowing the tail  $\sigma$ -field allows us to identify the Poisson boundary – see Section 3 for a definition of this object.

We observe that there is some similarity between the trickle-down description of the binary search tree process and the *internal diffusion limited aggregation model* that was first named as such in [36] after it was introduced in [11]. There particles are fed successively into a fixed state of some Markov chain and they then execute independent copies of the chain until they come to rest at the first unoccupied state they encounter. The digital search tree process that we discuss in Section 5 turns out to be internal diffusion limited aggregation model for the Markov chain on the complete rooted binary tree that from the state  $u$  moves to the states  $u0$  and  $u1$  with equal probability.

Finally, we note that there are a number of other papers that investigate the Doob-Martin boundary of Markov chains on various combinatorial structures such as Young diagrams and partitions – see, for ex., [43, 34, 19, 24, 21, 20].

## 2 The trickle-down construction

### 2.1 Routing instructions and clocks

We begin by introducing a class of directed graphs with features generalizing those of the complete binary tree  $\{0, 1\}^*$  considered in the Introduction.

Let  $\mathbf{I}$  be a countable directed acyclic graph. With a slight abuse of notation, write  $u \in \mathbf{I}$  to indicate that  $u$  is a vertex of  $\mathbf{I}$ . Given two vertices  $u, v \in \mathbf{I}$ , write  $u \rightarrow v$  if  $(u, v)$  is a directed edge in  $\mathbf{I}$ .

Suppose that there is a unique vertex  $\hat{0}$  such that for any other vertex  $u$  there is at least one finite *directed path*  $\hat{0} = v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_n = u$  from  $\hat{0}$  to  $u$ . Define a *partial order* on  $\mathbf{I}$  by declaring that  $u \leq v$  if  $u = v$  or there is a finite directed path  $u = w_0 \rightarrow w_1 \rightarrow \dots \rightarrow w_n = v$ . Note that  $\hat{0}$  is the unique minimal element of  $\mathbf{I}$ . Suppose further that the number of directed paths between any two vertices is finite: this is equivalent to supposing that the number of directed paths between  $\hat{0}$  and any vertex is finite.

For each vertex  $u \in \mathbf{I}$ , set

$$\alpha(u) := \{v \in \mathbf{I} : v \rightarrow u\}$$

and

$$\beta(u) := \{v \in \mathbf{I} : u \rightarrow v\}.$$

That is,  $\alpha(u)$  and  $\beta(u)$  are, respectively, the immediate predecessors and the immediate successors of  $u$ . Suppose that  $\beta(u)$  is non-empty for all  $u \in \mathbf{I}$ . Thus, any path  $\hat{0} = v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_n = u$  is the initial piece of a semi-infinite path  $v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_n \rightarrow v_{n+1} \rightarrow \dots$ .

We next introduce the notion of *routing instructions* that underlies the construction of a sequence of connected subsets of  $\mathbf{I}$  via a trickle-down mechanism analogous to that described in the Introduction for the BST: at each point in time a particle is fed into  $\hat{0}$  and trickles down through  $\mathbf{I}$  according to the routing instructions at the occupied vertices it encounters until it finds a vacant vertex to occupy.

Let  $(\mathbb{N}_0)^{\beta(u)}$  be the space of functions on the set of successors of  $u \in \mathbf{I}$  that take values in the non-negative integers. Let  $e_v, v \in \beta(u)$ , be the function that takes the value 1 at  $v$  and 0 elsewhere. That is, if we regard  $e_v$  as a vector indexed by  $\beta(u)$ , then  $e_v$  has 1 in the  $v^{\text{th}}$  coordinate and 0 elsewhere. Formally, a routing instruction for the vertex  $u \in \mathbf{I}$  is a sequence  $(\sigma_n^u)_{n \in \mathbb{N}_0}$  of elements of  $(\mathbb{N}_0)^{\beta(u)}$  with the properties:

- $\sigma_0^u = (0, 0, \dots)$ ,
- for each  $n \geq 1$ ,  $\sigma_n^u = \sigma_{n-1}^u + e_{v_n}$  for some  $v_n \in \beta(u)$ .

The interpretation of such a sequence is that, for each  $v \in \beta(u)$ , the component  $(\sigma_n^u)^v$  counts the number of particles out of the first  $n$  to pass through the vertex  $u$  that are routed onwards to vertex  $v \in \beta(u)$ . The equation  $\sigma_n^u = \sigma_{n-1}^u + e_{v_n}$  indicates that the  $n^{\text{th}}$  such particle is routed onwards to the vertex  $v_n \in \beta(u)$ .

For  $s = (s^v)_{v \in \beta(u)} \in (\mathbb{N}_0)^{\beta(u)}$  we put

$$|s| := \sum_{v \in \beta(u)} s^v. \tag{2.1}$$

Note that a routing instruction  $(\sigma_n^u)_{n \in \mathbb{N}_0}$  for the vertex  $u$  satisfies  $|\sigma_n^u| = n$  for all  $n \in \mathbb{N}_0$ .

For each vertex  $u \in \mathbf{I}$ , suppose that we have a non-empty set  $\Sigma^u$  of routing instructions for  $u$ . Put  $\Sigma := \prod_{u \in \mathbf{I}} \Sigma^u$ . Depending on convenience, we write a generic element of  $\Sigma$  in the form  $((\sigma_n^u)_{n \in \mathbb{N}_0})_{u \in \mathbf{I}}$  or the form  $((\sigma^u(n))_{n \in \mathbb{N}_0})_{u \in \mathbf{I}}$ . Recall that  $\sigma_n^u = \sigma^u(n)$  is an element of  $(\mathbb{N}_0)^{\beta(u)}$ , and so it has coordinates  $(\sigma_n^u)^w = (\sigma^u(n))^w$  for  $w \in \beta(u)$ .

Given  $\sigma \in \Sigma$ , each vertex  $u$  of  $\mathbf{I}$  has an associated *clock*  $(a_n^u(\sigma))_{n \in \mathbb{N}_0}$  such that  $a_n^u(\sigma)$  counts the number of particles that have passed through  $u$  by time  $n$  and been routed onwards to some vertex in  $\beta(u)$ . For each  $n \in \mathbb{N}$  and  $\sigma \in \Sigma$  the integers  $a_n^u(\sigma)$ ,  $u \in \mathbf{I}$ , are defined recursively (with respect to the partial order on  $\mathbf{I}$ ) as follows:

- (a)  $a_n^{\hat{0}}(\sigma) := n$ ,
- (b)  $a_n^u(\sigma) := (\sum_{v \in \alpha(u)} (\sigma^v(a_n^v(\sigma)))^u - 1)_+, u \neq \hat{0}$ .

In particular,  $a_0(\sigma) = (0, 0, \dots)$  for all  $\sigma \in \Sigma$ . The equation in (b) simply says that the number of particles that have been routed onwards from the vertex  $u$  by time  $n$  is equal to the number of particles that have passed through vertices  $v$  with  $v \rightarrow u$  and have been routed in the direction of  $u$ , excluding the first particle that reached the vertex  $u$  and occupied it.

We say that the sequence  $(x_n)_{n \in \mathbb{N}_0} = ((x_n^u)_{u \in \mathbf{I}})_{n \in \mathbb{N}_0}$  given by

$$x_n^u := \sigma^u(a_n^u(\sigma)) \tag{2.2}$$

is the *result of the trickle-down construction for the routing instruction*  $\sigma \in \Sigma$ .

**Example 2.1.** Suppose that the directed graph  $\mathbf{I}$  has  $\mathbb{N}_0^2$  as its set of vertices and directed edges of the form  $((i, j), (i + 1, j))$  and  $((i, j), (i, j + 1))$ . The root is  $(0, 0)$ .

(a) Figure 5 shows the state at time  $n = 12$  (that is, the values of  $x_{12}^u$  for  $u = (i, j)$  in  $\mathbf{I} = \mathbb{N}_0^2$ ) generated by routing instructions whose initial pieces are

$$\begin{aligned} \sigma_1^{(0,0)} &= (0, 1), \sigma_2^{(0,0)} = (0, 2), \sigma_3^{(0,0)} = (1, 2), \sigma_4^{(0,0)} = (2, 2), \\ \sigma_1^{(0,1)} &= (0, 1), \sigma_1^{(1,0)} = (1, 0), \sigma_2^{(1,0)} = (2, 0), \sigma_1^{(1,1)} = (0, 1), \end{aligned}$$

when the states  $(i + 1, j)$  and  $(i, j + 1)$  that comprise  $\beta(u)$ , the immediate successors of  $u$ , are taken in that order.

(b) The clock  $a^{(0,1)}$ , which translates from “real time” to the “local time” at the vertex  $(0, 1) \in \mathbf{I} = \mathbb{N}_0^2$  by counting the particles that pass through this vertex, has a corresponding sequence of states that begins  $a_0^{(0,1)} = a_1^{(0,1)} = a_2^{(0,1)} = a_3^{(0,1)} = 0$ ,  $a_4^{(0,1)} = a_5^{(0,1)} = 1$ .

(c) The configuration  $x_5$  consists of a pair  $x_5^u = x_5^{(i,j)} \in \mathbb{N}_0^2$  for every  $u = (i, j) \in \mathbf{I} = \mathbb{N}_0^2$ . Each such pair records the onward routings by time 5 to the immediate successors  $\beta(u) = \{(i + 1, j), (i, j + 1)\}$  of  $u$ . Following through the construction gives  $x_5^{(0,0)} = (2, 2)$ ,  $x_5^{(1,0)} = (0, 1)$ ,  $x_5^{(2,0)} = (2, 0)$ ,  $x_5^{(0,1)} = (0, 1)$ , with all the other components of  $x_5$  being  $(0, 0)$ . For example, the value  $x_5^{(0,1)} = (2, 0)$  indicates that by time 5 the vertex  $(0, 1)$  has been occupied, 2 particles have been sent onwards to the vertex  $(1, 1)$ , and 0 particles have been sent onwards to the other immediate successor  $(0, 2)$ .

(d) Looking at the state  $x_{12}^u$ ,  $u \in \mathbf{I}$ , at time  $n = 12$  we cannot reconstruct the relevant initial segments of the routing instructions but we can see, for example, that

- 13 particles have been fed into the root  $(0, 0)$ : the first of these stayed at the root, 6 of the remainder were routed onwards to  $(1, 0)$  and the other 6 were routed onwards to  $(0, 1)$  (that is,  $a_{12}^{(0,0)}(\sigma) = 12$  and  $\sigma_{12}^{(0,0)} = (6, 6)$ );
- of the 6 particles routed from the root towards  $(1, 0)$ , the first stayed there, 2 of the remainder were routed onwards to  $(2, 0)$  and the other 3 were routed onwards to  $(1, 1)$  (that is,  $a_{12}^{(1,0)}(\sigma) = 5$  and  $\sigma_5^{(1,0)} = (2, 3)$ );
- of the 6 particles routed from the root towards  $(0, 1)$ , the first stayed there, 3 of the remainder were routed onwards to  $(1, 1)$  and the other 2 were routed onwards to  $(0, 2)$  (that is,  $a_{12}^{(0,1)}(\sigma) = 5$  and  $\sigma_5^{(0,1)} = (3, 2)$ ).

For each vertex  $u \in \mathbf{I}$ , write  $\mathbf{S}^u \subseteq (\mathbb{N}_0)^{\beta(u)}$  for the set of vectors that can appear as an entry in an element of  $\Sigma^u$ . That is,  $s \in \mathbf{S}^u$  if and only if  $s = \sigma_m$  for some sequence  $(\sigma_n)_{n \in \mathbb{N}_0} \in \Sigma^u$ , where, of course,  $m = |s|$ . Note that the set  $\mathbf{S}^u$  is countable.

Let  $\mathbf{S}$  denote the subset of  $\prod_{u \in \mathbf{I}} \mathbf{S}^u$  consisting of points  $x = (x^u)_{u \in \mathbf{I}}$  that can be constructed as  $(x^u)_{u \in \mathbf{I}} = (\sigma^u(a_m^u(\sigma)))_{u \in \mathbf{I}}$  for some  $m \in \mathbb{N}_0$  and some  $\sigma = ((\sigma_n^v)_{n \in \mathbb{N}_0})_{v \in \mathbf{I}} \in \Sigma$ ; that is,  $x$  appears as the value at time  $m$  in the result of the trickle-down construction for the routing instruction  $\sigma$ . Clearly, if a sequence  $(x^u)_{u \in \mathbf{I}} \in \prod_{u \in \mathbf{I}} \mathbf{S}^u$  belongs to  $\mathbf{S}$ , then

$$\left( \sum_{v \in \alpha(u)} (x^v)^u - 1 \right)_+ = \sum_{w \in \beta(u)} (x^w)^u. \tag{2.3}$$

Given two points  $x, y \in \mathbf{S}$ , say that  $x \preceq y$  if for some  $m, n \in \mathbb{N}_0$  with  $m \leq n$  and some  $\sigma \in \Sigma$  we have  $x^u = \sigma^u(a_m^u(\sigma))$  and  $y^u = \sigma^u(a_n^u(\sigma))$  for all  $u \in \mathbf{I}$ .

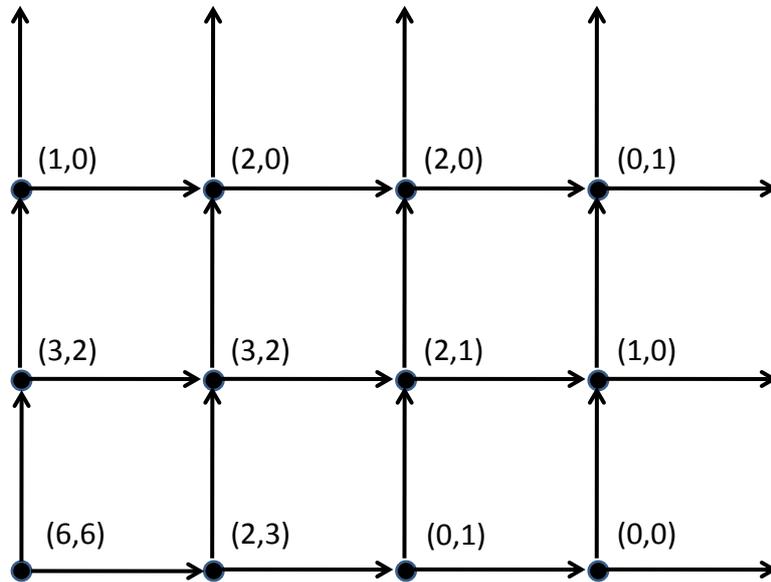


Figure 5: A possible result of the trickle-down construction at time  $n = 12$  on  $\mathbf{I} = \mathbb{N}_0^2$ . See the text for details.

**Remark 2.2.** Note that if  $x \preceq y$ , then  $(x^u)^v \leq (y^u)^v$  for all  $u \in \mathbf{I}$  and  $v \in \beta(u)$ . Moreover, if  $x \preceq y$ , then

$$\begin{aligned} & \left\{ \sigma \in \Sigma : (\sigma^u(a_m^u(\sigma)))_{u \in \mathbf{I}} = x \text{ and } (\sigma^u(a_n^u(\sigma)))_{u \in \mathbf{I}} = y \text{ for some } m \leq n \in \mathbb{N}_0 \right\} \\ &= \left\{ \sigma \in \Sigma : \left( \sigma^u \left( \sum_{v \in \beta(u)} (x^u)^v \right) \right)_{u \in \mathbf{I}} = x \text{ and } \left( \sigma^u \left( \sum_{v \in \beta(u)} (y^u)^v \right) \right)_{u \in \mathbf{I}} = y \right\} \\ &= \prod_{u \in \mathbf{I}} \left\{ \sigma^u \in \Sigma^u : \sigma^u \left( \sum_{v \in \beta(u)} (x^u)^v \right) = x^u \text{ and } \sigma^u \left( \sum_{v \in \beta(u)} (y^u)^v \right) = y^u \right\} \\ &= \prod_{u \in \mathbf{I}} \left\{ \sigma^u \in \Sigma^u : \sigma^u(p) = x^u \text{ and } \sigma^u(q) = y^u \text{ for some } p \leq q \in \mathbb{N}_0 \right\}. \end{aligned}$$

**Example 2.3.** Suppose that  $\mathbf{I}$  is a tree. This amounts to imposing the extra condition that for each vertex  $u \in \mathbf{I}$  there is a **unique** directed path from  $\hat{0}$  to  $u$ . For each  $u \in \mathbf{I}$  take  $\Sigma^u$  to be the set of all allowable routing instructions for  $u$ , so that the corresponding set  $\mathbf{S}^u$  is  $(\mathbb{N}_0)^{\beta(u)}$ . In this case, there is a bijection between  $\mathbf{S}$  and finite subtrees of  $\mathbf{I}$  that contain the root  $\hat{0}$ . An element  $x \in \mathbf{S}$  determines a finite rooted subtree  $\mathbf{t}$  by

$$\mathbf{t} = \{\hat{0}\} \cup \{v \in \mathbf{I} \setminus \{\hat{0}\} : (x^u)^v > 0 \text{ for some } u \in \alpha(v)\}.$$

In other words, the tree  $\mathbf{t}$  consists of those vertices of  $\mathbf{I}$  that are occupied by the first  $\sum_{v \in \beta(\hat{0})} (x^0)^v$  particles.

Conversely, if  $\mathbf{t}$  is a finite subtree of  $\mathbf{I}$  that contains  $\hat{0}$ , then the corresponding element of  $\mathbf{S}$  is

$$x = \left( (\#\{w \in \mathbf{t} : v \leq w\})_{v \in \beta(u)} \right)_{u \in \mathbf{I}};$$

that is,  $x$  appears as the result of the trickle down construction at some time  $n$  and for each pair of vertices  $u \in \mathbf{I}$  and  $v \in \beta(u)$  the integer  $\#\{w \in \mathbf{t} : v \leq w\}$  gives the number of particles that have been routed onwards from vertex  $u \in \mathbf{I}$  to vertex  $v \in \beta(u)$  by time  $n$ . The partial order  $\preceq$  on  $\mathbf{S}$  is equivalent to containment of the associated subtrees. From now on, when  $\mathbf{I}$  is a tree we sometimes do not mention this bijection explicitly and abuse terminology slightly by speaking of  $\mathbf{S}$  as the set of finite subtrees of  $\mathbf{I}$  that contain the root  $\hat{0}$ .

**Example 2.4.** In Example 2.3, the set  $\mathbf{S}^u$  of states for the routing instructions at any vertex  $u \in \mathbf{I}$  is all of  $(\mathbb{N}_0)^{\beta(u)}$ . At the other extreme we have what we call the single trail routing: as always, the first item is put into the root, but now, in the step from  $n$  to  $n+1$ , the new item follows the trail  $u_0, \dots, u_{n-1}$  left by the last one and then chooses  $u_n$  from  $\beta(u_{n-1})$ . In this case,  $\mathbf{S}^u = \{0\} \sqcup \bigsqcup_{v \in \beta(u)} \mathbb{N}e_v$ , where  $0$  is the zero vector in  $(\mathbb{N}_0)^{\beta(u)}$ . Examples of this type appear in Section 9.

**Remark 2.5.** In the setting of Example 2.3, the sequence  $(x_n)_{n \in \mathbb{N}_0}$  in  $\mathbf{S}$  constructed by setting  $x_n^u = \sigma^u(a_n^u(\sigma))$  for some  $\sigma \in \Sigma$  corresponds to a sequence of growing subtrees that begins with the trivial tree  $\{\hat{0}\}$  and successively add a single vertex that is connected by a directed edge to a vertex present in the current subtree, and this correspondence is bijective. In Example 2.4, a sequence  $(x_n)_{n \in \mathbb{N}_0}$  in  $\mathbf{S}$  corresponds to the sequence of initial segments of some infinite directed path,  $\hat{0} = u_0 \rightarrow u_1 \rightarrow u_2 \rightarrow \dots$  through  $\mathbf{I}$ , and this correspondence is also bijective.

## 2.2 Trickle-down chains

We now choose the routing instructions randomly in order to produce an  $\mathbf{S}$ -valued stochastic process.

For each  $u \in \mathbf{I}$ , let  $Q^u$  be a transition matrix whose rows and columns are indexed by some subset  $\mathbf{R}^u \subseteq (\mathbb{N}_0)^{\beta(u)}$  such that  $(0, 0, \dots) \in \mathbf{R}^u$ , and  $Q^u(s', s'') > 0$  for  $s', s'' \in \mathbf{R}^u$  implies that  $s'' = s' + e_v$  for some  $v \in \beta(u)$ . Let  $\Sigma^u$  be the set of sequences  $\sigma^u = (\sigma_n^u)_{n \in \mathbb{N}_0}$  in  $\mathbf{R}^u$  that satisfy  $\sigma_0^u = (0, 0, \dots)$  and  $Q^u(\sigma_n^u, \sigma_{n+1}^u) > 0$  for all  $n \in \mathbb{N}_0$ . Then  $\Sigma^u$  is a set of routing instructions for the vertex  $u$ . Define, as in the previous subsection,  $\mathbf{S}^u$  to be the set of elements of  $(\mathbb{N}_0)^{\beta(u)}$  that can appear as an entry in an element of  $\Sigma^u$ . Note that  $\mathbf{S}^u \subseteq \mathbf{R}^u$ : the set  $\mathbf{S}^u$  consists of the states that are reachable by a Markov chain with transition matrix  $Q^u$  started from the state  $(0, 0, \dots)$ . We will suppose from now on that  $\mathbf{R}^u = \mathbf{S}^u$ .

Write  $(Y_n^u)_{n \in \mathbb{N}_0}$  for the corresponding  $\mathbf{S}^u$ -valued Markov chain with its associated collection of probability measures  $\mathbb{Q}^{u, \xi}$ ,  $\xi \in \mathbf{S}^u$ . A realization of the process  $Y^u$  starting from the zero vector in  $(\mathbb{N}_0)^{\beta(u)}$  will serve as the routing instruction for the vertex  $u$ ; that is, the  $n^{\text{th}}$  particle that trickles down to  $u$  and finds  $u$  occupied will be routed onward to the immediate successor  $v \in \beta(u)$  specified by  $e_v = Y_n^u - Y_{n-1}^u$ . By assumption, and with  $0$  the zero vector in  $(\mathbb{N}_0)^{\beta(u)}$ ,  $Y^u$  has positive probability under  $\mathbb{Q}^{u, 0}$  of hitting any given state in  $\mathbf{S}^u$ . We will refer to  $Y^u$  as the *routing chain for the vertex  $u$* . Let

$$Y := (Y^u)_{u \in \mathbf{I}},$$

where the component processes  $Y^u$  are independent and have distribution  $\mathbb{Q}^{u, 0}$ .

With  $a_0, a_1, \dots$  the clocks defined in Section 2.1, set

$$A_n := \begin{cases} a_n(Y), & \text{if } Y_0 = (0, 0, \dots), \\ 0, & \text{otherwise.} \end{cases}$$

Thus,  $(A_n)_{n \in \mathbb{N}_0}$  is an  $(\mathbb{N}_0)^{\mathbf{I}}$ -valued stochastic process with non-decreasing paths and initial value  $(0, 0, \dots)$ . When  $Y_0 = (0, 0, \dots)$ , the value of the process  $A_n$  at time  $n$  is a

vector  $(A_n^u)_{u \in \mathbf{I}}$ : the non-negative integer  $A_n^u$  records the number of particles that have trickled down to the vertex  $u$  by time  $n$ , found  $u$  already occupied, and have been routed onwards.

Define

$$Z_n^u := Y_{A_n^u}^u, \quad u \in \mathbf{I}, n \in \mathbb{N}_0.$$

By construction,  $Z := (Z_n)_{n \in \mathbb{N}_0} = ((Z_n^u)_{u \in \mathbf{I}})_{n \in \mathbb{N}_0}$  is a Markov chain on the countable state space  $\mathbf{S}$  under the probability measure  $\bigotimes_{u \in \mathbf{I}} \mathbb{Q}^{u,0}$ . The paths of  $Z$  start from the state  $(0, 0, \dots)$  and increase strictly in the natural partial order on  $\mathbf{S}$ . The random vector  $Z_n^u$  gives for each immediate successor  $v \in \beta(u)$  of  $u$  the number of particles that have trickled down to  $u$  by time  $n$ , found  $u$  already occupied, and have been routed onwards towards  $v$ .

By standard arguments, we can construct a measurable space  $(\Omega, \mathcal{F})$ , a family of probability measures  $(\mathbb{P}^x)_{x \in \mathbf{S}}$  and an  $\mathbf{S}$ -valued stochastic process  $X = (X_n)_{n \in \mathbb{N}_0}$  such that  $X$  under  $\mathbb{P}^x$  is a Markov chain with  $X_0 = x$  and the same transition mechanism as  $Z$ .

**Remark 2.6.** Note that if  $\mathbf{J}$  is a subset of  $\mathbf{I}$  with the property that  $\{v \in \mathbf{I} : v \leq u\} \subseteq \mathbf{J}$  for all  $u \in \mathbf{J}$ , then  $((X_n^u)_{u \in \mathbf{J}})_{n \in \mathbb{N}_0}$  is a Markov chain under  $\mathbb{P}^x$ . Moreover, the law of the latter process under  $\mathbb{P}^x$  agrees with its law under  $\mathbb{P}^y$  for any  $y \in \mathbf{S}$  with  $x^u = y^u$  for all  $u \in \mathbf{J}$ .

### 3 Doob-Martin compactification background

We restrict the following sketch of Doob-Martin compactification theory for discrete time Markov chains to the situation of interest in the present paper. The primary reference is [16], but useful reviews may be found in [35, Chapter 10], [45, Chapter 7], [47], [55, Chapter IV], [46, Chapter III].

Suppose that  $(X_n)_{n \in \mathbb{N}_0}$  is a discrete time Markov chain with countable state space  $E$  and transition matrix  $P$ . Define the *Green kernel* or *potential kernel*  $G$  of  $P$  by  $G(i, j) := \sum_{n=0}^{\infty} P^n(i, j)$  for  $i, j \in E$  and assume that there is a reference state  $e \in E$  such that  $0 < G(e, j) < \infty$  for all  $j \in E$ . This implies that any state can be reached from  $e$  and that every state is transient. For the chains to which we apply the theory, the state space  $E$  is a partially ordered set with unique minimal element  $e$  and transition matrix  $P$  such that  $P(k, \ell) = 0$  unless  $k < \ell$ , so that the sample paths of the chain are *increasing* and

$$G(i, j) = \mathbb{P}^i\{X_n = j \text{ for some } n \in \mathbb{N}_0\} =: \mathbb{P}^i\{X \text{ hits } j\}$$

for all  $i, j \in E$ .

A function  $f : E \rightarrow \mathbb{R}_+$  is said to be *excessive* (respectively, *regular*) if

$$\sum_{j \in E} P(i, j)f(j) =: Pf(i) \leq f(i)$$

for all  $i \in E$  (respectively,  $Pf(i) = f(i)$  for all  $i \in E$ ). Excessive functions are also called *non-negative superharmonic functions*. Similarly, regular functions are also called *non-negative harmonic functions*. Given a finite measure  $\mu$  on  $E$ , define a function  $G\mu : E \rightarrow \mathbb{R}_+$  by  $G\mu(i) := \sum_{j \in E} G(i, j)\mu(\{j\})$  for  $i \in E$ . The function  $G\mu$  is excessive and is called the *potential* of the measure  $\mu$ . The Riesz decomposition says that any excessive function  $f$  has a unique decomposition  $f = h + p$ , where  $h$  is regular and  $p = G\nu$  is the potential of a unique measure  $\nu$ .

Note for any excessive function  $f$  that  $f(e) \geq \sup_{n \in \mathbb{N}_0} P^n(e, j)f(j)$ , and so  $f(e) = 0$  implies that  $f = 0$ . Therefore, any excessive function is a constant multiple of an

element of the set  $S$  of excessive functions that take the value 1 at  $e$ . The set  $S$  is a compact convex metrizable subset of the locally convex topological vector space  $\mathbb{R}^E$ .

The *Martin kernel with reference state  $e$*  is given by

$$K(i, j) := \frac{G(i, j)}{G(e, j)} = \frac{\mathbb{P}^i\{X \text{ hits } j\}}{\mathbb{P}^e\{X \text{ hits } j\}};$$

that is,  $K(\cdot, j)$  is the potential of the unit point mass at  $j$  normalized to have value 1 at the point  $e \in E$ . For each  $j \in E$  the function  $K(\cdot, j)$  belongs to  $S$  and is non-regular. Moreover,  $K(\cdot, j)$  is an extreme point of  $S$  and any extreme point of  $S$  that is not of the form  $K(\cdot, j)$  for some  $j \in E$  is regular. It also follows from the Riesz decomposition that the map  $\phi : E \rightarrow S$  given by  $\phi(j) := K(\cdot, j)$  is injective. Therefore, we can identify  $E$  with its image  $\phi(E) \subset S$  that sits densely inside the compact closure  $F$  of  $\phi(E)$  in  $S$ . With the usual slight abuse of terminology, we treat  $E$  as a subset of  $F$  and use the alternative notation  $\bar{E}$  for  $F$ . The construction of the compact metrizable space  $\bar{E}$  from  $E$  using the transition matrix  $P$  and the reference state  $e$  is the *Doob-Martin compactification* of  $E$  and the set

$$\partial E := \bar{E} \setminus E$$

is the *Doob-Martin boundary* of  $E$ .

By definition, a sequence  $(j_n)_{n \in \mathbb{N}}$  in  $E$  converges to a point in  $\bar{E}$  if and only if the sequence of real numbers  $(K(i, j_n))_{n \in \mathbb{N}}$  converges for all  $i \in E$ . Each function  $K(i, \cdot)$  extends continuously to  $\bar{E}$  and we call the resulting function  $K : E \times \bar{E} \rightarrow \mathbb{R}$  the *extended Martin kernel*.

The set of extreme points  $F_{\text{ex}}$  of the convex set  $F$  is a  $G_\delta$  subset of  $F$  and any regular function  $h \in S$  (that is, any regular function  $h$  with  $h(e) = 1$ ) has the representation

$$h = \int K(\cdot, y) \mu(dy)$$

for some unique probability measure on  $F$  that assigns all of its mass to  $F_{\text{ex}} \cap E^c \subseteq \partial E$ .

The primary probabilistic consequence of the Doob-Martin compactification is that for any initial state  $i$  the limit  $X_\infty := \lim_{n \rightarrow \infty} X_n$  exists  $\mathbb{P}^i$ -almost surely in the topology of  $F$  and the limit belongs to  $F_{\text{ex}}$ ,  $\mathbb{P}^i$ -almost surely.

If  $h$  is a regular function (not identical to 0), then the corresponding *Doob  $h$ -transform* is the Markov chain  $(X_n^{(h)})_{n \in \mathbb{N}_0}$  with state space  $E^h := \{i \in E : h(i) > 0\}$  and transition matrix

$$P^{(h)}(i, j) := h(i)^{-1} P(i, j) h(j), \quad i, j \in E^h.$$

When  $h$  is strictly positive, the Doob-Martin compactification of  $E$  and its set of extreme points are the same for  $P$  and  $P^{(h)}$ .

The regular function  $h$  is extremal if and only if the limit  $\lim_{n \rightarrow \infty} X_n^h$  is almost surely equal to a single point  $y$  for some  $y \in F$ , in which case  $y \in F_{\text{ex}} \cap E^c$  and  $h = K(\cdot, y)$ . In particular,  $h$  is extremal if and only if the tail  $\sigma$ -field of  $(X_n^{(h)})_{n \in \mathbb{N}_0}$  is trivial. In this case, the transformed chain  $(X_n^{(h)})_{n \in \mathbb{N}_0}$  may be thought of as the original chain  $(X_n)_{n \in \mathbb{N}_0}$  conditioned to converge to  $y$ . The original chain is a mixture of such conditioned chains, where the mixing measure is the unique probability measure  $\nu$  supported on

$$F_{\text{ex}} \cap E^c \subseteq \partial E$$

such that  $1 = \int K(\cdot, y) \nu(dy)$ . Further,  $\nu$  is the distribution of  $X_\infty$  under  $\mathbb{P}^e$ .

The Doob-Martin boundary provides a representation of the non-negative harmonic functions. We close this review section with a brief discussion of a measure theoretic boundary concept that has a more direct relation to tail  $\sigma$ -fields in the trickle-down case.

The set  $\mathbb{H}$  of all *bounded* harmonic functions is a linear space and indeed a Banach space when endowed with the supremum norm. The *Poisson boundary* is a measure space  $(M, \mathcal{A}, \mu)$  with the property that  $L^\infty(M, \mathcal{A}, \mu)$  and  $\mathbb{H}$  are isomorphic as Banach spaces. The Doob-Martin boundary  $\partial E$  together with its Borel  $\sigma$ -field and the distribution  $\nu$  of  $X_\infty$  under  $\mathbb{P}^e$  provides such a measure space.

Our models have the specific feature that, loosely speaking, ‘time is a function of space’: the state space  $E$  of a trickle-down chain  $(X_n)_{n \in \mathbb{N}_0}$  may be written as the disjoint union of the sets

$$E_n := \{x \in E : \mathbb{P}^e\{X_n = x\} > 0\}.$$

Let  $\mathcal{T}$  be the tail  $\sigma$ -field of the chain. Consider now the map that takes a bounded,  $\mathcal{T}$ -measurable random variable  $Z$  to the function  $h : E \rightarrow \mathbb{R}$  defined by

$$h(x) := \frac{1}{\mathbb{P}^e\{X_n = x\}} \int_{\{X_n = x\}} Z d\mathbb{P}^e,$$

for all  $x \in E_n$ , on each  $E_n$  separately. Note that  $h(X_n) = \mathbb{E}^e[Z|X_n]$ . Using martingale convergence and the Markov property, it follows that this map is a Banach space isomorphism between  $L^\infty(\Omega, \mathcal{T}, \mathbb{P}^e)$  and  $\mathbb{H}$ .

For any embedding in which the chain converges to a limit  $X_\infty$ , this limit is  $\mathcal{T}$ -measurable. The limit in the Doob-Martin compactification of a transient chain generates the invariant  $\sigma$ -field up to null sets, where for a chain  $(X_n)_{n \in \mathbb{N}_0}$  with state space  $E$ , an event  $A$  is invariant if there is a product measurable subset  $B \subseteq E^{\mathbb{N}_0}$  such that for all  $n \in \mathbb{N}_0$  the symmetric difference  $A \Delta \{(X_n, X_{n+1}, \dots) \in B\}$  has zero probability. In our models, the limit  $X_\infty$  in the Doob-Martin compactification generates the tail  $\sigma$ -field, because it is possible to reconstruct the value of the time parameter from the state of the process at an unspecified time. Conversely, from the tail  $\sigma$ -field we may obtain the Poisson boundary but not, in general, the Doob-Martin boundary.

#### 4 Compactification for trickle-down processes

For each  $u \in \mathbf{I}$ , let  $Q^u$  be a transition matrix on  $\mathbf{S}^u \subseteq \mathbb{N}_0^{\beta(u)}$  with the properties described in Section 2.2. The following result is immediate from the construction of the trickle-down chain  $X$  and Remark 2.2.

**Lemma 4.1.** *Consider elements  $x = (x^u)_{u \in \mathbf{I}}$  and  $y = (y^u)_{u \in \mathbf{I}}$  of  $\mathbf{S}$ . Write*

$$m^u = \sum_{v \in \beta(u)} (x^u)^v \quad \text{and} \quad n^u = \sum_{v \in \beta(u)} (y^u)^v.$$

Then,

$$\mathbb{P}^x\{X \text{ hits } y\} = \prod_{u \in \mathbf{I}} Q^{u, x^u} \{Y_{n^u - m^u}^u = y^u\} = \prod_{u \in \mathbf{I}} Q^{u, x^u} \{Y^u \text{ hits } y^u\}.$$

The product is zero unless  $x \preceq y$  (equivalently,  $x^u \leq y^u$  for all  $u \in \mathbf{I}$ ). Only finitely many terms in the product differ from 1, because  $x^u = y^u = (0, 0, \dots)$  (equivalently,  $m^u = n^u = 0$ ) for all but finitely many values of  $u \in \mathbf{I}$ .

**Corollary 4.2.** *The Martin kernel of the Markov chain  $X$  with respect to the reference state  $\hat{0}$  is given by*

$$K(x, y) = \prod_{u \in \mathbf{I}} K^u(x^u, y^u),$$

where  $K^u$  is the Martin kernel of the Markov chain  $Y^u$  with respect to reference state  $(0, 0, \dots) \in \mathbf{S}^u$ . The product is zero unless  $x \preceq y$  (equivalently,  $x^u \leq y^u$  for all  $u \in \mathbf{I}$ ). Only finitely many terms in the product differ from 1, because  $x^u = (0, 0, \dots)$  for all but finitely many values of  $u \in \mathbf{I}$ .

*Proof.* It suffices to note that

$$K(x, y) = \frac{\mathbb{P}^x\{X \text{ hits } y\}}{\mathbb{P}^{\hat{0}}\{X \text{ hits } y\}}$$

and

$$K^u(\xi, \zeta) = \frac{\mathbb{Q}^{u,\xi}\{Y^u \text{ hits } \zeta\}}{\mathbb{Q}^{u,0}\{Y^u \text{ hits } \zeta\}},$$

and then apply Lemma 4.1. □

**Example 4.3.** Consider the BST process from the Introduction. Recall that in this case the directed graph  $\mathbf{I}$  is the complete binary tree  $\{0, 1\}^*$  and each of the processes

$$(Y_n^u + (1, 1))_{n \in \mathbb{N}_0}$$

is the classical Pólya urn in which we have an urn consisting of black and white balls, we draw a ball uniformly at random at each step and replace it along with one of the same color, and we record the number of black and white balls present in the urn at each step. Note that if we start the Pólya urn with  $b$  black and  $w$  white balls, then the probability that we ever see  $B$  black balls and  $W$  white balls is the probability that after  $(B + W) - (b + w)$  steps we have added  $B - b$  black balls and  $W - w$  white balls. The probability of adding the extra balls in a particular specified order is

$$\frac{b(b+1) \cdots (B-1)w(w+1) \cdots (W-1)}{(b+w)(b+w+1) \cdots (B+W-1)}$$

(the fact that this probability is the same for all orders is the fundamental exchangeability fact regarding the Pólya urn). The probability of adding the required extra balls of each color in some order is therefore

$$\frac{((B+W) - (b+w))! b(b+1) \cdots (B-1)w(w+1) \cdots (W-1)}{(B-b)!(W-w)! (b+w)(b+w+1) \cdots (B+W-1)}.$$

Hence,

$$\begin{aligned} & \mathbb{Q}^{u,\xi}\{Y^u \text{ hits } \zeta\} \\ &= \frac{((\zeta^{u0} + \zeta^{u1}) - (\xi^{u0} + \xi^{u1}))!}{(\zeta^{u0} - \xi^{u0})!(\zeta^{u1} - \xi^{u1})!} \frac{(\xi^{u0} + 1) \cdots \zeta^{u0} \times (\xi^{u1} + 1) \cdots \zeta^{u1}}{(\xi^{u0} + \xi^{u1} + 2)(\xi^{u0} + \xi^{u1} + 1) \cdots (\zeta^{u0} + \zeta^{u1} + 1)} \end{aligned}$$

for  $\xi \leq \zeta$ , and so

$$\begin{aligned} K^u(\xi, \zeta) &= \frac{(\xi^{u0} + \xi^{u1} + 1)! (\zeta^{u0} - \xi^{u0} + 1) \cdots \zeta^{u0} \times (\zeta^{u1} - \xi^{u1} + 1) \cdots \zeta^{u1}}{\xi^{u0}! \xi^{u1}! ((\zeta^{u0} + \zeta^{u1}) - (\xi^{u0} + \xi^{u1}) + 1) \cdots (\zeta^{u0} + \zeta^{u1})} \\ &= \frac{(\xi^{u0} + \xi^{u1} + 1)!}{\xi^{u0}! \xi^{u1}!} \frac{\zeta^{u0}! \zeta^{u1}!}{(\zeta^{u0} + \zeta^{u1} + 1)!} \\ &\quad \times \frac{((\zeta^{u0} + \zeta^{u1}) - (\xi^{u0} + \xi^{u1}))!}{(\zeta^{u0} - \xi^{u0})!(\zeta^{u1} - \xi^{u1})!} (\zeta^{u0} + \zeta^{u1} + 1). \end{aligned}$$

Suppose that  $x, y \in \mathbf{S}$  with  $x \preceq y$ . It follows from Corollary 4.2 that

$$\begin{aligned} K(x, y) &= \prod_{u \in \mathbf{I}} \frac{((x^u)^{u0} + (x^u)^{u1} + 1)!}{(x^u)^{u0}! (x^u)^{u1}!} \frac{(y^u)^{u0}! (y^u)^{u1}!}{((y^u)^{u0} + (y^u)^{u1} + 1)!} \\ &\quad \times \frac{(((y^u)^{u0} + (y^u)^{u1}) - ((x^u)^{u0} + (x^u)^{u1}))!}{((y^u)^{u0} - (x^u)^{u0})! ((y^u)^{u1} - (x^u)^{u1})!} ((y^u)^{u0} + (y^u)^{u1} + 1). \end{aligned}$$

Recall from Example 2.3 that we may associate  $x$  and  $y$  with the two subtrees

$$\mathbf{s} = \{\emptyset\} \cup \{v \in \mathbf{I} : (x^u)^v > 0 \text{ for the unique } u \in \alpha(v)\}$$

and

$$\mathbf{t} = \{\emptyset\} \cup \{v \in \mathbf{I} : (y^u)^v > 0 \text{ for some the unique } u \in \alpha(v)\},$$

in which case  $(x^u)^v = \#\{w \in \mathbf{s} : v \leq w\} =: \#\mathbf{s}(v)$  for  $v \in \mathbf{s} \setminus \{\emptyset\}$  and  $u \in \alpha(v)$  (respectively,  $(y^u)^v = \#\{w \in \mathbf{t} : v \leq w\} =: \#\mathbf{t}(v)$  for  $v \in \mathbf{t} \setminus \{\emptyset\}$  and  $u \in \alpha(v)$ ). Note for  $\varepsilon = 0, 1$  that

$$(x^u)^{u\varepsilon} = \begin{cases} \#\mathbf{s}(u\varepsilon), & \text{if } u \in \mathbf{s}, \\ 0, & \text{otherwise,} \end{cases}$$

and that

$$(x^u)^{u0} + (x^u)^{u1} + 1 = \begin{cases} \#\mathbf{s}(u), & \text{if } u \in \mathbf{s}, \\ 1, & \text{otherwise.} \end{cases}$$

Similar relations exist for  $y$  and  $\mathbf{t}$ . It follows that

$$\begin{aligned} \prod_{u \in \mathbf{I}} \frac{((x^u)^{u0} + (x^u)^{u1} + 1)!}{(x^u)^{u0}!(x^u)^{u1}!} &= \#\mathbf{s}!, \\ \prod_{u \in \mathbf{I}} \frac{(y^u)^{u0}!(y^u)^{u1}!}{((y^u)^{u0} + (y^u)^{u1} + 1)!} &= \frac{1}{\#\mathbf{t}!}, \\ \prod_{u \in \mathbf{I}} \frac{(((y^u)^{u0} + (y^u)^{u1}) - ((x^u)^{u0} + (x^u)^{u1}))!}{((y^u)^{u0} - (x^u)^{u0})!((y^u)^{u1} - (x^u)^{u1})!} &= \frac{(\#\mathbf{t} - \#\mathbf{s})!}{\prod_{u \in \mathbf{t} \setminus \mathbf{s}} \#\mathbf{t}(u)}, \end{aligned}$$

and

$$\prod_{u \in \mathbf{I}} ((y^u)^{u0} + (y^u)^{u1} + 1) = \prod_{u \in \mathbf{t}} \#\mathbf{t}(u),$$

so we arrive at the simple formula

$$K(x, y) = \left(\frac{\#\mathbf{t}}{\#\mathbf{s}}\right)^{-1} \prod_{u \in \mathbf{s}} \#\mathbf{t}(u). \tag{4.1}$$

This formula may also be obtained without using Corollary 4.2 as follows. With a slight abuse of notation, we think of the process  $(X_n)_{n \in \mathbb{N}_0}$  as taking values in the set of finite subtrees of  $\{0, 1\}^*$  containing the root  $\emptyset$ . We first want a formula for  $\mathbb{P}^{\mathbf{s}}\{X \text{ hits } \mathbf{t}\}$  when  $\mathbf{s}$  and  $\mathbf{t}$  are two such trees with  $\mathbf{s} \subseteq \mathbf{t}$ . For ease of notation, set  $k := \#\mathbf{s}$  and  $n := \#\mathbf{t}$ . It is known (see, for example, [49, p.316]) that

$$\mathbb{P}^{\{\emptyset\}}\{X \text{ hits } \mathbf{t}\} = \mathbb{P}^{\{\emptyset\}}\{X_n = \mathbf{t}\} = \prod_{u \in \mathbf{t}} (\#\mathbf{t}(u))^{-1}, \tag{4.2}$$

Write  $v_1, \dots, v_{k+1}$  for the “external vertices” of  $\mathbf{s}$ ; that is, the elements of  $\{0, 1\}^*$  that are connected to a vertex of  $\mathbf{s}$  by a directed edge, but are not vertices of  $\mathbf{s}$  themselves (recall Figure 3). Denote by  $\mathbf{t}(v_j)$ ,  $j = 1, \dots, k + 1$  the subtrees of  $\mathbf{t}$  that are rooted at these vertices; that is, the  $\mathbf{t}(v_j)$  are the connected components of  $\mathbf{t} \setminus \mathbf{s}$ . In order for the BST process to pass from  $\mathbf{s}$  to  $\mathbf{t}$  it needs to place the correct number  $n_j := \#\mathbf{t}(v_j)$  of vertices into each of these subtrees and, moreover, the subtrees have to be equal to  $\mathbf{t}(v_j)$ , for  $j = 1, \dots, k + 1$ . The process that tracks the number of vertices in each subtree is, after we add the vector  $(1, \dots, 1)$ , a multivariate Pólya urn model starting with  $k + 1$  balls, all

of different colors. Thus, the probability that each subtree has the correct number of vertices is

$$\binom{n-k}{n_1, \dots, n_{k+1}} \frac{\prod_{i=1}^{k+1} n_i!}{(k+1) \cdots (n-1) \cdot n} = \left( \frac{\#\mathbf{t}}{\#\mathbf{s}} \right)^{-1},$$

using a standard argument for the Pólya urn [31, Chapter 4.5]. Moreover, it is apparent from the recursive structure of the BST process that, conditional on  $k+1$  subtrees receiving the correct number of vertices, the probability the subtrees are actually  $\mathbf{t}(v_1), \dots, \mathbf{t}(v_{k+1})$  is

$$\prod_{i=1}^{k+1} \prod_{v \in \mathbf{t}(v_i)} (\#\mathbf{t}(v))^{-1} = \prod_{v \in \mathbf{t} \setminus \mathbf{s}} (\#\mathbf{t}(v))^{-1}.$$

Thus,

$$\mathbb{P}^{\mathbf{s}}\{X \text{ hits } \mathbf{t}\} = \left( \frac{\#\mathbf{t}}{\#\mathbf{s}} \right)^{-1} \prod_{v \in \mathbf{t} \setminus \mathbf{s}} (\#\mathbf{t}(v))^{-1}, \tag{4.3}$$

and (4.1) follows upon taking the appropriate ratio.

With Example 4.3 in mind, we now begin to build a general framework for characterizing the Doob-Martin compactification of a trickle-down chain in terms of the compactifications of each of the routing chains.

**Proposition 4.4.** *Suppose  $(y_n)_{n \in \mathbb{N}_0}$  is a sequence in  $\mathbf{S}$  such that  $y_\infty^u := \lim_{n \rightarrow \infty} y_n^u$  exists in the Doob-Martin topology of  $\bar{\mathbf{S}}^u$  for each  $u \in \mathbf{I}$ . Then,  $(y_n)_{n \in \mathbb{N}_0}$  converges in the Doob-Martin topology of  $\mathbf{S}$  to a limit  $y_\infty$  and the value at  $(x, y_\infty)$  of the extended Martin kernel is  $K(x, y_\infty) = \prod_{u \in \mathbf{I}} K^u(x^u, y_\infty^u)$ .*

*Proof.* The assumption that  $y_\infty^u := \lim_{n \rightarrow \infty} y_n^u$  exists in the Doob-Martin topology of  $\bar{\mathbf{S}}^u$  for each  $u \in \mathbf{I}$  implies that  $\lim_{n \rightarrow \infty} K^u(\xi, y_n^u)$  exists for each  $u \in \mathbf{I}$  and  $\xi \in \mathbf{S}^u$ . This limit is, by definition, the value  $K^u(\xi, y_\infty^u)$  of the extended Martin kernel. We need to show for all  $x \in \mathbf{S}$  that  $\lim_{n \rightarrow \infty} K(x, y_n)$  exists and is given by  $\prod_{u \in \mathbf{I}} K^u(x^u, y_\infty^u)$ . It follows from Corollary 4.2 that  $K(x, y_n) = \prod_{u \in \mathbf{I}} K^u(x^u, y_n^u)$ . We also know from that result that we may restrict the product to the fixed, finite set of  $u$  for which  $x^u \neq (0, 0, \dots)$ , and hence we may interchange the limit and the product.  $\square$

**Remark 4.5.** *Proposition 4.4 shows that if the sequence  $(y_n)_{n \in \mathbb{N}_0}$  in  $\mathbf{S}$  is such that for each  $u \in \mathbf{I}$  the component sequence  $(y_n^u)_{n \in \mathbb{N}_0}$  converges in the Doob-Martin compactification of  $\mathbf{S}^u$ , then  $(y_n)_{n \in \mathbb{N}_0}$  converges in the Doob-Martin compactification of  $\mathbf{S}$ .*

*Establishing results in the converse direction is somewhat tricky, since  $K(x, y_n) = \prod_{u \in \mathbf{I}} K^u(x^u, y_n^u)$  might converge because  $K^v(x^v, y_n^v)$  converges to 0 for some particular  $v \in \mathbf{I}$ , and so we are not able to conclude that  $K^u(x^u, y_n^u)$  converges for all  $u \in \mathbf{I}$ . Instances of this possibility appear in Section 7 and Section 8.*

The following set of hypotheses gives one quite general setting in which it is possible to characterize the Doob-Martin compactification of  $\mathbf{S}$  in terms of the compactifications of the component spaces  $\mathbf{S}^u$ . These hypotheses are satisfied by a number of interesting examples such as the binary search tree and the random recursive tree processes (see Example 4.7 and Example 4.8 below as well as Section 5 and Section 6). The key condition is part (iii) of the following set of hypotheses: it requires that the Doob-Martin boundary of the routing chain for the vertex  $u$  may be thought of as a set of subprobability measures on  $\beta(u)$  that arise as the vector of limiting proportions of particles that have been routed onward to the various elements of  $\beta(u)$ .

**Hypothesis 4.6.** *Suppose that the following hold for all  $u \in \mathbf{I}$ .*

- (i) Writing  $|\xi| = \sum_{v \in \beta(u)} \xi^v$  for  $\xi \in \mathbf{S}^u$ , the sets  $\{\xi \in \mathbf{S}^u : |\xi| = m\}$  are finite for all  $m \in \mathbb{N}_0$ , so that if  $(\zeta_n)_{n \in \mathbb{N}_0}$  is a sequence from  $\mathbf{S}^u$ , then the two conditions

$$\#\{n \in \mathbb{N}_0 : \zeta_n = \zeta\} < \infty \text{ for all } \zeta \in \mathbf{S}^u \tag{4.4}$$

and

$$\lim_{n \rightarrow \infty} |\zeta_n| = \infty \tag{4.5}$$

are equivalent.

- (ii) In order that a sequence  $(\zeta_n)_{n \in \mathbb{N}_0}$  from  $\mathbf{S}^u$  is such that  $K^u(\xi, \zeta_n)$  converges as  $n \rightarrow \infty$  for all  $\xi \in \mathbf{S}^u$ , it is necessary and sufficient that either

$$\#\{n \in \mathbb{N}_0 : \zeta_n \neq \zeta\} < \infty \text{ for some } \zeta \in \mathbf{S}^u$$

or that the equivalent conditions (4.4) and (4.5) hold and, in addition,

$$\lim_{n \rightarrow \infty} \frac{\zeta_n^v}{|\zeta_n|} \text{ exists for all } v \in \beta(u). \tag{4.6}$$

- (iii) If  $(\zeta'_n)_{n \in \mathbb{N}_0}$  and  $(\zeta''_n)_{n \in \mathbb{N}_0}$  are two sequences from  $\mathbf{S}^u$  such that

$$\#\{n \in \mathbb{N}_0 : \zeta'_n = \zeta\} < \infty \text{ and } \#\{n \in \mathbb{N}_0 : \zeta''_n = \zeta\} < \infty$$

for all  $\zeta \in \mathbf{S}^u$  and both  $K^u(\xi, \zeta'_n)$  and  $K^u(\xi, \zeta''_n)$  converge for all  $\xi \in \mathbf{S}^u$ , then

$$\lim_{n \rightarrow \infty} K^u(\xi, \zeta'_n) = \lim_{n \rightarrow \infty} K^u(\xi, \zeta''_n)$$

for all  $\xi \in \mathbf{S}^u$  if and only if

$$\lim_{n \rightarrow \infty} \frac{(\zeta'_n)^v}{|\zeta'_n|} = \lim_{n \rightarrow \infty} \frac{(\zeta''_n)^v}{|\zeta''_n|}$$

for all  $v \in \beta(u)$ . It follows that there is a natural bijection between  $\partial \mathbf{S}^u := \overline{\mathbf{S}^u} \setminus \mathbf{S}^u$ , where  $\overline{\mathbf{S}^u}$  is the Doob-Martin compactification of  $\mathbf{S}^u$ , and the set  $\mathcal{S}^u$  of subprobability measures on  $\beta(u)$  that are limits in the vague topology of probability measures of the form

$$\frac{1}{|\zeta_n|} \sum_{v \in \beta(u)} \zeta_n^v \delta_v,$$

where  $(\zeta_n)_{n \in \mathbb{N}_0}$  is a sequence from  $\mathbf{S}^u$  that satisfies (4.4).

- (iv) The bijection between  $\partial \mathbf{S}^u$  and  $\mathcal{S}^u$  is a homeomorphism if the former set is equipped with the trace of the Doob-Martin topology and the latter set is equipped with the trace of the vague topology.
- (v) There is a collection  $\mathbf{R}^u \subseteq \{0, 1\}^{\beta(u)} \cap \mathbf{S}^u$  such that if  $(\zeta_n)_{n \in \mathbb{N}_0}$  is a sequence from  $\mathbf{S}^u$  that satisfies (4.4) and  $\lim_{n \rightarrow \infty} K^u(\eta, \zeta_n)$  exists for all  $\eta \in \mathbf{R}^u$ , then  $\lim_{n \rightarrow \infty} K^u(\xi, \zeta_n)$  exists for all  $\xi \in \mathbf{S}^u$ . Moreover, if  $(\zeta'_n)_{n \in \mathbb{N}_0}$  and  $(\zeta''_n)_{n \in \mathbb{N}_0}$  are two sequences from  $\mathbf{S}^u$  that both satisfy (4.4) and

$$\lim_{n \rightarrow \infty} K^u(\eta, \zeta'_n) = \lim_{n \rightarrow \infty} K^u(\eta, \zeta''_n)$$

for all  $\eta \in \mathbf{R}^u$ , then

$$\lim_{n \rightarrow \infty} K^u(\xi, \zeta'_n) = \lim_{n \rightarrow \infty} K^u(\xi, \zeta''_n)$$

for all  $\xi \in \mathbf{S}^u$ .

- (vi) Suppose that  $(\zeta_n)_{n \in \mathbb{N}_0}$  is a sequence from  $\mathbf{S}^u$  such that (4.4) holds and  $K^u(\xi, \zeta_n)$  converges as  $n \rightarrow \infty$  for all  $\xi \in \mathbf{S}^u$ . Let  $\rho = (\rho^v)_{v \in \beta(u)}$  be the subprobability vector of limiting proportions defined by (4.6). The extended Martin kernel is such that  $K^u(\xi, \rho) = 0$  whenever  $\xi^v \geq 2$  for some  $v \in \beta(u)$  with  $\rho^v = 0$ , whereas if  $\rho^v > 0$  for some  $v \in \beta(u)$ , then there exists a sequence  $(\xi_m)_{m \in \mathbb{N}}$  from  $\mathbf{S}^u$  such that  $\xi_m^v = m$ ,  $\xi_m^w \in \{0, 1\}$  for  $w \neq v$ , and  $K(\xi_m, \rho) > 0$ .
- (vii) A subprobability vector  $\rho$  belongs to  $\mathcal{S}^u$  iff there is a sequence  $(\sigma_n^u)_{n \in \mathbb{N}_0} \in \Sigma^u$  s.t.

$$\lim_{n \rightarrow \infty} \frac{\sigma_n^u}{|\sigma_n^u|} = \lim_{n \rightarrow \infty} \frac{\sigma_n^u}{n} = \rho.$$

**Example 4.7.** Hypothesis 4.6 holds if  $\#\beta(u) = 2$  for all  $u \in \mathbf{I}$  (for example, if  $\mathbf{I} = \{0, 1\}^*$ ),  $\mathbf{S}^u = (\mathbb{N}_0)^{\beta(u)}$ , and the Markov chains  $Y^u = (Y_n^u)_{n \in \mathbb{N}_0}$  are such that  $(Y_n^u + (1, 1))_{n \in \mathbb{N}_0}$  are all Pólya’s urns starting with one black ball and one white ball. This is a consequence of the results in [4]. Indeed, the same is true if for arbitrary  $\mathbf{I}$  with  $\beta(u)$  finite for all  $u \in \mathbf{I}$  we take  $\mathbf{S}^u = (\mathbb{N}_0)^{\beta(u)}$  and let  $Y^u$  be an urn scheme of the sort considered in [5] where there is a (not necessarily integer-valued) finite measure  $\nu_u$  on  $\beta(u)$  that describes the initial composition of an urn with balls whose “colors” are identified with the elements of  $\beta(u)$ , balls are drawn at random and replaced along with a new ball of the same color, and  $Y_n^u$  records the number of balls of the various colors that have been drawn by time  $n$ . In this general case, the extended Martin kernel is given by

$$\frac{(|\nu_u| + |\xi| - 1)(|\nu_u| + |\xi| - 2) \cdots |\nu_u|}{\prod_{v \in \beta(u)} [(\nu_u^v + \xi^v - 1)(\nu_u^v + \xi^v - 2) \cdots \nu_u^v]} \prod_{v \in \beta(u)} (\rho^v)^{\xi^v},$$

where  $|\nu_u| = \sum_{v \in \beta(u)} \nu_u^v$ ,  $|\xi| = \sum_{v \in \beta(u)} \xi^v$ , and  $(\rho^v)^{\xi^v}$  denotes the value  $\rho^v$  that the probability measure  $\rho$  assigns to  $\{v\}$  raised to the power  $\xi^v$ . We may take the set  $\mathbf{R}^u$  in this case to be the coordinate vectors  $e_v$ ,  $v \in \beta(u)$ , where  $e_v$  has a single 1 in the  $v^{\text{th}}$  component and 0 elsewhere. The set  $\mathcal{S}^u$  consists of all the probability measures on the finite set  $\beta(u)$ .

**Example 4.8.** Hypothesis 4.6 also holds if the set  $\beta(u)$  is finite for all  $u \in \mathbf{I}$ ,  $\mathbf{S}^u = (\mathbb{N}_0)^{\beta(u)}$ , and the routing chain  $Y^u$  is given by  $Y^u := (\sum_{k=1}^n W_k^u)_{n \in \mathbb{N}_0}$ , where the  $W_k^u$  are independent, identically distributed  $\mathbf{S}^u$ -valued random variables with distribution that has support the set of coordinate vectors. If  $p_u^v$  is the probability that the common distribution of the  $W_k^u$  assigns to the coordinate vector  $e_v$ , then the extended Martin kernel is given by

$$K^u(\xi, \rho) = \prod_{v \in \beta(u)} \left( \frac{\rho^v}{p_u^v} \right)^{\xi^v}.$$

Results of this type go back to [54] and are described in [35]. Once again, we may take  $\mathbf{R}^u$  to be the set of coordinate vectors, and once again  $\mathcal{S}^u$  consists of all the probability measures on the finite set  $\beta(u)$ .

In order to state a broadly applicable result in the converse direction of Proposition 4.4 we first need to develop some more notation and collect together some auxiliary results.

Adjoin a point  $\diamond$  to  $\mathbf{I}$  and write  $\mathbf{I}_\infty$  for the set of sequences of the form  $(u_n)_{n \in \mathbb{N}_0}$  where either  $u_n \in \mathbf{I}$  for all  $n \in \mathbb{N}_0$  and  $\hat{0} = u_0 \rightarrow u_1 \rightarrow \dots$  or, for some  $N \in \mathbb{N}_0$ ,  $u_n \in \mathbf{I}$  for  $n \leq N$ ,  $\hat{0} = u_0 \rightarrow \dots \rightarrow u_N$ , and  $u_n = \diamond$  for  $n > N$ . We think of  $\mathbf{I}_\infty$  as the space of directed paths through  $\mathbf{I}$  that start at  $\hat{0}$  and are possibly “killed” at some time and sent to the “cemetery”  $\diamond$ .

Write  $\mathcal{C}_\infty$  for the countable collection of subsets of  $\mathbf{I}_\infty$  of the form  $\{(v_n)_{n \in \mathbb{N}_0} \in \mathbf{I}_\infty : v_k = u_k, 0 \leq k \leq n\}$ , where  $n \in \mathbb{N}_0$ ,  $u_k \in \mathbf{I}$  for  $0 \leq k \leq n$ , and  $\hat{0} = u_0 \rightarrow \dots \rightarrow u_n$ . Denote by  $\mathcal{I}_\infty$  the  $\sigma$ -field generated by  $\mathcal{C}_\infty$ . The following result is elementary and we leave its proof to the reader.

**Lemma 4.9.** *Any probability measure on the measurable space  $(\mathbf{I}_\infty, \mathcal{I}_\infty)$  is specified by its values on the sets in  $\mathcal{C}_\infty$ . The space of such probability measures equipped with the coarsest topology that makes each of the maps  $\mu \mapsto \mu(C)$ ,  $C \in \mathcal{C}_\infty$ , continuous is compact and metrizable.*

Consider the case of Lemma 4.9 where the measure  $\mu$  describes the dynamics of a Markov process. That is, for each  $u \in \mathbf{I}$  there is a subprobability measure  $r^u$  on  $\beta(u)$  such that if the process is in state  $u$ , then the next step is with probability  $(r^u)^v$  to  $v$ , and with probability  $1 - \sum_{v \in \beta(u)} (r^u)^v$  to  $\diamond$ .

Label  $u \in \mathbf{I}$  with  $\downarrow$  if  $u$  is *reachable* from  $\hat{0}$  (in the classical sense of Markov chains), and with  $\dagger$  otherwise. Denote by  $\mathbf{J}^\downarrow$  and  $\mathbf{J}^\dagger$  the sets of vertices labeled with  $\downarrow$  and  $\dagger$ , respectively.

Clearly, in order to specify the distribution  $\mu$  of the Markovian path starting from  $\hat{0}$  it suffices to have the subprobability measures  $r^u$  only for  $u \in \mathbf{J}^\downarrow$ .

Note that the labeling  $(\mathbf{J}^\downarrow, \mathbf{J}^\dagger)$  has the two properties

- the vertex  $\hat{0}$  is labeled with  $\downarrow$ ;
- if for some  $v \neq \hat{0}$  every vertex  $u \in \alpha(v)$  is labeled with  $\dagger$ , then  $v$  is also labeled with  $\dagger$ .

Let us now switch perspectives and start from a labeling instead of a collection of subprobability measures.

**Definition 4.10.** *Say that a labeling of  $\mathbf{I}$  with the symbols  $\downarrow$  and  $\dagger$  is admissible if it satisfies the above two properties. Write  $\mathbf{I}^\downarrow$  (resp.  $\mathbf{I}^\dagger$ ) for the subset of vertices labeled with  $\downarrow$  (resp.  $\dagger$ ).*

Note that if  $(\mathbf{I}^\downarrow, \mathbf{I}^\dagger)$  is an admissible labeling of  $\mathbf{I}$ ,  $(u_n)_{n \in \mathbb{N}_0}$  is a directed path in  $\mathbf{I}$  with  $u_0 = \hat{0}$ , and we define a sequence  $(\tilde{u}_n)_{n \in \mathbb{N}_0}$  in  $\mathbf{I} \cup \{\diamond\}$  by

$$\tilde{u}_n := \begin{cases} u_n, & \text{if } u_n \in \mathbf{I}^\downarrow, \\ \diamond, & \text{if } u_n \in \mathbf{I}^\dagger, \end{cases}$$

then  $(\tilde{u}_n)_{n \in \mathbb{N}_0}$  is an element of  $\mathbf{I}_\infty$ .

**Definition 4.11.** *Given an admissible labeling  $(\mathbf{I}^\downarrow, \mathbf{I}^\dagger)$  of  $\mathbf{I}$ , say that a collection  $(r^u)_{u \in \mathbf{I}^\downarrow}$ , where  $r^u$  is a subprobability measure on  $\beta(u)$  for  $u \in \mathbf{I}^\downarrow$ , is compatible with the labeling if a vertex  $v \in \mathbf{I} \setminus \{\hat{0}\}$  is in  $\mathbf{I}^\dagger$  if and only if  $\alpha(v) \cap \mathbf{I}^\downarrow = \emptyset$  or  $(r^u)^v = 0$  for  $u \in \alpha(v) \cap \mathbf{I}^\downarrow$ .*

**Remark 4.12.** *For an admissible labeling  $(\mathbf{I}^\downarrow, \mathbf{I}^\dagger)$  of  $\mathbf{I}$  and a collection of subprobability measures as in Definition 4.11, compatibility of the subprobability measures with the labeling is equivalent to the equality  $\mathbf{I}^\downarrow = \mathbf{J}^\downarrow$ , where  $\mathbf{J}^\downarrow$  is the set of vertices that are reachable from  $\hat{0}$  under the Markovian dynamics specified by the subprobability measures.*

The assertions (i), (ii) and (iii) in the following lemma, with  $\mathbf{J}^\downarrow$  and  $\mathbf{J}^\dagger$  instead of  $\mathbf{I}^\downarrow$  and  $\mathbf{I}^\dagger$ , are obvious. The proof of the lemma is then clear from the previous remark.

**Lemma 4.13.** *Consider an admissible labeling of  $\mathbf{I}$  with the symbols  $\downarrow$  and  $\dagger$  and a compatible collection of subprobability measures  $(r^u)_{u \in \mathbf{I}^\downarrow}$ .*

- (i) There is a unique probability measure  $\mu$  on  $(\mathbf{I}_\infty, \mathcal{I}_\infty)$  for which the mass assigned to the set  $\{(v_n)_{n \in \mathbb{N}_0} \in \mathbf{I}_\infty : v_k = u_k, 0 \leq k \leq n\} \in \mathcal{C}_\infty$  is

$$\begin{cases} \prod_{k=0}^{n-1} (r^{u_k})^{u_{k+1}}, & \text{if } u_k \in \mathbf{I}^\downarrow \text{ for } 0 \leq k \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

- (ii) The vertex  $u$  belongs to  $\mathbf{I}^\dagger$  if and only if  $\mu\{(v_n)_{n \in \mathbb{N}_0} \in \mathbf{I}_\infty : v_k = u_k, 0 \leq k \leq n\} = 0$  whenever  $\hat{0} = u_0 \rightarrow \dots \rightarrow u_n = u$ .  
 (iii) If  $u \in \mathbf{I}^\downarrow$  and  $v \in \beta(u)$ , then

$$(r^u)^v = \frac{\mu\{(v_n)_{n \in \mathbb{N}_0} \in \mathbf{I}_\infty : v_k = u_k, 0 \leq k \leq n+1\}}{\mu\{(v_n)_{n \in \mathbb{N}_0} \in \mathbf{I}_\infty : v_k = u_k, 0 \leq k \leq n\}}$$

for any choice of  $\hat{0} = u_0 \rightarrow \dots \rightarrow u_n = u \rightarrow u_{n+1} = v$  such that the denominator is positive. In particular, it is possible to recover the labeling and the collection  $(r^u)_{u \in \mathbf{I}^\downarrow}$  from the probability measure  $\mu$ .

**Theorem 4.14.** Suppose that Hypothesis 4.6 holds. Denote by  $\mathcal{R}_\infty$  the set of pairs  $((\mathbf{I}^\downarrow, \mathbf{I}^\dagger), (r^u)_{u \in \mathbf{I}^\downarrow})$ , such that  $(\mathbf{I}^\downarrow, \mathbf{I}^\dagger)$  is an admissible labeling of  $\mathbf{I}$  and  $(r^u)_{u \in \mathbf{I}^\downarrow} \in \prod_{u \in \mathbf{I}^\downarrow} \mathcal{S}^u$  is a compatible collection of subprobability measures.

- (i) If a sequence  $(y_n)_{n \in \mathbb{N}_0}$  in  $\mathbf{S}$  converges to a point in the Doob-Martin boundary  $\partial\mathbf{S} = \bar{\mathbf{S}} \setminus \mathbf{S}$ , then there exists  $((\mathbf{I}^\downarrow, \mathbf{I}^\dagger), (r^u)_{u \in \mathbf{I}^\downarrow}) \in \mathcal{R}_\infty$  satisfying

$$\lim_{n \rightarrow \infty} \frac{y_n^u}{|y_n^u|} = r^u \in \mathcal{S}^u, \quad \text{for all } u \in \mathbf{I}^\downarrow. \tag{4.7}$$

Moreover, if two such sequences converge to the same point then the corresponding elements of  $\mathcal{R}_\infty$  coincide.

- (ii) Conversely, if  $((\mathbf{I}^\downarrow, \mathbf{I}^\dagger), (r^u)_{u \in \mathbf{I}^\downarrow}) \in \mathcal{R}_\infty$ , then there is a sequence  $(y_n)_{n \in \mathbb{N}_0}$  in  $\mathbf{S}$  that converges to a point in the Doob-Martin boundary  $\partial\mathbf{S} = \bar{\mathbf{S}} \setminus \mathbf{S}$  and satisfies (4.7). Moreover, any two such sequences converge to the same point, establishing a bijection between  $\mathcal{R}_\infty$  and  $\partial\mathbf{S}$ .  
 (iii) For  $x \in \mathbf{S}$  and  $((\mathbf{I}^\downarrow, \mathbf{I}^\dagger), (r^u)_{u \in \mathbf{I}^\downarrow}) \in \mathcal{R}_\infty \cong \partial\mathbf{S}$ , the value of the extended Martin kernel is

$$\begin{cases} \prod_{u \in \mathbf{I}^\downarrow} K^u(x^u, r^u), & \text{if } x^v = (0, 0, \dots) \text{ for all } v \notin \mathbf{I}^\downarrow, \\ 0, & \text{otherwise.} \end{cases}$$

- (iv) Let  $\mathcal{P}_\infty$  be the set of probability measures on  $\mathbf{I}_\infty$  constructed from elements of  $\mathcal{R}_\infty$  via the bijection of Lemma 4.13. Equip  $\mathcal{P}_\infty$  with the trace of the metrizable topology introduced in Lemma 4.9. The composition of the bijection between  $\mathcal{P}_\infty$  and  $\mathcal{R}_\infty$  and the bijection between  $\mathcal{R}_\infty$  and  $\partial\mathbf{S}$  is a homeomorphism between  $\mathcal{P}_\infty$  and  $\partial\mathbf{S}$ .

*Proof.* Consider part (i). Suppose that the sequence  $(y_n)_{n \in \mathbb{N}_0}$  converges to a point in  $\partial\mathbf{S}$ ; that is,

$$\lim_{n \rightarrow \infty} K(x, y_n) \text{ exists for all } x \in \mathbf{S} \tag{4.8}$$

and no subsequence converges in the discrete topology on  $\mathbf{S}$  to a point of  $\mathbf{S}$ . Thus,

$$\#\{n \in \mathbb{N}_0 : y_n = y\} < \infty \text{ for any } y \in \mathbf{S}. \tag{4.9}$$

Because of (4.9) and Hypothesis 4.6(i), it follows that

$$\lim_{n \rightarrow \infty} |y_n^{\hat{0}}| = \infty. \tag{4.10}$$

Consider  $\eta \in \mathbf{R}^{\hat{0}}$ . Define  $x \in \mathbf{S}$  by setting  $x^{\hat{0}} = \eta$ . By the consistency condition (2.3), this completely specifies  $x$ . Note that  $x^w = 0$  if  $w \neq \hat{0}$ . By Corollary 4.2,

$$K(x, y_n) = K^{\hat{0}}(\eta, y_n^{\hat{0}}),$$

and so  $\lim_{n \rightarrow \infty} K^{\hat{0}}(\eta, y_n^{\hat{0}})$  exists. Since this is true for all  $\eta \in \mathbf{R}^{\hat{0}}$ , it follows from Hypothesis 4.6(v) that  $\lim_{n \rightarrow \infty} K^{\hat{0}}(\xi, y_n^{\hat{0}})$  exists for all  $\xi \in \mathbf{S}^{\hat{0}}$ . Hence, by Hypothesis 4.6(ii)

$$\lim_{n \rightarrow \infty} \frac{(y_n^{\hat{0}})^v}{|y_n^{\hat{0}}|}$$

exists for all  $v \in \beta(\hat{0})$ . Write  $r^{\hat{0}} = ((r^{\hat{0}})^v)_{v \in \beta(\hat{0})} \in \mathcal{S}^{\hat{0}}$  for the subprobability vector defined by the limits.

If  $(r^{\hat{0}})^v = 0$  for some  $v \in \beta(\hat{0})$ , then, from Hypothesis 4.6(vi),  $\lim_{n \rightarrow \infty} K(x, y_n) = 0$  for any  $x \in \mathbf{S}$  with  $(x^{\hat{0}})^v \geq 2$  – no matter what the values of  $y_n^u$  are for  $u > \hat{0}$ . Consequently, in order to understand what further constraints are placed on the sequence  $(y_n)_{n \in \mathbb{N}_0}$  by the assumption that (4.8) holds, we need only consider choices of  $x \in \mathbf{S}$  with the property that  $(x^{\hat{0}})^v \in \{0, 1\}$  for all  $v \in \beta(\hat{0})$  such that  $(r^{\hat{0}})^v = 0$ . Note from the consistency condition (2.3) that for this restricted class of  $x$  we must have  $x^w = 0$  for all  $w \in \mathbf{I}$  such that all directed path from  $\hat{0}$  to  $w$  necessarily passes through  $v \in \beta(\hat{0})$  with  $(r^{\hat{0}})^v = 0$ .

Suppose that  $r^{\hat{0}} \neq 0$ . Fix a vertex  $u \in \beta(\hat{0})$  such that  $(r^{\hat{0}})^u > 0$  and  $\eta \in \mathbf{R}^u$ . From Hypothesis 4.6(vi), there exists  $\theta \in \mathbf{S}^{\hat{0}}$  such that  $\theta^u = |\eta| + 1$ , and  $\theta^w \in \{0, 1\}$  for  $w \neq u$ . Define  $x \in \mathbf{S}$  by setting  $x^{\hat{0}} = \theta$  and  $x^u = \eta$ . By the consistency condition (2.3), this completely specifies  $x$ . Note that  $x^w = 0$  if  $w \notin \{\hat{0}, u\}$ . By Corollary 4.2,

$$K(x, y_n) = K^{\hat{0}}(\theta, y_n^{\hat{0}})K^u(\eta, y_n^u),$$

and, by the choice of  $\theta$ ,  $K^{\hat{0}}(\theta, y_n^{\hat{0}})$  converges to a non-zero value as  $n \rightarrow \infty$ . Therefore,  $\lim_{n \rightarrow \infty} K^u(\eta, y_n^u)$  exists. Since this is true for all  $\eta \in \mathbf{R}^u$ , it follows from Hypothesis 4.6(v) that  $\lim_{n \rightarrow \infty} K^u(\xi, y_n^u)$  exists for all  $\xi \in \mathbf{S}^u$ . Hence, by Hypothesis 4.6(ii),

$$\lim_{n \rightarrow \infty} \frac{(y_n^u)^v}{|y_n^u|}$$

exists for all  $v \in \beta(u)$ . Write  $r^u \in \mathcal{S}^u$  for the resulting subprobability measure.

Continuing in this way, we see that, under the assumption (4.9), if (4.8) holds then there is a labeling of  $\mathbf{I}$  with the symbols  $\downarrow$  and  $\uparrow$  such that the following are true:

- the vertex  $\hat{0}$  is in  $\mathbf{I}^{\downarrow}$ ;
- if a vertex  $u$  is in  $\mathbf{I}^{\downarrow}$ , then the limiting subprobability measure

$$\lim_{n \rightarrow \infty} \frac{y_n^u}{|y_n^u|} =: r^u \in \mathcal{S}^u$$

exists;

- a vertex  $v \neq \hat{0}$  belongs to  $\mathbf{I}^{\uparrow}$  if and only if every vertex  $u \in \alpha(v)$  belongs to  $\mathbf{I}^{\uparrow}$  or  $(r^u)^v = 0$  for every vertex  $u \in \alpha(v) \cap \mathbf{I}^{\downarrow}$ .

Thus, the labeling  $(\mathbf{I}^{\downarrow}, \mathbf{I}^{\uparrow})$  is admissible and the collection  $(r^u)_{u \in \mathbf{I}^{\downarrow}} \in \prod_{u \in \mathbf{I}^{\downarrow}} \mathcal{S}^u$  are compatible, so  $(I^{\downarrow}, I^{\uparrow}), (r^u)_{u \in \mathbf{I}^{\downarrow}}$  is an element of  $\mathcal{R}_{\infty}$ .

Suppose that  $(y_n)_{n \in \mathbb{N}_0}$  and  $(z_n)_{n \in \mathbb{N}_0}$  are two sequences from  $\mathbf{S}$  that converge to the same point in  $\partial \mathbf{S}$ . Then,  $|y_n^{\hat{0}}| \rightarrow \infty$  and  $|z_n^{\hat{0}}| \rightarrow \infty$  as  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} K(x, y_n) \text{ exists for all } x \in \mathbf{S},$$

$$\lim_{n \rightarrow \infty} K(x, z_n) \text{ exists for all } x \in \mathbf{S},$$

and

$$\lim_{n \rightarrow \infty} K(x, z_n) = \lim_{n \rightarrow \infty} K(x, z_n) \text{ for all } x \in \mathbf{S}.$$

It is clear that the vertices of  $\mathbf{I}$  that are labeled with the symbol  $\downarrow$  (resp.  $\uparrow$ ) for the sequence  $(y_n)_{n \in \mathbb{N}_0}$  must coincide with the vertices of  $\mathbf{I}$  that are labeled with the symbol  $\downarrow$  (resp.  $\uparrow$ ) for the sequence  $(z_n)_{n \in \mathbb{N}_0}$ , and

$$\lim_{n \rightarrow \infty} \frac{y_n^u}{|y_n^u|} = \lim_{n \rightarrow \infty} \frac{z_n^u}{|z_n^u|}$$

for the common set of vertices  $u \in \mathbf{I}$  labeled with  $\downarrow$ . This completes the proof of part (i).

Moreover, it follows from what we have just done that if  $x \in \mathbf{S}$  and the convergent sequence  $(y_n)_{n \in \mathbb{N}_0}$  is associated with  $(\mathbf{I}^\downarrow, \mathbf{I}^\uparrow), (r^u)_{u \in \mathbf{I}^\downarrow}$ , then

$$\lim_{n \rightarrow \infty} K(x, y_n) = \begin{cases} \prod_{u \in \mathbf{I}^\downarrow} K^u(x^u, r^u), & \text{if } x^v = (0, 0, \dots) \text{ for all } v \notin \mathbf{I}^\downarrow, \\ 0, & \text{otherwise.} \end{cases} \quad (4.11)$$

This establishes part (iii) once we show part (ii).

Now consider part (ii). Fix  $(\mathbf{I}^\downarrow, \mathbf{I}^\uparrow), (r^u)_{u \in \mathbf{I}^\downarrow} \in \mathcal{R}_\infty$ . By Hypothesis 4.6(vii), for each  $u \in \mathbf{I}^\downarrow$  there is a sequence  $(\sigma_n^u)_{n \in \mathbb{N}_0} \in \Sigma^u$  such that

$$\lim_{n \rightarrow \infty} \frac{\sigma_n^u}{|\sigma_n^u|} = r^u.$$

Choose sequences  $(\sigma_n^u)_{n \in \mathbb{N}_0} \in \Sigma^u$  for  $u \notin \mathbf{I}^\downarrow$  arbitrarily and set  $\sigma = (\sigma^u)_{u \in \mathbf{I}} \in \Sigma$ . Define a sequence  $(y_n)_{n \in \mathbb{N}_0}$  from  $\mathbf{S}$  by setting  $y_n^u = \sigma^u(a_n^u(\sigma))$  for  $n \in \mathbb{N}_0$  and  $u \in \mathbf{I}$ . It is clear from the arguments for part (i) that  $(y_n)_{n \in \mathbb{N}_0}$  converges to a point in  $\partial\mathbf{S}$  and (4.7) holds. Moreover, it follows from the same arguments that any two convergent sequences satisfying (4.7) must converge to the same point. This establishes (ii).

The proof of (iv) is straightforward and we omit it. □

## 5 Binary search tree and digital search tree processes

Recall the binary search tree (BST) process from the Introduction. We observed in Example 4.7 that Hypothesis 4.6 holds for the BST process. Recall from Example 2.3 that we can identify  $\mathbf{S}$  in this case with the set of finite subtrees of the complete binary tree  $\{0, 1\}^*$  that contain the root  $\emptyset$ . Moreover, it follows from the discussion in Section 4 that  $\partial\mathbf{S}$  is homeomorphic to the set of probability measures on  $\{0, 1\}^\infty$  equipped with the weak topology corresponding to the usual product topology on  $\{0, 1\}^\infty$ .

We therefore abuse notation slightly and take  $\mathbf{S}$  to be set of finite subtrees of  $\{0, 1\}^*$  rooted at  $\emptyset$  and take  $\partial\mathbf{S}$  to be the probability measures on  $\{0, 1\}^\infty$ .

With this identification the partial order  $\preceq$  on  $\mathbf{S}$  is just subset containment and the Martin kernel is given by

$$K(\mathbf{s}, \mathbf{t}) = \begin{cases} \left(\frac{\#\mathbf{t}}{\#\mathbf{s}}\right)^{-1} \prod_{u \in \mathbf{s}} \#\mathbf{t}(u), & \text{if } \mathbf{s} \subseteq \mathbf{t}, \\ 0, & \text{otherwise,} \end{cases} \quad (5.1)$$

where we recall from Example 4.3 that  $\#\mathbf{t}(u) = \#\{v \in \mathbf{t} : u \leq v\}$ .

A sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}}$  in  $\mathbf{S}$  with  $\#\mathbf{t}_n \rightarrow \infty$  converges in the Doob-Martin compactification of  $\mathbf{S}$  if and only if  $\#\mathbf{t}_n(u)/\#\mathbf{t}_n$  converges for all  $u \in \{0, 1\}^*$ . Moreover, if the sequence converges, then the limit can be identified with the probability measure  $\mu$  on  $\{0, 1\}^\infty$  such that

$$\mu\{v \in \{0, 1\}^\infty : u < v\} = \lim_{n \rightarrow \infty} \frac{\#\mathbf{t}_n(u)}{\#\mathbf{t}_n}$$

for all  $u \in \{0, 1\}^*$ .

Recall that the partial order on  $\{0, 1\}^*$  is such that if  $u = u_1 \dots u_k$  and  $v = v_1 \dots v_\ell$  are two words, then  $u \leq v$  if and only if  $u$  is an initial segment of  $v$ , that is, if and only if  $k \leq \ell$  and  $u_i = v_i$  for  $i = 1, \dots, k$ . Extend this partial order to  $\{0, 1\}^* \sqcup \{0, 1\}^\infty$  by declaring that any two elements of  $\{0, 1\}^\infty$  are not comparable and  $u < v$  for  $u = u_1 \dots u_k \in \{0, 1\}^*$  and  $v = v_1 v_2 \dots \in \{0, 1\}^\infty$  when  $u_i = v_i$  for  $i = 1, \dots, k$ . Given  $\mu \in \partial \mathbf{S}$ , set

$$\mu_u := \mu\{v \in \{0, 1\}^\infty : u < v\}. \tag{5.2}$$

That is,  $\mu_u$  is the mass assigned by  $\mu$  to the set of infinite paths in the complete binary tree that begin at the root and that pass through the vertex  $u$ . The extended Martin kernel is given by

$$K(\mathbf{s}, \mu) = (\#\mathbf{s})! \prod_{u \in \mathbf{s}} \mu_u, \quad \mathbf{s} \in \mathbf{S}, \mu \in \partial \mathbf{S}. \tag{5.3}$$

Note from the construction of the BST process that its transition matrix is

$$P(\mathbf{s}, \mathbf{t}) = \begin{cases} \frac{1}{\#\mathbf{s}+1}, & \text{if } \mathbf{s} \subset \mathbf{t} \text{ and } \#(\mathbf{t} \setminus \mathbf{s}) = 1, \\ 0, & \text{otherwise,} \end{cases}$$

(this is also apparent from (4.3)). Set  $h_\mu := K(\cdot, \mu)$  for  $\mu \in \partial \mathbf{S}$ . The Doob  $h$ -transform process corresponding to the regular function  $h_\mu$  has state space

$$\{\mathbf{t} \in \mathbf{S} : \mu_u > 0 \text{ for all } u \in \mathbf{t}\}$$

and transition matrix

$$P^{(h_\mu)}(\mathbf{s}, \mathbf{t}) = \begin{cases} \mu_u, & \text{if } \mathbf{t} = \mathbf{s} \sqcup \{u\}, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that the  $h$ -transformed process results from a trickle-down construction. For simplicity, we only verify this in the case when  $\mu_u > 0$  for all  $u \in \{0, 1\}^* = \mathbf{I}$ , so that the state-space of the  $h$ -transformed process is all of  $\mathbf{S}$ , and leave the formulation of the general case to the reader. The routing chain on  $\mathbf{S}^u = \mathbb{N}_0^{\{u0, u1\}}$  has transition matrix  $Q^u$  given by

$$Q^u((m, n), (m + 1, n)) = \frac{\mu_{u0}}{\mu_u}$$

and

$$Q^u((m, n), (m, n + 1)) = \frac{\mu_{u1}}{\mu_u}.$$

In other words, we can regard the routing chain as the space-time chain corresponding to the one-dimensional simple random walk that has probability  $\mu_{u0}/\mu_u$  of making a  $-1$  step and probability  $\mu_{u1}/\mu_u$  of making a  $+1$  step.

We have the following “trickle-up” construction of the  $h$ -transformed process. Suppose on some probability space that there is a sequence of independent identically distributed  $\{0, 1\}^\infty$ -valued random variables  $(V^n)_{n \in \mathbb{N}}$  with common distribution  $\mu$ . For an initial finite rooted subtree  $\mathbf{w}$  in the state space of the  $h$ -transformed process, define a sequence  $(W_n)_{n \in \mathbb{N}_0}$  of random finite subsets of  $\{0, 1\}^*$  inductively by setting  $W_0 := \mathbf{w}$  and  $W_{n+1} := W_n \cup \{V_1^{n+1} \dots V_{H(n+1)+1}^{n+1}\}$ ,  $n \geq 0$ , where  $H(n + 1) := \max\{l \in \mathbb{N} : V_1^{n+1} \dots V_l^{n+1} \in W_n\}$  with the convention  $\max \emptyset = 0$ . That is, at each point in time we start a particle at a “leaf” of the complete binary tree  $\{0, 1\}^*$  picked according to  $\mu$  and then let that particle trickle up the tree until it can go no further because its path is blocked by previous particles that have come to rest. It is clear that  $(W_n)_{n \in \mathbb{N}_0}$  is a Markov chain with state space the appropriate set of finite rooted subtrees of  $\{0, 1\}^*$ , initial state  $\mathbf{w}$ , and transition matrix  $P^{(h_\mu)}$ .

It follows from the trickle-up construction and Kolmogorov's zero-one law that the tail  $\sigma$ -field of the  $h$ -transformed process is trivial, and hence  $\mu$  is an extremal point of  $\mathbf{S}$ . Alternatively,  $\mu$  is extremal because it is clear from the strong law of large numbers that the  $h$ -transformed process converges to  $\mu$ .

Consider the special case of the  $h$ -transform construction when the boundary point  $\mu$  is the "uniform" or "fair coin-tossing" measure on  $\{0, 1\}^\infty$ ; that is,  $\mu$  is the infinite product of copies of the measure on  $\{0, 1\}$  that assigns mass  $\frac{1}{2}$  to each of the subsets  $\{0\}$  and  $\{1\}$ . In this case, the transition matrix of the  $h$ -transformed process is

$$P^{(h_\mu)}(\mathbf{s}, \mathbf{t}) = \begin{cases} 2^{-|u|}, & \text{if } \mathbf{t} = \mathbf{s} \sqcup \{u\}, \\ 0, & \text{otherwise,} \end{cases}$$

where we write  $|u|$  for the length of the word  $u$ ; that is,  $|u| = k$  when  $u = u_1 \dots u_k$ . This transition mechanism is that of the *digital search tree (DST) process*. We have therefore established the following result.

**Theorem 5.1.** *The digital search tree process is the Doob  $h$ -transform of the binary search tree process associated with the regular function  $h(\mathbf{s}) := (\#\mathbf{s})! \prod_{u \in \mathbf{s}} 2^{-|u|}$ ,  $\mathbf{s} \in \mathbf{S}$ . The regular function  $h$  is extremal and corresponds to the uniform probability measure on  $\{0, 1\}^\infty$  thought of as an element of the Doob-Martin compactification of the state space  $\mathbf{S}$  of the BST process. Consequently, the Doob-Martin compactification of the DST process coincides with that of the BST process.*

**Remark 5.2.** *The digital search tree (DST) algorithm is discussed in [33, p.496ff] and in [38, Chapter 6]. The process in Theorem 5.1 appears as the output of the DST algorithm if the input is a sequence of independent and identically distributed random 0-1 sequences with distribution  $\mu$ , where  $\mu$  is the fair coin tossing measure. In the literature this assumption is also known as the symmetric Bernoulli model; in the general Bernoulli model the probability  $1/2$  for an individual digit 1 is replaced by an arbitrary  $p \in (0, 1)$ . In our approach we do not need any assumptions on the internal structure of the random 0-1 sequences and we can work with a general distribution  $\mu$  on  $\{0, 1\}^\infty$ . Any such DST processes "driven by  $\mu$ " is an  $h$ -transform of the BST process, provided that  $\mu_u > 0$  for all  $u \in \mathbf{I}$ , and the trickle-up construction shows that the conditional distribution of the BST process, given that its limit is  $\mu$ , is the same as the distribution of the DST process driven by  $\mu$ .*

In the symmetric Bernoulli model, the sample paths of the DST process converge almost surely to the single boundary point  $\mu$  in the Doob-Martin topology, where  $\mu$  is the uniform measure on  $\{0, 1\}^\infty$ . We now investigate the distribution of the limit of the sample paths of the BST process. There are several routes we could take.

Recall that the routing chains for the BST process are essentially Pólya urns; that is, the routing chain  $Y^u = ((Y^u)^{u0}, (Y^u)^{u1})$  for the vertex  $u \in \{0, 1\}^*$  makes the transition  $(g, d) \rightarrow (g + 1, d)$  with probability  $(g + 1)/(g + d + 2)$  and the transition  $(g, d) \rightarrow (g, d + 1)$  with probability  $(d + 1)/(g + d + 2)$ . It is a well-known fact about the Pólya urn that, when  $((Y_0^u)^{u0}, (Y_0^u)^{u1}) = (0, 0)$ , the sequence  $((Y_n^u)^{u0} + (Y_n^u)^{u1})^{-1}((Y_n^u)^{u0}, (Y_n^u)^{u1})$ ,  $n \in \mathbb{N}_0$ , converges almost surely to a random variable of the form  $(U, 1 - U)$ , where  $U$  is uniformly distributed on  $[0, 1]$ . It follows that if we write  $(T_n)_{n \in \mathbb{N}}$  for the BST process, then almost surely

$$\frac{\#T_n(u)}{\#T_n} \rightarrow \prod_{\emptyset < v \leq u} U_v, \quad u \in \{0, 1\}^*,$$

where the pairs  $(U_{u0}, U_{u1})$ ,  $u \in \{0, 1\}^*$ , are independent, the random variables  $U_{u0}$  and  $U_{u1}$  are uniformly distributed on  $[0, 1]$ , and  $U_{u0} + U_{u1} = 1$ . Thus, the limit of the BST chain is the random measure  $M$  on  $\{0, 1\}^\infty$  such that  $M_u = \prod_{\emptyset < v \leq u} U_v$  for all  $u \in \{0, 1\}^*$ .

Another approach is to observe that, from the trickle-up description of the  $h$ -transformed processes described above and the extremality of all the boundary points, we only need to find a random measure on  $\{0, 1\}^\infty$  such that if we perform the trickle-up construction from a realization of the random measure, then we produce the BST process. It follows from the main result of [5] that the random measure  $M$  has the correct properties.

Yet another perspective is to observe that, by the general theory outlined in Section 3, the distribution of the limit is the unique probability measure  $\mathbb{M}$  on  $\partial\mathbf{S}$  such that

$$1 = \int_{\partial\mathbf{S}} K(\mathbf{s}, \mu) \mathbb{M}(d\mu).$$

In the present situation the right hand side evaluates to

$$\int_{\partial\mathbf{S}} (\#\mathbf{s})! \prod_{u \in \mathbf{s}} \mu_u \mathbb{M}(d\mu) = (\#\mathbf{s})! \mathbb{E} \left[ \prod_{u \in \mathbf{s}} \tilde{M}_u \right],$$

where  $\tilde{M}$  is a random measure on  $\{0, 1\}^\infty$  with distribution  $\mathbb{M}$ . Rather than simply verify that taking  $\tilde{M} = M$ , where  $M_u = \prod_{\emptyset < v \leq u} U_v$  as above, has the requisite property, we consider a more extensive class of random probability measures with similar structure, compute the corresponding regular functions, and identify the transition matrices of the resulting  $h$ -transform processes.

Let the pairs  $(R_{u0}, R_{u1})$ ,  $u \in \{0, 1\}^*$ , be independent and take values in the set  $\{(a, b) : a, b \geq 0, a + b = 1\}$ . Define a random probability measure  $N$  on  $\{0, 1\}^\infty$  by setting  $N_u := \prod_{\emptyset < v \leq u} R_v$  for all  $u \in \{0, 1\}^*$ . The corresponding regular function is

$$\begin{aligned} h(\mathbf{s}) &= \mathbb{E}[K(\mathbf{s}, N)] \\ &= (\#\mathbf{s})! \mathbb{E} \left[ \prod_{u \in \mathbf{s}} N_u \right] \\ &= (\#\mathbf{s})! \mathbb{E} \left[ \prod_{u \in \mathbf{s}} \prod_{\emptyset < v \leq u} R_v \right] \\ &= (\#\mathbf{s})! \mathbb{E} \left[ \prod_{u \in \mathbf{s} \setminus \{\emptyset\}} R_u^{\#\mathbf{s}(u)} \right] \\ &= (\#\mathbf{s})! \prod_u A_u(\#\mathbf{s}(u0), \#\mathbf{s}(u1)), \end{aligned}$$

where the last product is over  $\{0, 1\}^*$  and

$$A_u(j, k) := \mathbb{E} \left[ R_{u0}^j R_{u1}^k \right].$$

With this notation, the probability that the resulting  $h$ -transform of the BST process makes a transition from  $\mathbf{s}$  to  $\mathbf{t} := \mathbf{s} \sqcup \{v\}$  is

$$\begin{aligned} h(\mathbf{s})^{-1} \frac{1}{\#\mathbf{s} + 1} h(\mathbf{t}) &= \frac{1}{\#\mathbf{s} + 1} \frac{(\#\mathbf{s} + 1)! \prod_u A_u(\#\mathbf{t}(u0), \#\mathbf{t}(u1))}{(\#\mathbf{s})! \prod_u A_u(\#\mathbf{s}(u0), \#\mathbf{s}(u1))} \\ &= \prod_u \frac{A_u(\#\mathbf{t}(u0), \#\mathbf{t}(u1))}{A_u(\#\mathbf{s}(u0), \#\mathbf{s}(u1))} \\ &= \prod_{\emptyset \leq u < v} \frac{A_u(\#\mathbf{t}(u0), \#\mathbf{t}(u1))}{A_u(\#\mathbf{s}(u0), \#\mathbf{s}(u1))}, \end{aligned} \tag{5.4}$$

because  $\#s(u) = \#t(u)$  unless  $u \leq v$ .

The ratios in (5.4) have a simple form: if  $\#s(u0) = j$  and  $\#s(u1) = k$ , then

$$\frac{A_u(\#t(u0), \#t(u1))}{A_u(\#s(u0), \#s(u1))} = \frac{A_u(j+1, k)}{A_u(j, k)}, \quad \text{if } u0 \leq v, \quad (5.5)$$

and

$$\frac{A_u(\#t(u0), \#t(u1))}{A_u(\#s(u0), \#s(u1))} = \frac{A_u(j, k+1)}{A_u(j, k)}, \quad \text{if } u1 \leq v. \quad (5.6)$$

Suppose now that each  $R_{u0}$  has a beta distribution with parameters  $\theta_u$  and  $\eta_u$ , (so that  $R_{u1} = 1 - R_{u0}$  has a beta distribution with parameters  $\eta_u$  and  $\theta_u$  and the pair  $(R_{u0}, R_{u1})$  has a Dirichlet distribution with parameters  $\theta_u$  and  $\eta_u$ ). Then,

$$A_u(j, k) = \frac{\theta_u(\theta_u + 1) \cdots (\theta_u + j - 1) \times \eta_u(\eta_u + 1) \cdots (\eta_u + k - 1)}{(\theta_u + \eta_u)(\theta_u + \eta_u + 1) \cdots (\theta_u + \eta_u + j + k - 1)},$$

and the factors in (5.5) and (5.6) are

$$\frac{\theta_u + j}{\theta_u + \eta_u + j + k} \quad \text{and} \quad \frac{\eta_u + k}{\theta_u + \eta_u + j + k},$$

respectively. As expected, the BST chain arises as the special case  $\theta_u = \eta_u = 1$  for all  $u$ .

**Remark 5.3.** *The chain with  $\theta_u = \eta_u = \ell$  for some fixed  $\ell \in \mathbb{N}$  appears in connection with the median-of- $(2\ell - 1)$  version of the algorithms Quicksort and Quickselect (Find) – see [26].*

A special case of the above construction arises in connection with Dirichlet random measures. Recall that a *Dirichlet random measure* (sometimes called a *Ferguson random measure*) directed by a finite measure  $\nu$  on  $\{0, 1\}^\infty$  is a random probability measure  $\tilde{N}$  on  $\{0, 1\}^\infty$  with the property that, for any Borel partition  $B_1, \dots, B_k$  of  $\{0, 1\}^\infty$ , the random vector  $(\tilde{N}(B_1), \dots, \tilde{N}(B_k))$  has a Dirichlet distribution with parameters  $(\nu(B_1), \dots, \nu(B_k))$ . In particular,  $\tilde{N}(B)$  has a beta distribution with parameters  $\nu(B)$  and  $\nu(B^c)$ . It follows easily from Lemma 5.4 below that if  $\theta_u = \nu_{u0}$  and  $\eta_u = \nu_{u1}$  for all  $u$  in the above construction of a random probability measure using beta distributed weights, then the result is a Dirichlet random measure directed by  $\nu$ . (We note that the random measures that appear as the limit of the BST and median-of- $(2\ell - 1)$  processes are **not** Dirichlet.)

**Lemma 5.4.** *Suppose that  $(D_1, D_2, D_3, D_4)$  is a Dirichlet distributed random vector with parameters  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ . Then, the three pairs*

$$\left( \frac{D_1}{D_1 + D_2}, \frac{D_2}{D_1 + D_2} \right), \left( \frac{D_3}{D_3 + D_4}, \frac{D_4}{D_3 + D_4} \right) \quad \text{and} \quad (D_1 + D_2, D_3 + D_4)$$

*are independent Dirichlet distributed random vectors with respective parameters*

$$(\alpha_1, \alpha_2), \quad (\alpha_3, \alpha_4), \quad \text{and} \quad (\alpha_1 + \alpha_2, \alpha_3 + \alpha_4).$$

*Proof.* Note that  $(D_1, D_2, D_3, D_4)$  has the same distribution as

$$\left( \frac{G_1}{G_1 + \cdots + G_4}, \dots, \frac{G_4}{G_1 + \cdots + G_4} \right),$$

where the  $G_1, \dots, G_4$  are independent and  $G_i$  has the Gamma distribution with parameters  $(\alpha_i, 1)$ . Moreover, the latter random vector is independent of the sum  $G_1 + \cdots + G_4$ .

Now,

$$\left( D_1 + D_2, D_3 + D_4, \frac{D_1}{D_1 + D_2}, \frac{D_2}{D_1 + D_2}, \frac{D_3}{D_3 + D_4}, \frac{D_4}{D_3 + D_4} \right)$$

has the same distribution as

$$\left( \frac{G_1 + G_2}{G_1 + \dots + G_4}, \frac{G_3 + G_4}{G_1 + \dots + G_4}, \frac{G_1}{G_1 + G_2}, \frac{G_2}{G_1 + G_2}, \frac{G_3}{G_3 + G_4}, \frac{G_4}{G_3 + G_4} \right).$$

By the fact above,

$$\begin{aligned} & G_1 + G_2, \\ & G_3 + G_4, \\ & \left( \frac{G_1}{G_1 + G_2}, \frac{G_2}{G_1 + G_2} \right), \end{aligned}$$

and

$$\left( \frac{G_3}{G_3 + G_4}, \frac{G_4}{G_3 + G_4} \right)$$

are independent, and so

$$\begin{aligned} & \left( \frac{G_1 + G_2}{G_1 + \dots + G_4}, \frac{G_3 + G_4}{G_1 + \dots + G_4} \right), \\ & \left( \frac{G_1}{G_1 + G_2}, \frac{G_2}{G_1 + G_2} \right), \end{aligned}$$

and

$$\left( \frac{G_3}{G_3 + G_4}, \frac{G_4}{G_3 + G_4} \right)$$

are independent. □

## 6 Random recursive trees and nested Chinese restaurant processes

### 6.1 Random recursive trees from another encoding of permutations

Recall from the Introduction how the binary search tree process arises from a classical bijection between permutations of  $[n] := \{1, 2, \dots, n\}$  and a suitable class of labeled rooted trees. The random recursive tree process arises from a similar, but slightly less well-known, bijection that we now describe.

We begin with a definition similar to that of the complete binary tree in the Introduction. Denote by  $\mathbb{N}^* := \bigsqcup_{k=0}^{\infty} \mathbb{N}^k$  the set of finite tuples or *words* drawn from the alphabet  $\mathbb{N}$  (with the empty word  $\emptyset$  allowed). Write an  $\ell$ -tuple  $(v_1, \dots, v_\ell) \in \mathbb{N}^*$  more simply as  $v_1 \dots v_\ell$ . Define a directed graph with vertex set  $\mathbb{N}^*$  by declaring that if  $u = u_1 \dots u_k$  and  $v = v_1 \dots v_\ell$  are two words, then  $(u, v)$  is a directed edge (that is,  $u \rightarrow v$ ) if and only if  $\ell = k + 1$  and  $u_i = v_i$  for  $i = 1, \dots, k$ . Call this directed graph the *complete Harris-Ulam tree*. A *finite rooted Harris-Ulam tree* is a subset  $\mathbf{t}$  of  $\mathbb{N}^*$  with properties:

- $\emptyset \in \mathbf{t}$ ,
- if  $v = u_1 \dots u_k \in \mathbf{t}$ , then  $u_1 \dots u_j \in \mathbf{t}$  for  $1 \leq j \leq k - 1$  and  $u_1 \dots u_{k-1}m \in \mathbf{t}$  for  $1 \leq m \leq u_k - 1$ .

As in the binary case there is a canonical way to draw a finite rooted Harris-Ulam tree in the plane, see Figure 6 for an example. Further, we can similarly define a vertex  $u \in \mathbb{N}^*$  to be an external vertex of the tree  $\mathbf{t}$  if  $u \notin \mathbf{t}$  and if  $\mathbf{t} \sqcup \{u\}$  is again a Harris-Ulam tree. Note that, in contrast to the binary case, external vertices are now specified by their

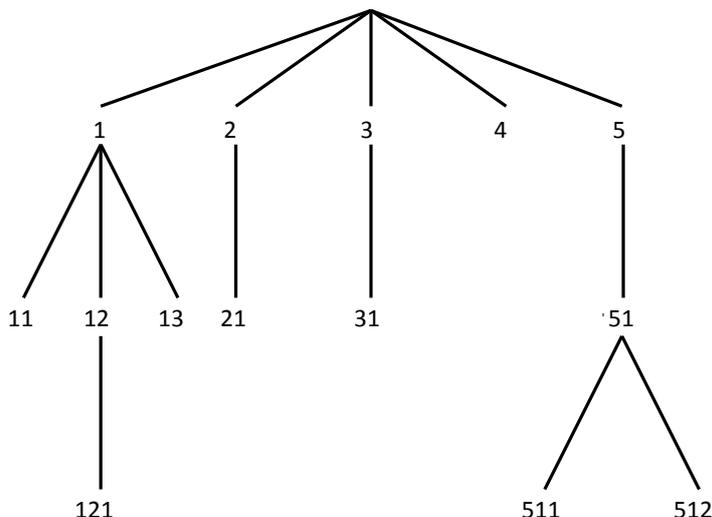


Figure 6: An example of a finite rooted Harris-Ulam tree.

immediate predecessor; in particular, a Harris-Ulam tree with  $n$  vertices has  $n$  external vertices.

Given a permutation  $\pi$  of  $[n]$ , set  $r(i) = \pi^{-1}(i)$  for  $1 \leq i \leq n$ . Construct a finite rooted Harris-Ulam tree with  $n + 1$  vertices labeled by  $[n] \cup \{0\}$  from  $r(1), \dots, r(n)$  recursively, as follows. Denote by  $\mathbf{t}_0$  the tree consisting of just the root  $\emptyset$  labeled with 0. Suppose for  $1 \leq i \leq n - 1$  that a tree  $\mathbf{t}_i$  with  $i$  vertices labeled by  $\{0, \dots, i - 1\}$  has already been defined. Assume that  $i = r(\ell)$ . If  $\{j : 1 \leq j < \ell, r(j) < i\} = \emptyset$ , set  $s := 0$ . Otherwise, set  $s := r(k)$ , where  $k := \max\{j : 1 \leq j < \ell, r(j) < i\}$ . Let  $u$  be the vertex in  $\mathbf{t}_i$  labeled by  $s$ . Put  $q := \max\{p \in \mathbb{N} : up \in \mathbf{t}_i\} + 1$ , adjoin the vertex  $uq$  to  $\mathbf{t}_i$  to create the tree  $\mathbf{t}_{i+1}$ , and label this new vertex with  $i$ .

For example, 1 is always the first child of 0 (occupying the vertex 1 in the complete Harris-Ulam tree) and 2 is either the second child of 0 (occupying the vertex 2 in the complete Harris-Ulam tree) or the first child of 1 (occupying the vertex 11 in the complete Harris-Ulam tree), depending on whether 2 appears before or after 1 in the list  $r(1), \dots, r(n)$ . See Figure 7 for an instance of the construction with  $n = 9$ .

Clearly,  $\pi$  can be reconstructed from the tree and its vertex labels.

As in the Introduction, given a sequence  $(U_n)_{n \in \mathbb{N}}$  of independent identically distributed random variables that are uniform on the unit interval  $[0, 1]$ , define a random permutation  $\Pi_n$  of  $[n]$  for each positive integer  $n$  by setting  $\Pi_n(k) = \#\{1 \leq \ell \leq n : U_\ell \leq U_k\}$ . Applying the bijection to  $\Pi_n$ , we obtain a random labeled rooted tree and a corresponding unlabeled rooted tree that we again denote by  $L_n$  and  $T_n$ , respectively. Both of these processes are Markov chains with simple transition probabilities. For example, given  $T_n$  we pick one of its  $n + 1$  vertices uniformly at random and connect a new vertex to it to form  $T_{n+1}$ . Thus,  $(T_n)_{n \in \mathbb{N}}$  is the simplest *random recursive tree* process (see, for example, [50] for a survey of such models).

As with the BST and DST processes, we think of building the sequence  $(T_n)_{n \in \mathbb{N}}$  by first building a growing sequence of finite rooted Harris-Ulam trees labeled with the values of the input sequence  $U_1, U_2, \dots$  and then ignoring the labels. The transition rule for the richer process takes a simple form: attach a new vertex labeled with  $U_{n+1}$  to the root if  $U_{n+1}$  is smaller than each of the previous variables  $U_1, \dots, U_n$ ; if not, then

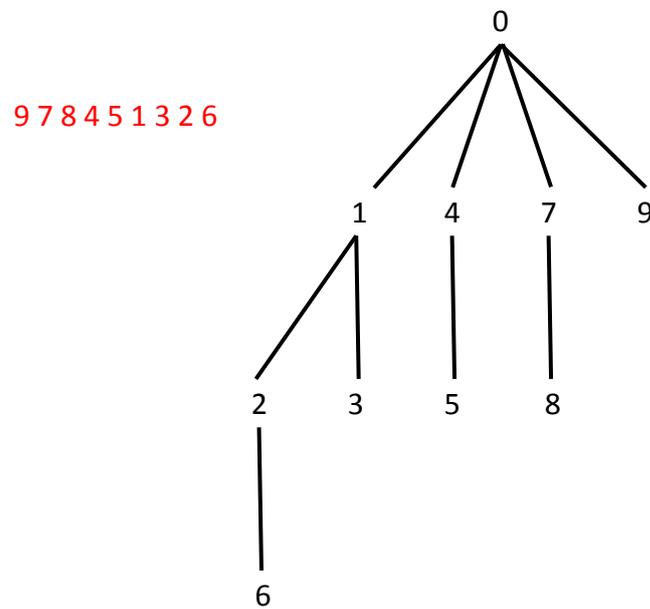


Figure 7: The labeled Harris-Ulam tree corresponding to the permutation of  $[9]$  with  $r(1), \dots, r(9) = 9, 7, 8, 4, 5, 1, 3, 2, 6$ . For the sake of clarity, the Harris-Ulam coding of the vertices as elements of  $\mathbb{N}^*$  is not shown. The correspondence between the labels from  $[9] \cup \{0\}$  and the coding of the vertices by elements of  $\mathbb{N}^*$  is  $0 \leftrightarrow \emptyset, 1 \leftrightarrow 1, 2 \leftrightarrow 11, 3 \leftrightarrow 12, 4 \leftrightarrow 2, 5 \leftrightarrow 21, 6 \leftrightarrow 111, 7 \leftrightarrow 3, 8 \leftrightarrow 31, 9 \leftrightarrow 4$ .

attach a new vertex labeled with  $U_{n+1}$  to the existing vertex that is labeled with the rightmost of the smaller elements. In contrast to the binary search tree situation, the labeled versions of the trees  $T_1, \dots, T_{n-1}$  can now be determined from the labeled version of  $T_n$ . However, if we remove the labels then we are in the same situation as in the BST case: the next tree is obtained by choosing an external vertex of the current tree uniformly at random and attaching it to the current tree.

## 6.2 Chinese restaurant processes

Suppose that in the tree  $T_n$  the root has  $k$  offspring. Let  $n_1, \dots, n_k$  denote the number of vertices in the subtrees rooted at each of these offspring, so that  $n_1 + \dots + n_k = n$ . Note that in constructing  $T_{n+1}$  from  $T_n$ , either a new vertex is attached to the  $j^{\text{th}}$  subtree with probability  $n_j/(n+1)$  or it is attached to the root and begins a new subtree with probability  $1/(n+1)$ . Thus, the manner in which the number and sizes of subtrees rooted at offspring of the root evolve is given by the number and sizes of tables in the simplest *Chinese restaurant process*: the  $n^{\text{th}}$  customer to enter the restaurant finds  $k$  tables in use with respective numbers of occupants  $n_1, \dots, n_k$  and the customer either sits at the  $j^{\text{th}}$  table with probability  $n_j/(n+1)$  or starts a new table with probability  $1/(n+1)$ .

It is clear from the construction of  $(T_n)_{n \in \mathbb{N}}$  that if we begin observing the subtree below one of the offspring of the root at the time the offspring first appears and only record the state of the subtree at each time it grows, then the resulting tree-valued process has the same dynamics as  $(T_n)_{n \in \mathbb{N}}$ . Iterating this observation, we see that we may think of  $(T_n)_{n \in \mathbb{N}}$  as an infinite collection of hierarchically nested Chinese restaurant processes and, in particular, that  $(T_n)_{n \in \mathbb{N}}$  arises as an instance of the trickle-down construction.

Rather than just investigate the Doob-Martin compactification of  $(T_n)_{n \in \mathbb{N}}$  we first recall the definition of Pitman's two-parameter family of processes to which the simple Chinese restaurant process belongs – see [41] for background and an extensive treatment of the properties of these processes. We then apply the trickle-down construction to build a tree-valued Markov chain that uses these more general processes as routing instructions. Analogous finitely nested Chinese restaurant processes have been used in hierarchical Bayesian inference [52].

A member of the family of Chinese restaurant processes is specified by two parameters  $\alpha$  and  $\theta$  that satisfy the constraints

$$\alpha < 0 \text{ and } \theta = -M\alpha \text{ for some } M \in \mathbb{N}$$

or

$$0 \leq \alpha < 1 \text{ and } \theta > -\alpha.$$

At time  $p$  the state of the process is a partition  $\mathcal{B}$  of the set  $[p]$  with  $\#\mathcal{B}$  blocks that are thought of as describing the composition of  $\#\mathcal{B}$  occupied tables. The next customer arrives at time  $p+1$  and decides either to sit at an empty table with probability

$$\frac{\theta + \alpha\#\mathcal{B}}{p + \theta},$$

thereby adjoining an extra block  $\{p+1\}$  to the partition and increasing the number of blocks by 1, or else to sit at an occupied table  $B \in \mathcal{B}$  of size  $\#B$  with probability

$$\frac{\#B - \alpha}{p + \theta},$$

thereby replacing the block  $B$  by the block  $B \cup \{p+1\}$  and leaving the number of blocks unchanged.

The probability that the partition of  $[q]$  we see at time  $q$  is  $\mathcal{B} = \{B_1, \dots, B_n\}$  with block sizes  $b_k = \#B_k$  is

$$\frac{(\theta + \alpha)(\theta + 2\alpha) \cdots (\theta + (n - 1)\alpha)}{(\theta + 1)(\theta + 2) \cdots (\theta + q - 1)} \prod_{k=1}^n (1 - \alpha)(2 - \alpha) \cdots (b_k - 1 - \alpha).$$

Note that if  $\alpha < 0$  and  $\theta = -M\alpha$  for some  $M \in \mathbb{N}$ , then, with probability one, the number of blocks in the partition is always at most  $M$ .

We are only interested in the process that records the number and size of the blocks. This process is also Markov. The probability that the random partition at time  $q$  has block sizes  $b_1, b_2, \dots, b_n$  is

$$\begin{aligned} & \frac{(\theta + \alpha)(\theta + 2\alpha) \cdots (\theta + (n - 1)\alpha)}{(\theta + 1)(\theta + 2) \cdots (\theta + q - 1)} \prod_{k=1}^n (1 - \alpha)(2 - \alpha) \cdots (b_k - 1 - \alpha) \\ & \times \binom{q - 1}{b_1 - 1} \binom{q - b_1 - 1}{b_2 - 1} \cdots \binom{q - b_1 - \cdots - b_{n-2} - 1}{b_{n-1} - 1}. \end{aligned}$$

The ordering of the blocks in this formula is their order of appearance:  $b_1$  is the size of the initial table,  $b_2$  is the size of the table that began receive customers next, and so on.

More generally, the probability that we go from the partition  $\mathcal{A} = \{A_1, \dots, A_m\}$  at time  $p$  to the partition  $\mathcal{B} = \{B_1, \dots, B_n\}$  at time  $q > p$  is

$$\begin{aligned} & \frac{(\theta + m\alpha)(\theta + 2\alpha) \cdots (\theta + (n - 1)\alpha)}{(\theta + p)(\theta + p + 1) \cdots (\theta + q - 1)} \\ & \times \prod_{k=1}^m (a_k - \alpha)(a_k + 1 - \alpha) \cdots (b_k - 1 - \alpha) \prod_{k=m+1}^n (1 - \alpha)(2 - \alpha) \cdots (b_k - 1 - \alpha). \end{aligned}$$

The corresponding probability that we go from a partition with block sizes  $a_1, \dots, a_m$  at time  $p$  to one with block sizes  $b_1, \dots, b_n$  at time  $q > p$  is

$$\begin{aligned} & \frac{(\theta + m\alpha)(\theta + 2\alpha) \cdots (\theta + (n - 1)\alpha)}{(\theta + p)(\theta + p + 1) \cdots (\theta + q - 1)} \\ & \times \prod_{k=1}^m (a_k - \alpha)(a_k + 1 - \alpha) \cdots (b_k - 1 - \alpha) \prod_{k=m+1}^n (1 - \alpha)(2 - \alpha) \cdots (b_k - 1 - \alpha) \\ & \times \binom{q - p}{b_1 - a_1} \binom{(q - b_1) - (p - a_1)}{b_2 - a_2} \\ & \times \cdots \binom{(q - b_1 - \cdots - b_{m-1}) - (p - a_1 - \cdots - a_{m-1})}{b_m - a_m} \\ & \times \binom{q - b_1 - \cdots - b_m - 1}{b_{m+1} - 1} \binom{q - b_1 - \cdots - b_{m+1} - 1}{b_{m+2} - 1} \\ & \times \cdots \binom{q - b_1 - \cdots - b_{n-2} - 1}{b_{n-1} - 1}. \end{aligned}$$

We can think of the block size process as a Markov chain with state space

$$\mathbf{E} := \{(0, 0, \dots)\} \sqcup \bigsqcup_{m \in \mathbb{N}} \mathbb{N}^m \times \{0\} \times \{0\} \times \cdots \subset \mathbb{N}_0^{\mathbb{N}}$$

when  $0 \leq \alpha < 1$  and  $\theta > -\alpha$ , or

$$\mathbf{E} := \{(0, 0, \dots)\} \sqcup \bigsqcup_{m=1}^M \mathbb{N}^m \times \{0\}^{M-m} \subset \mathbb{N}_0^M$$

when  $\alpha < 0$  and  $\theta = -M\alpha$  for some  $M \in \mathbb{N}$ . For two states  $\mathbf{a} = (a_1, \dots, a_m, 0, 0, \dots) \in \mathbf{E}$  and  $\mathbf{b} = (b_1, \dots, b_n, 0, 0, \dots) \in \mathbf{E}$  with  $1 \leq m \leq n$ ,  $b_i \geq a_i > 0$  when  $1 \leq i \leq m$ ,  $b_j > 0$  when  $m + 1 \leq j \leq n$ ,  $\sum_{i=1}^m a_i = p$ , and  $\sum_{j=1}^n b_j = q$ , the Martin kernel is

$$\begin{aligned}
 K(\mathbf{a}, \mathbf{b}) &= \frac{(\theta + 1)(\theta + 2) \cdots (\theta + p - 1)}{(\theta + \alpha)(\theta + 2\alpha) \cdots (\theta + (m - 1)\alpha)} \\
 &\times \left[ \prod_{k=1}^m (1 - \alpha)(2 - \alpha) \cdots (a_k - 1 - \alpha) \right]^{-1} \\
 &\times \frac{(q - p)!}{(b_1 - a_1)!((q - p) - (b_1 - a_1))!} \frac{(b_1 - 1)!(q - b_1)!}{(q - 1)!} \\
 &\times \frac{((q - p) - (b_1 - a_1))!}{(b_2 - a_2)!((q - p) - (b_1 - a_1) - (b_2 - a_2))!} \frac{(b_2 - 1)!(q - b_1 - b_2)!}{(q - b_1 - 1)!} \\
 &\dots \\
 &\times \frac{((q - p) - (b_1 - a_1) - \cdots - (b_{m-1} - a_{m-1}))!}{(b_m - a_m)!((q - p) - (b_1 - a_1) - \cdots - (b_m - a_m))!} \\
 &\times \frac{(b_m - 1)!(q - b_1 - \cdots - b_m)!}{(q - b_1 - \cdots - b_{m-1} - 1)!}.
 \end{aligned}$$

This expression can be rearranged to give

$$\begin{aligned}
 &\frac{(\theta + 1)(\theta + 2) \cdots (\theta + p - 1)}{(\theta + \alpha)(\theta + 2\alpha) \cdots (\theta + (m - 1)\alpha)} \left[ \prod_{k=1}^m (1 - \alpha)(2 - \alpha) \cdots (a_k - 1 - \alpha) \right]^{-1} \\
 &\times \frac{(q - p)!}{((q - p) - (b_1 - a_1) - \cdots - (b_m - a_m))!} \\
 &\times \frac{(b_1 - 1)! \cdots (b_m - 1)!}{(b_1 - a_1)! \cdots (b_m - a_m)!} \\
 &\times \frac{(q - b_1)!(q - b_1 - b_2)! \cdots (q - b_1 - b_2 - \cdots - b_m)!}{(q - 1)!(q - b_1 - 1)! \cdots (q - b_1 - b_2 - \cdots - b_{m-1} - 1)!} \\
 &= \frac{(\theta + 1)(\theta + 2) \cdots (\theta + p - 1)}{(\theta + \alpha)(\theta + 2\alpha) \cdots (\theta + (m - 1)\alpha)} \left[ \prod_{k=1}^m (1 - \alpha)(2 - \alpha) \cdots (a_k - 1 - \alpha) \right]^{-1} \\
 &\times [(q - p + 1)(q - p + 2) \cdots (q - 1)]^{-1} \\
 &\times \prod_{k=1}^m (b_k - a_k + 1) \cdots (b_k - 1) \prod_{k=1}^{m-1} \left( q - \sum_{\ell=1}^k b_\ell \right).
 \end{aligned}$$

If  $(\mathbf{b}_N)_{N \in \mathbb{N}} = ((b_{N,1}, b_{N,2}, \dots))_{N \in \mathbb{N}}$  is a sequence from  $\mathbf{E}$  such that

$$\#\{N \in \mathbb{N} : \mathbf{b}_N = \mathbf{b}\} < \infty$$

for all  $\mathbf{b} \in \mathbf{E}$ , then  $\lim_{N \rightarrow \infty} \sum_{k=1}^{\infty} b_{N,k} = \infty$ . In this case, it is not hard to see that  $\lim_{N \rightarrow \infty} K(\mathbf{a}, \mathbf{b}_N)$  exists for  $\mathbf{a} \in \mathbf{E}$  if and only if

$$\lim_{N \rightarrow \infty} \frac{b_{N,k}}{\sum_{\ell=1}^{\infty} b_{N,\ell}} =: \rho_k$$

exists for all  $k \in \mathbb{N}$ . Furthermore, for  $\mathbf{a} = (a_1, a_2, \dots, a_m, 0, \dots)$  as above

$$\begin{aligned} \lim_{N \rightarrow \infty} K(\mathbf{a}, \mathbf{b}_N) &= \frac{(\theta + 1)(\theta + 2) \cdots (\theta + p - 1)}{(\theta + \alpha)(\theta + 2\alpha) \cdots (\theta + (m - 1)\alpha)} \\ &\quad \times \left[ \prod_{k=1}^m (1 - \alpha)(2 - \alpha) \cdots (a_k - 1 - \alpha) \right]^{-1} \\ &\quad \times \rho_1^{\alpha_1 - 1} \cdots \rho_m^{\alpha_m - 1} \\ &\quad \times (1 - \rho_1)(1 - \rho_1 - \rho_2) \cdots (1 - \rho_1 - \rho_2 - \cdots - \rho_{m-1}) \\ &=: K(\mathbf{a}, \rho). \end{aligned}$$

Note that  $\lim_{N \rightarrow \infty} K(\mathbf{a}, \mathbf{b}_N)$  exists for all  $\mathbf{a} \in \mathbf{E}$  if and only if the limit exists for all  $\mathbf{a} \in \mathbf{E}$  of the form  $(1, \dots, 1, 0, 0, \dots)$  (that is, for all  $\mathbf{a} \in \mathbf{E}$  with entries in  $\{0, 1\}$ ). Note also that the extended Martin kernel has the property that

$$K(\mathbf{a}, \rho) = 0 \Leftrightarrow \begin{cases} a_k \geq 1 \text{ for some } k \text{ with } \sum_{j=1}^{k-1} \rho_j = 1, \\ a_k \geq 2 \text{ for some } k \text{ with } \rho_k = 0. \end{cases}$$

Recall that if  $\mathbf{a}$  is as above, then the transition probabilities of the block size process are given by

$$P(\mathbf{a}, \mathbf{b}) = \begin{cases} \frac{\theta + \alpha m}{\theta + p}, & \text{if } \mathbf{b} = (a_1, \dots, a_m, 1, 0, \dots), \\ \frac{a_k - \alpha}{\theta + p}, & \text{if } \mathbf{b} = (a_1, \dots, a_{k+1}, \dots, a_m, 0, \dots). \end{cases}$$

The Doob  $h$ -transform corresponding to the regular function  $h_\rho := K(\cdot, \rho)$  therefore has transition probabilities given by

$$\begin{aligned} P^{(h_\rho)}(\mathbf{a}, \mathbf{b}) &= \begin{cases} \frac{\theta + \alpha m}{\theta + p} \frac{\theta + p}{\theta + m\alpha} (1 - \rho_1 - \cdots - \rho_m), & \text{if } \mathbf{b} = (a_1, \dots, a_m, 1, 0, \dots), \\ \frac{a_k - \alpha}{\theta + p} (\theta + p) (a_k - \alpha)^{-1} \rho_k, & \text{if } \mathbf{b} = (a_1, \dots, a_k + 1, \dots, a_m, 0, \dots). \end{cases} \end{aligned}$$

That is,

$$\begin{aligned} P^{(h_\rho)}(\mathbf{a}, \mathbf{b}) &= \begin{cases} (1 - \rho_1 - \cdots - \rho_m), & \text{if } \mathbf{b} = (a_1, \dots, a_m, 1, 0, \dots), \\ \rho_k, & \text{if } \mathbf{b} = (a_1, \dots, a_k + 1, \dots, a_m, 0, \dots). \end{cases} \end{aligned}$$

Note that the parameters  $\alpha$  and  $\theta$  do not appear in this expression for the transition probabilities. It follows that for a given  $M$  the block size chains all arise as Doob  $h$ -transforms of each other.

We can build a Markov chain  $(W_n)_{n \in \mathbb{N}_0}$  with transition matrix  $P^{(h_\rho)}$  and initial state  $\mathbf{c}$  as follows. Let  $(V_n)_{n \in \mathbb{N}}$  be a sequence of independent identically distributed random variables taking values in  $[M] \cup \{\infty\}$  with  $\mathbb{P}\{V_n = k\} = \rho_k$  for  $k \in [M]$  and  $\mathbb{P}\{V_n = \infty\} = 1 - \sum_{\ell} \rho_\ell$  (the latter probability is always 0 when  $M$  is finite). Define  $(W_n)_{n \in \mathbb{N}_0}$  inductively by setting  $W_0 = \mathbf{c}$  and, writing  $N_n := \inf\{j \in [M] : W_{nj} = 0\}$  with the usual convention that  $\inf \emptyset = \infty$ ,

$$W_{n+1} = \begin{cases} (W_{n1}, \dots, W_{nN_n}, 1, 0, \dots), & \text{if } V_{n+1} > N_n, \\ (W_{n1}, \dots, W_{nk} + 1, \dots, W_{nN_n}, 0, \dots), & \text{if } V_{n+1} = k \leq N_n, \end{cases}$$

for  $n \geq 0$ . It is clear from this construction and Kolmogorov's zero-one law that the tail  $\sigma$ -field of the chain is trivial, and so the regular function  $h_\rho$  is extremal.

### 6.3 Chinese restaurant trees

Fix an admissible pair of parameters  $\alpha$  and  $\theta$  for the two-parameter Chinese restaurant process. Set  $M := \infty$  when  $0 \leq \alpha < 1$  and  $M := -\theta/\alpha \in \mathbb{N}$  when  $\alpha < 0$ . Put  $[M] := \mathbb{N}$  for  $M = \infty$  and  $[M] := \{1, \dots, M\}$  otherwise.

Consider the trickle-down construction with the following ingredients. The underlying directed acyclic graph  $\mathbf{I}$  has vertex set  $[M]^* := \bigsqcup_{k=0}^{\infty} [M]^k$ , the set of finite tuples or words drawn from the alphabet  $[M]$  (with the empty word  $\emptyset$  allowed) and directed edges are defined in a manner analogous to that in Subsection 6.1 – when  $[M] = \mathbb{N}$  we just recover the complete Harris-Ulam tree of Subsection 6.1. Thus,  $\mathbf{I}$  is a tree rooted at  $\emptyset$  in which we may identify  $\beta(u)$ , the set of offspring of vertex  $u \in \mathbf{I}$ , with  $[M]$  for every vertex  $u$ . With this identification, we take the routing chain for every vertex to be the Chinese restaurant block size process with parameters  $\alpha$  and  $\theta$ .

We may think of the state space  $\mathbf{S}$  of the trickle-down chain  $(X_n)_{n \in \mathbb{N}_0}$  as the set of finite subsets  $\mathbf{t}$  of  $\mathbf{I}$  with the property that if a word  $v = v_1 \dots v_\ell \in \mathbf{t}$ , then  $v_1 \dots v_{\ell-1} \in \mathbf{t}$  and  $v_1 \dots v_{\ell-1}k \in \mathbf{t}$  for  $1 \leq k < v_\ell$ . That is, when  $[M] = \mathbb{N}$  we may think of  $\mathbf{S}$  as the set of finite rooted Harris-Ulam trees from Subsection 6.1, and when  $M$  is finite we get an analogous collection in which each individual has at most  $M$  offspring.

The partial order on  $\mathbf{I} = [M]^*$  is the one we get by declaring that  $u \leq v$  for two words  $u, v \in \mathbf{I}$  if and only if  $u = u_1 \dots u_k$  and  $v = v_1 \dots v_\ell$  with  $k \leq \ell$  and  $u_1 \dots u_k = v_1 \dots v_k$ , just as for the complete binary tree. By analogy with the notation introduced in Example 4.3 for finite rooted binary trees, write  $\#\mathbf{t}(u) := \#\{v \in \mathbf{t} : u \leq v\}$  for  $\mathbf{t} \in \mathbf{S}$  and  $u \in [M]^*$ .

It follows from the discussion in Subsection 6.2 that Hypothesis 4.6 holds. We may identify the set  $\mathbf{I}_\infty$  with

$$[M]^\infty \sqcup \bigsqcup_{k=0}^{\infty} ([M]^k \times \{\diamond\}^\infty)$$

For each vertex  $u \in \mathbf{I}$  the collection  $\mathcal{S}^u$  consists of all probability measures on  $\beta(u)$  when  $M$  is finite and all subprobability measures on  $\beta(u)$  when  $M = \infty$ . We may therefore identify  $\partial\mathbf{S}$  with the probability measures on  $\mathbf{I}_\infty$  that assign all of their mass to  $[M]^\infty$  when  $M$  is finite and with the set of all probability measures on  $\mathbf{I}_\infty$  when  $M = \infty$ . We may extend the partial order by declaring that  $u < v$  for  $u \in \mathbf{I} = [M]^*$  and  $v \in \mathbf{I}_\infty = [M]^\infty \sqcup \bigsqcup_{k=0}^{\infty} ([M]^k \times \{\diamond\}^\infty)$  if and only if  $u = u_1 \dots u_k$  and  $v = v_1 v_2 \dots$  with  $u_1 \dots u_k = v_1 \dots v_k$ .

The following result summarizes the salient conclusions of the above discussion.

**Theorem 6.1.** *Consider the Chinese restaurant tree process with parameters  $(\alpha, \theta)$ , where  $\alpha < 0$  and  $\theta = -M\alpha$  for some  $M \in \mathbb{N}$  or  $0 \leq \alpha < 1$  and  $\theta > -\alpha$ , in which case we define  $M = \infty$ . We may identify the state space  $\mathbf{S}$  of this process as the set of finite rooted Harris-Ulam trees where the vertices are composed of digits drawn from  $[M]$ . When  $M < \infty$  (resp.  $M = \infty$ ), the Doob-Martin boundary  $\partial\mathbf{S}$  is homeomorphic to the space of probability measures on  $[M]^\infty$  (resp.  $[M]^\infty \sqcup \bigsqcup_{k=0}^{\infty} ([M]^k \times \{\diamond\}^\infty)$ ) equipped with the topology of weak convergence. With this identification, a sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}}$  of finite rooted Harris-Ulam trees converges in the topology of the Doob-Martin compactification  $\bar{\mathbf{S}}$  to the (sub)probability measure  $\mu$  in the Doob-Martin boundary  $\partial\mathbf{S}$  if and only if  $\lim_{n \rightarrow \infty} \#\mathbf{t}_n = \infty$  and*

$$\lim_{n \rightarrow \infty} \frac{\#\mathbf{t}_n(u)}{\#\mathbf{t}_n} = \mu \left\{ v \in [M]^\infty \sqcup \bigsqcup_{k=0}^{\infty} ([M]^k \times \{\diamond\}^\infty) : u < v \right\}$$

for all  $u \in [M]^*$ .

**Example 6.2.** *Suppose that  $M = \infty$ . Consider the sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}}$  of finite rooted Harris-Ulam trees given by  $\mathbf{t}_n := \{\emptyset, 1, 2, \dots, n, 21, 211, \dots, 21^{n-1}\}$ , where the notation*

$21^k$  indicates 2 followed by  $k$  1s. This sequence of trees converges in the topology of  $\bar{\mathbf{S}}$  to the probability measures on  $[M]^\infty \sqcup \bigsqcup_{k=0}^\infty ([M]^k \times \{\diamond\}^\infty)$  that puts mass  $\frac{1}{2}$  at the point  $\diamond\diamond\dots$  and mass  $\frac{1}{2}$  at the point  $2111\dots$

**Remark 6.3.** The calculations of the extended Martin kernel and Doob  $h$ -transform transition probabilities associated with a given  $\mu \in \partial\mathbf{S}$  are straightforward but notationally somewhat cumbersome, so we omit them. They show that there is the following “trickle-up” construction of a Markov chain  $(W_n)_{n \in \mathbb{N}_0}$  with initial state  $\mathbf{w} \in \mathbf{S}$  and the  $h$ -transform transition probabilities (compare the analogous construction for the Chinese restaurant process itself in Subsection 6.2).

Let  $(V^n)_{n \in \mathbb{N}}$  be a sequence of independent, identically distributed  $\mathbf{I}_\infty$ -valued random variables with common distribution  $\mu$ . Suppose that  $\mathbf{S}$ -valued random variables

$$\mathbf{w} =: W_0 \subset \dots \subset W_n$$

have already been defined. Put

$$H(n+1) := \max\{h \in \mathbb{N} : V_1^{n+1} \dots V_h^{n+1} \in W_n\},$$

with the convention  $\max \emptyset = 0$ , and

$$M(n+1) := \max\{m \in \mathbb{N} : V_1^{n+1} \dots V_{H(n+1)}^{n+1} m \in W_n\},$$

again with the convention  $\max \emptyset = 0$ . Set

$$W_{n+1} := W_n \cup \{V_1^{n+1} \dots V_{H(n+1)}^{n+1} (M(n+1) + 1)\}.$$

For example, if  $\mathbf{w} = \emptyset$  and  $\mu$  is the unit point mass at the sequence  $\diamond\diamond\dots$ , then

$$W_n = \{\emptyset, 1, \dots, n\}$$

for  $n \geq 1$ ; that is,  $W_n$  consists of the root  $\emptyset$  and the first  $n$  children of the root.

It is clear from the Kolmogorov zero-one law that the tail  $\sigma$ -field of  $(W_n)_{n \in \mathbb{N}_0}$  is trivial for any  $\mu$ , and so any  $\mu$  is extremal.

**Remark 6.4.** By analogy with the definition of the BST process in Section 5, we define  $T_n$  to be the set of vertices occupied by time  $n$  (so that  $T_0 = \{\emptyset\}$ ). Put, for each vertex  $u$ ,  $T_n(u) := \{v \in T_n : u \leq v\}$ . The distribution of the random probability measure  $R$  on  $[M]^\infty$  defined by  $R\{w \in [M]^\infty : u < w\} := \lim_{n \rightarrow \infty} \#T_n(u) / \#T_n$ ,  $u \in [M]^*$ , may be derived from known properties of the two-parameter Chinese restaurant process (see, for example, Theorem 3.2 of [41]). For  $v \in [M]^*$  put

$$(U_{v1}, U_{v2}, U_{v3}, \dots) := (B_{v1}, (1 - B_{v1})B_{v2}, (1 - B_{v1})(1 - B_{v2})B_{v3}, \dots),$$

where the random variables  $B_{vk}$ ,  $v \in [M]^*$ ,  $k \in [M]$ , are independent and  $B_{uk}$  has the beta distribution with parameters  $(1 - \alpha, \theta + k\alpha)$ . That is, the sequence  $(U_{vk})_{k \in [M]}$  has a Griffiths–Engen–McCloskey (GEM) distribution with parameters  $(\alpha, \theta)$ . Then,  $R$  is distributed as the random probability measure on  $[M]^\infty$  that for each  $u \in [M]^*$  assigns mass  $\prod_{\emptyset < v \leq u} U_v$  to the set  $\{w \in [M]^\infty : u < w\}$ .

## 7 Mallows chains

### 7.1 Mallows’ $\phi$ model for random permutations and the associated tree

The  $\phi$  model of Mallows [39] produces a random permutation of the set  $[n]$  for some integer  $n \in \mathbb{N}$ . One way to describe the model is the following.

We place the elements of  $[n]$  successively into  $n$  initially vacant “slots” labeled by  $[n]$  to obtain a permutation of  $[n]$  (if the number  $i$  goes into slot  $j$ , then the permutation sends  $i$  to  $j$ ). To begin with, each slot is equipped with a Bernoulli random variable. These random variables are obtained by taking  $n$  independent Bernoulli random variables with common success probability  $0 < p < 1$  and conditioning on there being at least 1 success. The number 1 is placed in the first slot for which the associated Bernoulli random variable is a success. Thus, the probability that there are  $k$  vacant slots to the left of 1 is

$$\frac{(1-p)^k p}{1-(1-p)^n}, \quad 0 \leq k \leq n-1.$$

Now equip the remaining  $n-1$  vacant slots (that is, every slot except the one in which 1 was placed) with a set of Bernoulli random variables that is independent of the first set. These random variables are obtained by taking  $n-1$  independent Bernoulli random variables with common success probability  $p$  and conditioning on there being at least 1 success. Place the number 2 in the first vacant slot for which the associated Bernoulli is a success. The probability that there are  $k$  vacant slots to the left of 2 is

$$\frac{(1-p)^k p}{1-(1-p)^{n-1}}, \quad 0 \leq k \leq n-2.$$

Continue in this fashion until all the slots have been filled.

The analogous procedure can be used to produce a permutation of  $\mathbb{N}$ . Now the procedure begins with infinitely many slots labeled by  $\mathbb{N}$ , and at each stage there is no need to condition on the almost sure event that there is at least one success. After each  $m \in \mathbb{N}$  is inserted, the current number of vacant slots to the left of the slot in which  $m$  is placed is distributed as the number of failures before the first success in independent Bernoulli trials with common success probability  $p$ , and these random variables are independent. We note that this distribution on permutations of  $\mathbb{N}$  appears in [23] in connection with  $q$ -analogues of de Finetti’s theorem.

Suppose now that  $\pi$  is a permutation of the set  $S$ , where  $S = [n]$  or  $S = \mathbb{N}$ . Let  $I(\pi) := \pi(1)$ . That is, if we think of  $\pi$  as a list of the elements of  $S$  in some order, then  $I(\pi)$  is the index of 1. Put  $S^L(\pi) := \{i : \pi(i) < \pi(1)\}$  and  $S^R(\pi) := \{i : \pi(i) > \pi(1)\}$ . Note that  $\pi$  maps  $S^L(\pi)$  to  $\{1, \dots, I(\pi) - 1\}$  and  $S^R(\pi)$  to  $I(\pi) + \{1, \dots, n - I(\pi)\}$  or  $I(\pi) + \mathbb{N}$ , and that  $S^L(\pi)$  (respectively,  $S^R(\pi)$ ) is the set of elements of  $S$  that appear before (respectively, after) 1 in the ordered listing of  $S$  defined by  $\pi$ .

If  $S = [n]$ , write  $\psi^L(\pi)$  for the unique increasing bijection from  $\{1, \dots, I(\pi) - 1\}$  to  $S^L(\pi)$  and  $\psi^R(\pi)$  for the unique increasing bijection from  $\{1, \dots, n - I(\pi)\}$  to  $S^R(\pi)$ . If  $S = \mathbb{N}$ , define  $\psi^L(\pi)$  and  $\psi^R(\pi)$  similarly, except that now  $\psi^R(\pi)$  maps  $\mathbb{N}$  to  $S^R(\pi)$ .

Define permutations  $\sigma^L(\pi)$  and  $\sigma^R(\pi)$  of  $\{1, \dots, I(\pi) - 1\}$  and  $\{1, \dots, n - I(\pi)\}$  (if  $S = [n]$ ) or  $\{1, \dots, I(\pi) - 1\}$  and  $\mathbb{N}$  (if  $S = \mathbb{N}$ ) by requiring that  $\pi$  restricted to  $S^L(\pi)$  is  $\psi^L(\pi) \circ \sigma^L(\pi) \circ (\psi^L(\pi))^{-1}$  and that  $\pi$  restricted to  $S^R(\pi)$  is  $\psi^R(\pi) \circ \sigma^R(\pi) \circ (\psi^R(\pi))^{-1}$ . In other words,  $\sigma^L(\pi)(i)$  is the index of the  $i^{\text{th}}$  smallest element of  $S^L(\pi)$  in the ordered listing of  $S$  defined by  $\pi$ , and  $I(\pi) + \sigma^R(\pi)(i)$  is the index of the  $i^{\text{th}}$  smallest element of  $S^R(\pi)$  in the ordered listing of  $S$  defined by  $\pi$ .

Note that  $\pi$  is uniquely specified by the objects  $I(\pi)$ ,  $S^L(\pi)$ ,  $S^R(\pi)$ ,  $\sigma^L(\pi)$ , and  $\sigma^R(\pi)$ .

The following lemma is immediate from the construction of the Mallows model.

**Lemma 7.1.** *Suppose that  $\Pi$  is a random permutation of either  $[n]$  or  $\mathbb{N}$  that is distributed according to the Mallows model with parameter  $p$ . Then, conditional on*

$$(I(\Pi), S^L(\Pi), S^R(\Pi)),$$

*the permutations  $\sigma^L(\Pi)$  and  $\sigma^R(\Pi)$  are independent and distributed according to the Mallows model with parameter  $p$ .*

Recall from the description of the BST process in the Introduction how it is possible to construct from a permutation  $\pi$  of  $[n]$  a subtree of the complete binary tree  $\{0, 1\}^*$  that contains the root  $\emptyset$  and has  $n$  vertices. The procedure actually produces a tree labeled with the elements of  $[n]$ , but we are only interested in the underlying unlabeled tree. Essentially the same construction produces an infinite rooted binary tree labeled with  $\mathbb{N}$  from a permutation  $\pi$  of  $\mathbb{N}$ . This tree has the property that if a vertex  $u = u_1 \dots u_k$  belongs to the tree, then there only finitely many vertices  $v$  such that  $u_1 \dots u_k 0 \leq v$ .

The following result is immediate from Lemma 7.1 and the recursive nature of the procedure that produces a rooted subtree of  $\{0, 1\}^*$  from a permutation.

**Proposition 7.2.** *Let  $(X_n)_{n \in \mathbb{N}_0}$  be the Markov chain that results from the trickle-down construction applied when the directed graph  $\mathbf{I}$  is the infinite complete binary tree  $\{0, 1\}^*$  and all the routing chains have the common transition matrix  $Q$  on the state space  $\mathbb{N}_0 \times \mathbb{N}_0$ , where*

$$Q((i, 0), (i + 1, 0)) := (1 - p), \quad \text{for all } i \geq 0,$$

$$Q((i, 0), (i, 1)) := p, \quad \text{for all } i \geq 0,$$

and

$$Q((i, j), (i, j + 1)) := 1, \quad \text{for all } i \geq 0 \text{ and } j \geq 1.$$

We may regard  $(X_n)_{n \in \mathbb{N}_0}$  as a Markov chain taking values in the set of finite subtrees of  $\{0, 1\}^*$  that contain the root  $\emptyset$ , in which case  $\{\emptyset\} = X_0 \subseteq X_1 \subseteq \dots$  and  $X_\infty := \bigcup_{n \in \mathbb{N}_0} X_n$  is an infinite subtree of  $\{0, 1\}^*$  that contains  $\emptyset$ . Then,  $X_\infty$  has the same distribution as the random tree constructed from a random permutation of  $\mathbb{N}$  that is distributed according to the Mallows model with parameter  $p$ .

We call the Markov chain  $(X_n)_{n \in \mathbb{N}_0}$  of Proposition 7.2 the *Mallows tree process*.

### 7.2 Mallows urns

Consider the Markov chain on  $\mathbb{N}_0 \times \mathbb{N}_0$  with transition matrix  $Q$  introduced in Proposition 7.2. We call this chain the *Mallows urn*, because its role as a routing chain for the Mallows tree process is similar to that played by the Pólya urn in the construction of the BST process. When started from  $(0, 0)$ , a sample path of the Mallows urn process looks like  $(0, 0), (1, 0), \dots, (K, 0), (K, 1), (K, 2), \dots$ , where  $\mathbb{P}\{K = k\} = (1 - p)^k p$  for  $k \in \mathbb{N}_0$ .

The probability that the Mallows urn process visits the state  $(k, \ell)$  starting from the state  $(i, j)$  is

$$\begin{cases} (1 - p)^{k-i}, & \text{if } i \leq k, j = 0 \text{ and } \ell = 0, \\ (1 - p)^{k-i} p, & \text{if } i \leq k, j = 0 \text{ and } \ell \geq 1, \\ 1, & \text{if } i = k \text{ and } 1 \leq j \leq \ell, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, the probability that the process visits  $(k, \ell)$  starting from  $(0, 0)$  is

$$\begin{cases} (1 - p)^k, & \text{if } \ell = 0, \\ (1 - p)^k p, & \text{if } \ell \geq 1. \end{cases}$$

Taking, as usual,  $(0, 0)$  as the reference state, the Martin kernel for the Mallows urn process is thus

$$K((i, j), (k, \ell)) := \begin{cases} (1 - p)^{-i}, & \text{if } i \leq k, j = 0 \text{ and } \ell = 0, \\ (1 - p)^{-i}, & \text{if } i \leq k, j = 0 \text{ and } \ell \geq 1, \\ (1 - p)^{-k} p^{-1}, & \text{if } i = k \text{ and } 1 \leq j \leq \ell, \\ 0, & \text{otherwise,} \end{cases}$$

or, equivalently,

$$K((i, j), (k, \ell)) = \begin{cases} (1 - p)^{-i}, & \text{if } i \leq k \text{ and } j = 0, \\ (1 - p)^{-i}p^{-1}, & \text{if } i = k \text{ and } 1 \leq j \leq \ell, \\ 0, & \text{otherwise.} \end{cases} \quad (7.1)$$

It follows that if  $((k_n, \ell_n))_{n \in \mathbb{N}_0}$  is a sequence for which  $k_n + \ell_n \rightarrow \infty$  then, in order for the sequence  $(K((i, j), (k_n, \ell_n)))_{n \in \mathbb{N}_0}$  to converge, it must either be that  $k_n = k_\infty$  for some  $k_\infty$  for all  $n$  sufficiently large and  $\ell_n \rightarrow \infty$ , in which case the limit is

$$\begin{cases} (1 - p)^{-i}, & \text{if } i \leq k_\infty \text{ and } j = 0, \\ (1 - p)^{-i}p^{-1}, & \text{if } i = k_\infty \text{ and } j \geq 1, \\ 0, & \text{otherwise,} \end{cases}$$

or that  $k_n \rightarrow \infty$  with no restriction on  $\ell_n$ , in which case the limit is

$$\begin{cases} (1 - p)^{-i}, & \text{if } j = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, the Doob-Martin compactification  $\overline{\mathbb{N}_0 \times \mathbb{N}_0}$  of the state space of the Mallows urn process is such that the Doob-Martin boundary

$$\partial(\mathbb{N}_0 \times \mathbb{N}_0) := \overline{\mathbb{N}_0 \times \mathbb{N}_0} \setminus \mathbb{N}_0 \times \mathbb{N}_0$$

can be identified with  $\mathbb{N}_0 \cup \{\infty\}$ , the usual one-point compactification of  $\mathbb{N}_0$ .

With this identification, the state space of the  $h$ -transformed process corresponding to the boundary point  $k \in \mathbb{N}_0$  is  $\{(0, 0), (1, 0), \dots, (k, 0)\} \cup \{(k, 1), (k, 2), \dots\}$  and the transition probabilities are

$$Q^h((i, 0), (i + 1, 0)) = ((1 - p)^{-i})^{-1}(1 - p)(1 - p)^{-(i+1)} = 1, \quad \text{for } 0 \leq i \leq k - 1,$$

$$Q((k, 0), (k, 1)) = ((1 - p)^{-i})^{-1}p(1 - p)^{-i}p^{-1} = 1,$$

and

$$Q((k, j), (k, j + 1)) = ((1 - p)^{-i}p^{-1})^{-1}1(1 - p)^{-i}p^{-1} = 1, \quad \text{for all } j \geq 1.$$

Thus, a realization of the  $h$ -transformed process starting from  $(0, 0)$  is the deterministic path  $(0, 0), (1, 0), \dots, (k, 0), (k, 1), (k, 2), \dots$

Similarly, the state space of the  $h$ -transformed process corresponding to the boundary point  $\infty$  is  $\{(0, 0), (1, 0), (2, 0), \dots\}$  and a realization of the  $h$ -transformed process starting from  $(0, 0)$  is the deterministic path  $(0, 0), (1, 0), (2, 0), \dots$

### 7.3 Mallows tree process

Recall from Example 2.3 that we may identify the state space  $\mathbf{S}$  of the Mallows tree process with the set of finite subtrees of the complete binary tree  $\mathbf{I} = \{0, 1\}^*$  that contain the root  $\emptyset$ , and with this identification the partial order  $\preceq$  is just subset containment.

Consider  $s$  in  $\mathbf{S}$  and a sequence  $(t_n)_{n \in \mathbb{N}_0}$  from  $\mathbf{S}$  such that  $\#t_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Given a vertex  $u$  of  $\{0, 1\}^*$  write, as in Section 5,  $\#s(u) := \{v \in s : u \preceq v\}$  and define  $\#t_n(u)$  similarly. Note that in this setting the consistency condition (2.3) becomes

$$(\#s(u) - 1)_+ = \#s(u0) + \#s(u1)$$

and

$$(\#t_n(u) - 1)_+ = \#t_n(u0) + \#t_n(u1).$$

Write

$$L(\mathbf{s}) := \sum_{u \in \{0,1\}^*} \#\mathbf{s}(u0).$$

When  $\mathbf{s} \subseteq \mathbf{t}_n$ , put

$$M(\mathbf{s}, \mathbf{t}_n) := \#\{u \in \{0,1\}^* : \#\mathbf{s}(u0) = \#\mathbf{t}_n(u0), \#\mathbf{s}(u1) \geq 1\}$$

and

$$I(\mathbf{s}, \mathbf{t}_n) := \begin{cases} 1, & \text{if } \#\mathbf{s}(u0) = \#\mathbf{t}_n(u0) \text{ whenever } \#\mathbf{s}(u1) \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

From Corollary 4.2 and (7.1), the Martin kernel of the Mallows tree process is

$$K(\mathbf{s}, \mathbf{t}_n) := \begin{cases} (1-p)^{-L(\mathbf{s})} p^{-M(\mathbf{s}, \mathbf{t}_n)} I(\mathbf{s}, \mathbf{t}_n), & \text{if } \mathbf{s} \subseteq \mathbf{t}_n, \\ 0, & \text{otherwise.} \end{cases}$$

Note that if  $\mathbf{s} \subseteq \mathbf{t}_n$  and  $\#\mathbf{s}(u0) = \#\mathbf{t}_n(u0)$ , then  $\{v \in \mathbf{s} : u0 \leq v\} = \{v \in \mathbf{t}_n : u0 \leq v\}$ . Therefore, when  $\mathbf{s} \subseteq \mathbf{t}_n$ ,  $M(\mathbf{s}, \mathbf{t}_n)$  counts the number of vertices of the form  $u0$  such that the subtree below  $u0$  in  $\mathbf{s}$  is the same as the subtree below  $u0$  in  $\mathbf{t}_n$  and  $u1 \in \mathbf{s}$ . Similarly,  $I(\mathbf{s}, \mathbf{t}_n) = 1$  if and only if for all vertices of the form  $u0$ , the subtree below  $u0$  in  $\mathbf{s}$  is the same as the subtree below  $u0$  in  $\mathbf{t}_n$  whenever  $u1 \in \mathbf{s}$ . Hence, if  $\mathbf{s} \subseteq \mathbf{t}_n$ , then

$$p^{-M(\mathbf{s}, \mathbf{t}_n)} I(\mathbf{s}, \mathbf{t}_n) = p^{-N(\mathbf{s})} I(\mathbf{s}, \mathbf{t}_n),$$

where  $N(\mathbf{s}) := \#\{u \in \{0,1\}^* : u1 \in \mathbf{s}\}$ . Thus,

$$K(\mathbf{s}, \mathbf{t}_n) = \begin{cases} (1-p)^{-L(\mathbf{s})} p^{-N(\mathbf{s})} I(\mathbf{s}, \mathbf{t}_n), & \text{if } \mathbf{s} \subseteq \mathbf{t}_n, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that  $\#\mathbf{t}_n(0) \rightarrow \infty$ . For any  $\mathbf{s}$  such that  $1 \in \mathbf{s}$ ,  $I(\mathbf{s}, \mathbf{t}_n)$  must be 0 for all  $n$  sufficiently large, because the subtree below 0 in  $\mathbf{s}$  cannot equal the subtree below 0 in  $\mathbf{t}_n$  for all  $n$ .

On the other hand, if  $1 \notin \mathbf{s}$ , then  $K(\mathbf{s}, \mathbf{t}_n) = K(\mathbf{s}, \tilde{\mathbf{t}}_n)$ , where  $\tilde{\mathbf{t}}_n$  is the tree obtained from  $\mathbf{t}_n$  by deleting all vertices  $v$  with  $1 \leq v$ . Consequently, if  $\#\mathbf{t}_n(0) \rightarrow \infty$ , then in order to check whether  $K(\mathbf{s}, \mathbf{t}_n)$  converges for all  $\mathbf{s} \in \mathbf{S}$ , it suffices to replace  $\mathbf{t}_n$  by  $\tilde{\mathbf{t}}_n$  and restrict consideration to  $\mathbf{s}$  such that  $1 \notin \mathbf{s}$ . Moreover, the limits of  $K(\mathbf{s}, \mathbf{t}_n)$  and  $K(\mathbf{s}, \tilde{\mathbf{t}}_n)$  are the same, so the sequences  $(\mathbf{t}_n)_{n \in \mathbb{N}}$  and  $(\tilde{\mathbf{t}}_n)_{n \in \mathbb{N}}$  correspond to the same point in the Doob-Martin compactification.

Now suppose that  $\#\mathbf{t}_n(0) \not\rightarrow \infty$  (so that  $\#\mathbf{t}_n(1) \rightarrow \infty$  must hold). It is clear that if  $K(\mathbf{s}, \mathbf{t}_n)$  converges for all  $\mathbf{s} \in \mathbf{S}$  with  $1 \notin \mathbf{s}$ , then the sets  $\{v \in \mathbf{t}_n : 0 \leq v\}$  are equal for all  $n$  sufficiently large.

Let  $\hat{\mathbf{t}}_m$  be the subtree of  $\mathbf{t}_m$  obtained by deleting from  $\mathbf{t}_m$  any vertex  $v$  such that  $u1 \leq v$  for some  $u$  with  $\#\mathbf{t}_n(u0) \rightarrow \infty$ . Applying the above arguments recursively, a necessary and sufficient condition for the sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}_0}$  to converge to a point in the Doob-Martin compactification is that whenever  $\#\hat{\mathbf{t}}_n(u0) \not\rightarrow \infty$  for some  $u$ , then the set  $\{v \in \hat{\mathbf{t}}_n : u0 \leq v\}$  are equal for all  $n$  sufficiently large. Moreover, the sequences  $(\mathbf{t}_n)_{n \in \mathbb{N}_0}$  and  $(\hat{\mathbf{t}}_n)_{n \in \mathbb{N}_0}$  converge to the same limit point.

Suppose that  $(\mathbf{t}_n)_{n \in \mathbb{N}_0}$  and hence  $(\hat{\mathbf{t}}_n)_{n \in \mathbb{N}_0}$  converges in the Doob-Martin compactification. Set

$$\mathbf{t}_\infty = \bigcup_{m \in \mathbb{N}_0} \bigcap_{n \geq m} \hat{\mathbf{t}}_n.$$

Note that  $\mathbf{t}_\infty$  is an infinite subtree of  $\{0,1\}^*$  containing the root  $\emptyset$  and if  $\#\mathbf{t}_\infty(u0) = \infty$  for some  $u \in \{0,1\}^*$ , then  $\#\mathbf{t}_\infty(u1) = 0$  (that is,  $u1 \notin \mathbf{t}_\infty$ ). Equivalently, there is a unique

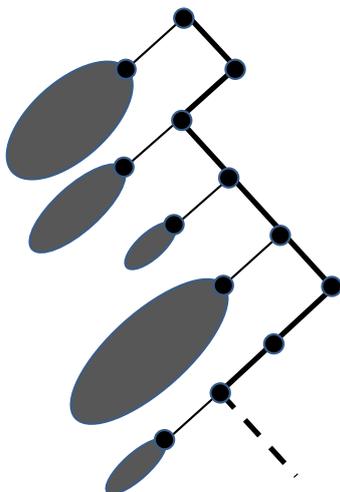


Figure 8: A typical element of the set  $\mathbf{T}$  of infinite rooted binary trees with a single infinite spine. The beginning of the infinite spine is the thick line. The “blobs” hanging off the left side of the spine represent finite subtrees. Any vertex that has a “left” child with infinitely many descendants has no “right” child.

infinite path  $\emptyset = u_0 \rightarrow u_1 \rightarrow u_2 \rightarrow \dots$  in  $\mathbf{t}_\infty$  and this path is such that if  $u_n = w_1 \cdots w_{n-1}0$ , then  $w_1 \cdots w_{n-1}1 \notin \mathbf{t}_\infty$ . Let  $\mathbf{T}$  be the set of subtrees with this property. We can think of a subtree  $\mathbf{t} \in \mathbf{T}$  as consisting of the infinite “spine”  $\emptyset = v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow \dots$  to which are attached the finite subtrees  $\{v \in \mathbf{t} : v_n 0 \leq v\}$  for those  $n \in \mathbb{N}_0$  such that  $v_{n+1} = v_n 1$  – see Figure 8.

We have

$$\lim_{n \rightarrow \infty} K(\mathbf{s}, \mathbf{t}_n) = \lim_{n \rightarrow \infty} K(\mathbf{s}, \hat{\mathbf{t}}_n) = \begin{cases} (1-p)^{-L(\mathbf{s})} p^{-N(\mathbf{s})} I(\mathbf{s}, \mathbf{t}_\infty), & \text{if } \mathbf{s} \subset \mathbf{t}_\infty, \\ 0, & \text{otherwise,} \end{cases}$$

where  $I(\mathbf{s}, \mathbf{t}_\infty)$  is defined to be 1 or 0 depending on whether or not for all vertices of the form  $u0$  with  $u1 \in \mathbf{s}$  the subtree below  $u0$  in  $\mathbf{s}$  is the same as the subtree below  $u0$  in  $\mathbf{t}_\infty$ .

Recall that we write  $|u|$  for the length of a word  $u \in \{0, 1\}^*$ ; that is,  $|u| = k$  when  $u = u_1 \dots u_k$ . Note that if  $\mathbf{t} \in \mathbf{T}$ , then the sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}_0}$  in  $\mathbf{S}$  defined by

$$\mathbf{t}_n := \{u \in \mathbf{t} : |u| \leq n\}$$

converges in the Doob-Martin compactification of  $\mathbf{S}$  and the tree  $\mathbf{t}_\infty$  constructed from this sequence is just  $\mathbf{t}$ .

Finally, observe that if we extend  $K(\mathbf{s}, \mathbf{t})$  for  $\mathbf{s} \in \mathbf{S}$  and  $\mathbf{t} \in \mathbf{T}$  by

$$K(\mathbf{s}, \mathbf{t}) := \begin{cases} (1-p)^{-L(\mathbf{s})} p^{-N(\mathbf{s})} I(\mathbf{s}, \mathbf{t}), & \text{if } \mathbf{s} \subset \mathbf{t}, \\ 0, & \text{otherwise,} \end{cases}$$

then for any distinct  $\mathbf{t}', \mathbf{t}'' \in \mathbf{T}$  there exists  $\mathbf{s} \in \mathbf{S}$  such that  $K(\mathbf{s}, \mathbf{t}') \neq K(\mathbf{s}, \mathbf{t}'')$ .

The important elements of the above discussion are contained in the following result.

**Theorem 7.3.** *Consider the Mallows tree chain with state space  $\mathbf{S}$  consisting of the set of finite rooted binary trees. Let  $\mathbf{T}$  be the set of infinite rooted binary trees  $\mathbf{t}$  such that*

$u1 \in \mathbf{t}$  for some  $u \in \{0, 1\}^*$  implies  $\#\mathbf{t}(u0) < \infty$ . Equip  $\mathbf{S} \sqcup \mathbf{T}$  with the topology generated by the maps  $\Pi_n : \mathbf{S} \sqcup \mathbf{T} \rightarrow \mathbf{S}$ ,  $n \in \mathbb{N}_0$ , defined by  $\Pi_n(\mathbf{t}) := \{u \in \mathbf{t} : |u| \leq n\}$ , where on the right we equip the countable set  $\mathbf{S}$  with the discrete topology. The Doob-Martin compactification  $\bar{\mathbf{S}}$  is homeomorphic to  $\mathbf{S} \sqcup \mathbf{T}$ , and this homeomorphism identifies the Doob-Martin boundary  $\partial\mathbf{S}$  with  $\mathbf{T}$ .

**Remark 7.4.** The limit in the Doob-Martin topology of the Mallows tree chain  $(X_n)_{n \in \mathbb{N}_0}$  started from the trivial tree  $\emptyset$  is just the  $\mathbf{T}$ -valued random variable  $X_\infty := \bigcup_{n \in \mathbb{N}_0} X_n$  introduced in Proposition 7.2. Almost surely, the spine of  $X_\infty$  (that is, the unique infinite path from the root  $\emptyset$ ) is equal to the rightmost path  $\emptyset \rightarrow 1 \rightarrow 11 \rightarrow 111 \dots$  in the complete infinite binary tree.

**Remark 7.5.** It is straightforward to check that each of the harmonic functions  $K(\cdot, \mathbf{t})$ ,  $\mathbf{t} \in \mathbf{T}$  is extremal. If we order the alphabet  $\{0, 1\}$  so that 0 comes before 1 and equip the set of words  $\{0, 1\}^*$  with the corresponding lexicographic order, then the state space of the  $h$ -transformed process corresponding to an infinite tree  $\mathbf{t} \in \mathbf{T}$  is the set of finite subtrees  $\mathbf{s}$  of  $\mathbf{t}$  such that if  $u \in \mathbf{s}$ , then every predecessor of  $u$  in the lexicographic order also belongs to  $\mathbf{s}$ . A realization of the  $h$ -transformed process started from  $\emptyset$  is the deterministic path that adds the vertices of  $\mathbf{t}$  one at a time in increasing lexicographic order.

**Remark 7.6.** As in the BST and DST cases, the Mallows tree process can be regarded as a Markov chain which moves from a tree  $\mathbf{t}$  to a tree  $\mathbf{s}$  of the form  $\mathbf{s} = \mathbf{t} \sqcup \{v\}$ , where the new vertex  $v$  is an external vertex of  $\mathbf{t}$  (see the discussion following (4.2)). This implies that the transition probabilities can be coded by a function  $p$  that maps pairs  $(\mathbf{t}, v)$ ,  $\mathbf{t} \in \mathbf{I}$  and  $v$  an external vertex of  $\mathbf{t}$ , to the probability that the chain moves from  $\mathbf{t}$  to  $\mathbf{t} \sqcup \{v\}$ .

In the BST case one of the  $|\mathbf{t}|+1$  external vertices of  $\mathbf{t}$  is chosen uniformly at random, that is,  $p(v|\mathbf{t}) = 1/(|\mathbf{t}|+1)$ , whereas we have  $p(v|\mathbf{t}) = 2^{-|v|}$  in the DST case. For Mallows trees, we have the following stochastic mechanism. Let  $u$  be the vertex of  $\mathbf{t}$  that is greatest in the lexicographic order. Denote by  $i_1 < \dots < i_\ell$  the indices at which the corresponding entry of  $u$  is a 0 (we set  $\ell = 0$  if every entry of  $u$  is a 1). Write  $v_j$ ,  $1 \leq j \leq \ell$ , for the external vertices of  $\mathbf{t}$  that arise if the 0 in position  $i_j$  is changed to 1. Put  $v_{\ell+1} := v_1$  and  $v_{\ell+2} := v_0$ . Then, we choose  $v_j$  with probability  $p^{i_j}$ ,  $j = 1, \dots, \ell$ , and  $v_{\ell+1}$  and  $v_{\ell+2}$  with probabilities  $rp$  and  $r(1-p)$  respectively, where  $r := 1 - \sum_{j=1}^{\ell} p^{i_j}$ .

Note that not all Markov chains of the vertex-adding type can be represented as trickle-down processes. Indeed, a distinguishing feature of the trickle-down chains within this larger class is the fact that the restriction of the function  $v \rightarrow p(v|\mathbf{t})$  to the external vertices of the left subtree of  $\mathbf{t}$  depends on  $\mathbf{t}$  only via the number of vertices in the right subtree of  $\mathbf{t}$ . Similar restrictions hold with left and right interchanged, and also for the subtrees of non-root vertices.

## 8 $q$ -binomial chains

### 8.1 $q$ -binomial urns

Fix parameters  $0 < q < 1$  and  $0 < r < 1$ , and define a transition matrix  $Q$  for the state space  $\mathbb{N}_0 \times \mathbb{N}_0$  by

$$Q((i, j), (i + 1, j)) = rq^j$$

and

$$Q((i, j), (i, j + 1)) = 1 - rq^j$$

for  $(i, j) \in \mathbb{N}_0 \times \mathbb{N}_0$ . We note that this 2-parameter family of processes is a special case of the 3-parameter family studied in [9], where it is shown to have a number of interesting

connections with graph theory. In the next subsection, we use Markov chains with the transition matrix  $Q$  as the routing chains for a trickle-down process on  $\mathbf{I} = \{0, 1\}^*$  in the same way that we have used the Pólya and Mallows urn processes.

Note that, by a simple Borel-Cantelli argument, almost surely any sample path of a Markov chain  $(Y_n)_{n \in \mathbb{N}_0} = ((Y'_n, Y''_n))_{n \in \mathbb{N}_0}$  with transition matrix  $Q$  is such that

$$Y'_N = Y'_{N+1} = Y'_{N+2} = \dots$$

for some  $N$  (so that  $Y''_{N+1} = Y''_N + 1$ ,  $Y''_{N+2} = Y''_N + 2$ , ...).

We want to compute the probability that the chain goes from  $(i, j)$  to  $(k, \ell)$  for  $i \leq k$  and  $j \leq \ell$ . Observe that the probability the chain goes from  $(i, j)$  to  $(k, \ell)$  via  $(k, j)$  is

$$R((i, j), (k, \ell)) := (rq^j)^{k-i}(1 - rq^j)(1 - rq^{j+1}) \dots (1 - rq^{\ell-1}).$$

Observe also that if  $S(i, j)$  is the probability the chain goes from  $(i, j)$  to  $(i + 1, j + 1)$  via  $(i + 1, j)$  and  $T(i, j)$  is the probability the chain goes from  $(i, j)$  to  $(i + 1, j + 1)$  via  $(i, j + 1)$ , then  $T(i, j) = qS(i, j)$ . It follows by repeated applications of this observation that the probability the chain goes from  $(i, j)$  to  $(k, \ell)$  along some “north-east” lattice path  $\sigma$  is

$$q^{A(\sigma)} R((i, j), (k, \ell)),$$

where  $A(\sigma)$  is the area in the plane above the line segment  $[i, k] \times \{j\}$  and below the curve obtained by a piecewise linear interpolation of  $\sigma$ . Hence, the probability that the chain hits  $(k, \ell)$  starting from  $(i, j)$  is

$$\sum_{\sigma} q^{A(\sigma)} R((i, j), (k, \ell)),$$

where the sum is over all “north-east” lattice paths  $\sigma$  from  $(i, j)$  to  $(k, \ell)$ .

As explained in [1, Chapter 10], the evaluation of the sum is a consequence of the non-commutative  $q$ -binomial theorem of [48] (see also [42]), and

$$\sum_{\sigma} q^{A(\sigma)} = \frac{(1 - q)(1 - q^2) \dots (1 - q^{(k-i)+(l-j)})}{(1 - q)(1 - q^2) \dots (1 - q^{(k-i)}) \times (1 - q)(1 - q^2) \dots (1 - q^{(l-j)})}.$$

Taking, as usual,  $(0, 0)$  as the reference state, the Martin kernel for the chain is thus

$$K((i, j), (k, \ell)) = \frac{(1 - q^{k-i+1}) \dots (1 - q^k) \times (1 - q^{\ell-j+1}) \dots (1 - q^{\ell})}{(1 - q^{(k-i)+(l-j)+1}) \dots (1 - q^{k+l})} \\ \times r^{-i} q^{j(k-i)} \frac{1}{(1 - r)(1 - rq) \dots (1 - rq^{j-1})},$$

for  $i \leq k$  and  $j \leq \ell$  (and 0 otherwise).

The Doob-Martin compactification of a chain with transition matrix  $Q$  is identified in [22, Section 4], but for the sake of completeness we present the straightforward computations. If  $((k_n, \ell_n))_{n \in \mathbb{N}_0}$  is a sequence such that  $k_n + \ell_n \rightarrow \infty$ , then, in order for  $K((i, j), (k_n, \ell_n))$  to converge, we must have either that  $k_n = k_{\infty}$  for some  $k_{\infty}$  for all  $n$  sufficiently large and  $\ell_n \rightarrow \infty$ , in which case the limit is

$$(1 - q^{k_{\infty}-i+1}) \dots (1 - q^{k_{\infty}}) \times r^{-i} q^{j(k_{\infty}-i)} \frac{1}{(1 - r)(1 - rq) \dots (1 - rq^{j-1})}$$

for  $i \leq k_{\infty}$  (and 0 otherwise), or that  $k_n \rightarrow \infty$  with no restriction on  $\ell_n$ , in which case the limit is

$$\begin{cases} r^{-i}, & \text{if } j = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, the Doob-Martin compactification  $\overline{\mathbb{N}_0 \times \mathbb{N}_0}$  of the state space is such that  $\partial(\mathbb{N}_0 \times \mathbb{N}_0) := \overline{\mathbb{N}_0 \times \mathbb{N}_0} \setminus \mathbb{N}_0 \times \mathbb{N}_0$  can be identified with  $\mathbb{N}_0 \cup \{\infty\}$ , the usual one-point compactification of  $\mathbb{N}_0$ .

With this identification, the  $h$ -transformed process corresponding to the boundary point  $k \in \mathbb{N}_0$  has state space  $\{0, \dots, k\} \times \mathbb{N}_0$ , and transition probabilities

$$Q^h((i, j), (i + 1, j)) = (1 - q^{k-i}), \quad i < k,$$

$$Q^h((i, j), (i, j + 1)) = q^{k-i}, \quad i < k,$$

and

$$Q^h((k, j), (k, j + 1)) = 1.$$

Similarly, the  $h$ -transformed process corresponding to the boundary point  $\infty$  has state space  $\mathbb{N}_0 \times \{0\}$  and transition probabilities

$$Q^h((i, 0), (i + 1, 0)) = 1.$$

## 8.2 $q$ -binomial trees

Suppose that we apply the trickle-down construction with  $\mathbf{I} = \{0, 1\}^*$  and all of the routing chains given by the  $q$ -binomial urn of Subsection 8.1, in the same manner that the BST process and the Mallows tree process were built from the Pólya urn and the Mallows urn, respectively. Just as for the latter two processes, we may identify the state space  $\mathbf{S}$  with the set of finite subtrees of  $\{0, 1\}^*$  that contain the root  $\emptyset$ . We call the resulting tree-valued Markov chain the  $q$ -binomial tree process.

Recalling Theorem 7.3 and comparing the conclusions of Subsection 8.1 with those of Subsection 7.2, the following result should come as no surprise. We leave the details to the reader.

**Theorem 8.1.** *Consider the  $q$ -binomial tree chain with state space  $\mathbf{S}$  consisting of the set of finite rooted binary trees. Let  $\mathbf{T}$  be the set of infinite rooted binary trees  $\mathbf{t}$  such that  $u1 \in \mathbf{t}$  for some  $u \in \{0, 1\}^*$  implies  $\#\mathbf{t}(u0) < \infty$ . Equip  $\mathbf{S} \sqcup \mathbf{T}$  with the topology generated by the maps  $\Pi_n : \mathbf{S} \sqcup \mathbf{T} \rightarrow \mathbf{S}$ ,  $n \in \mathbb{N}_0$ , defined by  $\Pi_n(\mathbf{t}) := \{u \in \mathbf{t} : |u| \leq n\}$ , where on the right we equip the countable set  $\mathbf{S}$  with the discrete topology. The Doob-Martin compactification  $\bar{\mathbf{S}}$  is homeomorphic to  $\mathbf{S} \sqcup \mathbf{T}$ , and this homeomorphism identifies the Doob-Martin boundary  $\partial\mathbf{S}$  with  $\mathbf{T}$ . Moreover, each boundary point is extremal.*

## 9 Chains with perfect memory

Recall the Mallows urn model of Subsection 7.2 and the  $q$ -binomial urn model of Subsection 8.1. These Markov chains have the interesting feature that if we know the state of the chain at some time, then we know the whole path of the process up to that time. In this section we examine the Doob-Martin compactifications of such chains with a view towards re-deriving the results of Subsection 7.2 and Subsection 8.1 in a general context. We also analyze a trickle-down process resulting from a composition-valued Markov chain.

We return to the notation of Section 3:  $X = (X_n)_{n \in \mathbb{N}_0}$  is a transient Markov chain with countable state space  $E$ , transition matrix  $P$  and reference state  $e \in E$  such that

$$\rho(j) := \mathbb{P}^e\{X \text{ hits } j\} > 0, \quad \text{for all } j \in E.$$

We suppose that the chain  $X$  has *perfect memory*, by which we mean that the sets

$$E_n := \{j \in E : \mathbb{P}^e\{X_n = j\} > 0\}, \quad n \in \mathbb{N}_0,$$

are disjoint, and that there is a map  $f : E \setminus \{e\} \rightarrow E$  with the property that

$$\mathbb{P}^e\{f(X_n) = X_{n-1}\} = 1, \quad \text{for all } n \in \mathbb{N}.$$

Note that this implies that the tail  $\sigma$ -field associated with the process  $X$  is the same as the  $\sigma$ -field  $\sigma(\{X_n : n \in \mathbb{N}_0\})$  generated by the full collection of variables of the process.

Suppose that we construct a directed graph  $T$  that has  $E$  as its set of vertices and contains a directed edge  $(i, j)$  if and only if  $P(i, j) > 0$ . By the assumption on  $e$ , for any  $j \in E_n$ ,  $n \in \mathbb{N}$ , there is a directed path  $e = i_0 \rightarrow \dots \rightarrow i_n = j$ . Also, it follows from the perfect memory assumption that a directed edge  $(i, j)$  must have  $i \in E_n$  and  $j \in E_{n+1}$  for some  $n$ . Moreover, if  $(i, j)$  is such a directed edge, then there is no  $h \in E_n$  for which  $(h, j)$  is also a directed edge. Combining these observations, we see that the directed graph  $T$  is a rooted tree with root  $e$ . The function  $f$  is simply the map that assigns to any vertex  $j \in E \setminus \{e\}$  its parent. For  $j \in E_n$ ,  $n \in \mathbb{N}$ , the unique directed path from  $e$  to  $j$  is  $e = f^n(j) \rightarrow f^{n-1}(j) \rightarrow \dots \rightarrow f(j) \rightarrow j$ .

Suppose from now on that the tree  $T$  is *locally finite*; that is, for each  $i \in E$ , there are only finitely many  $j \in E$  with  $P(i, j) > 0$ .

As usual, we define a partial order  $\leq$  on  $T (= E)$  by declaring that  $i \leq j$  if  $i$  appears on the unique directed path from the root  $e$  to  $j$ .

We now recall the definition of the *end compactification* of  $T$ . This object can be defined in a manner reminiscent of the definition of the Doob-Martin compactification as follows. We map  $T$  injectively into the space  $\mathbb{R}^T$  of real-valued functions on  $T$  via the map that takes  $j \in T$  to the indicator function of the set  $\{i \in T : i \leq j\}$ . The closure of the image of  $T$  is a compact subset of  $\mathbb{R}^T$ . We identify  $T$  with its image and write  $\bar{T}$  for the closure. The compact space  $\bar{T}$  is metrizable and a sequence  $(j_n)_{n \in \mathbb{N}}$  from  $T$  converges in  $\bar{T}$  if and only if  $\mathbf{1}_{\{i \leq j_n\}}$  converges for all  $i \in T$ , where  $\mathbf{1}_{\{i \leq \cdot\}}$  is the indicator function of the set  $\{j \in T : i \leq j\}$ . The boundary  $\partial T := \bar{T} \setminus T$  can be identified with the infinite directed paths from the root  $e$ . We can extend the function  $\mathbf{1}_{\{i \leq \cdot\}}$  continuously to  $\bar{T}$ . We can also extend the partial order  $\leq$  to  $\bar{T}$  by declaring that  $\xi \not\leq \zeta$  for any  $\xi \neq \zeta \in \partial T$  and  $i \leq \xi$  for  $\xi \in \partial T$  if and only if  $\mathbf{1}_{\{i \leq \xi\}} = 1$ .

**Theorem 9.1.** *Let  $X$  be a chain with state space  $E$ , reference state  $e$ , perfect memory, and locally finite associated tree  $T$ . Then, the associated Martin kernel is given by*

$$K(i, j) = \begin{cases} \rho(i)^{-1}, & \text{if } i \leq j, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i, j \in E.$$

The Doob-Martin compactification of  $E$  is homeomorphic to the end compactification of  $T$ . The extended Martin kernel is given by

$$K(i, \zeta) = \begin{cases} \rho(i)^{-1}, & \text{if } i \leq \zeta, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i \in E, \zeta \in \partial E \cong \partial T.$$

*Proof.* By definition,

$$K(i, j) = \frac{\mathbb{P}^i\{X \text{ hits } j\}}{\mathbb{P}^e\{X \text{ hits } j\}}.$$

By assumption, the numerator is 0 unless  $i \leq j$ . If  $i \leq j$ , then the denominator is

$$\mathbb{P}^e\{X \text{ hits } j\} = \mathbb{P}^e\{X \text{ hits } i\} \mathbb{P}^i\{X \text{ hits } j\}$$

and the claimed formula for the Doob-Martin kernel follows.

The remainder of the proof is immediate from the observation that the manner in which the end compactification is constructed from the functions  $\mathbf{1}_{\{i \leq \cdot\}}$ ,  $i \in E$ , is identical to the manner in which the Doob-Martin compactification is constructed from the functions  $K(i, \cdot) = \rho(i)^{-1} \mathbf{1}_{\{i \leq \cdot\}}$ ,  $i \in E$ .  $\square$

**Example 9.2.** The Mallows urns process satisfies the conditions of Theorem 9.1. The tree  $T$  has  $\mathbb{N}_0^2$  as its set of vertices, and directed edges of the form  $((i, 0), (i + 1, 0))$  and  $((i, j), (i, j + 1))$ ,  $i, j \in \mathbb{N}_0$ . The perfect memory property survives the lift from urn to tree. The “parenthood” function  $f$  takes a tree  $\mathbf{t}$  in the state space of the Mallows tree process and simply removes the vertex of  $\mathbf{t}$  that is greatest in the lexicographic order.

This description of the state space of the Mallows tree process as a “tree-of-trees” also makes its Doob-Martin compactification easier to understand. We know from Section 7.3 that points in the Doob-Martin boundary can be identified with rooted binary trees with a single infinite path – the “spine” – with nothing dangling off to the right of the spine. It is, of course, easy to construct a sequence of finite rooted binary trees that tries to grow more than one infinite path: for example, let  $\mathbf{t}_n$  be the tree that consists of the two vertices  $00 \dots 0, 11 \dots 1 \in \{0, 1\}^n$  and the vertices in  $\{0, 1\}^*$  on the directed paths connecting them to the root  $\emptyset$ . The sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}}$  must have a subsequence with a limit point in the compact space  $\bar{S}$  or, equivalently, it must have a subsequence that converges to a limit in the end compactification  $\bar{T}$  of the tree  $T$ . From the above description of the parenthood function  $f$ , we see for a tree  $\mathbf{s} \in T$  that  $\mathbf{s} \leq \mathbf{t}_n$  if and only if one of the following three conditions hold:

- $\mathbf{s}$  consists of the two vertices  $00 \dots 0 \in \{0, 1\}^n$  and  $11 \dots 1 \in \{0, 1\}^m$  for some  $m \leq n$  and their prefixes in  $\{0, 1\}^*$ ;
- $\mathbf{s}$  consists of the vertex  $00 \dots 0 \in \{0, 1\}^m$  for some  $m \leq n$  and its prefixes in  $\{0, 1\}^*$ ;
- $\mathbf{s}$  consists of the single vertex  $\emptyset \in \{0, 1\}^*$ .

It follows that  $\mathbf{s} \leq \mathbf{t}_n$  for all  $n$  sufficiently large if and only if  $\mathbf{s}$  is the tree consisting of some element of  $\{0\}^*$  and its prefixes in  $\{0, 1\}^*$ . Thus,  $\mathbf{t}_n$  converges in the end compactification of  $T$  to  $\mathbf{t}_\infty \in \bar{T} \setminus T$  as  $n \rightarrow \infty$ , where we can regard  $\mathbf{t}_\infty$  as the single infinite path tree consisting solely of the infinite spine  $\emptyset \rightarrow 0 \rightarrow 00 \rightarrow \dots$ .

We note that the sequence  $(\mathbf{t}_n)_{n \in \mathbb{N}}$  of finite rooted binary trees converges even in the Doob-Martin compactification of the binary search tree process to a point in the boundary. Indeed (see the first paragraph of Section 5), we can identify this latter point with the probability measure on  $\{0, 1\}^\infty$  that puts mass  $\frac{1}{2}$  at each of the points  $00 \dots$  and  $11 \dots$ .

**Example 9.3.** A composition of an integer  $n \in \mathbb{N}$  is an element  $c = (c_1, \dots, c_k)$  of  $\mathbb{N}^*$  with the property that  $\sum_{i=1}^k c_i = n$ . We recall the standard proof of the fact that there are  $2^{n-1}$  such compositions for a given  $n$ : one thinks of placing  $n$  balls on a string and defines a composition by placing separators into some of the  $n - 1$  gaps between the balls. A combinatorially equivalent bijection arises from deleting the last of these balls, labeling the balls to the left of each separator by 1 and labeling the remaining balls by 0. We can now construct a Markov chain  $(X_n)_{n \in \mathbb{N}}$  such that  $X_n$  is uniformly distributed on the set of compositions of  $n$  and  $X_n$  is a prefix of  $X_{n+1}$  for all  $n \in \mathbb{N}$ : the state space is  $E = \{0, 1\}^*$  and the allowed transitions are of the form

$$(u_1, \dots, u_{n-1}) \rightarrow (u_1, \dots, u_{n-1}, 1), \quad (u_1, \dots, u_{n-1}) \rightarrow (u_1, \dots, u_{n-1}, 0),$$

both with probability  $1/2$ . Here  $X_1 = \emptyset$  represents the only composition  $1 = 1$  of  $n = 1$ . Attaching the digit 1 to the state representing a composition of  $n$  means that the new composition, now of  $n + 1$ , has an additional summand of size 1 at the end, whereas adding 0 corresponds to increasing the last summand of the old composition by 1. A construction of this type, which relates random compositions to samples from a geometric distribution, has been used in [28] – see also the references given there.

The chain  $(X_n)_{n \in \mathbb{N}}$  certainly has the perfect memory property and the associated tree  $T$  is just the complete rooted binary tree structure on  $\{0, 1\}^*$  from the Introduction.

It follows from Theorem 9.1 that the Doob-Martin compactification is homeomorphic to  $\{0, 1\}^* \sqcup \{0, 1\}^\infty$ , the end compactification of  $\{0, 1\}^*$ .

Note that we can also think of the chain  $(X_n)_{n \in \mathbb{N}}$  as a result of the trickle-down construction in which the underlying directed acyclic graph  $\mathbf{I}$  is the complete rooted binary tree, the routing instruction chains all have state space  $\{(0, 0)\} \sqcup (\mathbb{N} \times \{0\}) \sqcup (\{0\} \times \mathbb{N})$ , and transition matrices are all of the form

$$\begin{aligned} Q((0, 0), (1, 0)) &= \frac{1}{2}, \\ Q((0, 0), (0, 1)) &= \frac{1}{2}, \\ Q((i, 0), (i + 1, 0)) &= 1, \quad i \geq 1, \\ Q((0, j), (0, j + 1)) &= 1, \quad j \geq 1. \end{aligned}$$

The chain is of the single trail type described in Example 2.4. For processes of this type there are usually several possibilities for the underlying directed graph; here we may take  $\mathbf{I} = \mathbb{N}_0 \times \mathbb{N}_0$  instead of the complete rooted binary tree if we interpret appending 0 as a move to the right and appending 1 as a move up.

**Remark 9.4.** For several of the chains  $(X_n)_{n \in \mathbb{N}_0}$  that we have considered in the previous sections there is a “background chain”  $(\tilde{X}_n)_{n \in \mathbb{N}_0}$  with the perfect memory property in the sense that there is a function  $\Psi : \tilde{S} \rightarrow S$  with  $X_n = \Psi(\tilde{X}_n)$  for all  $n \in \mathbb{N}$ , where  $S$  and  $\tilde{S}$  are the respective state spaces. For example, random recursive trees are often considered together with their labels and are then of the perfect memory type – see Figure 7.

Conversely, we can always extend the state space  $S$  of a given chain by including the previous states, taking the new state space  $\tilde{S}$  to be the set of words from the alphabet  $S$ , to obtain a background chain of the perfect memory type. For example, the Pólya urn then leads to a single trail chain in the sense of Example 2.4, with underlying directed graph  $\mathbb{N} \times \mathbb{N}$  and transitions  $Q((i, j), (i + 1, j)) = i/(i + j)$  and  $Q((i, j), (i, j + 1)) = j/(i + j)$ .

## 10 Another approach to tail $\sigma$ -fields

As mentioned in the Introduction, our initial motivation for studying the Doob-Martin compactifications of various trickle-down chains was to understand the chains’ tail  $\sigma$ -fields. Determining the compactification requires a certain amount of knowledge about the hitting probabilities of a chain, and this information may not always be easy to come by. In this section we consider a family of trickle-down chains for which it is possible to describe the tail  $\sigma$ -field directly without recourse to the more extensive information provided by the Doob-Martin compactification. The class of processes to which this approach applies includes the Mallows tree and  $q$ -binomial tree process that we have already analyzed, as well as the Catalan tree process of Section 11 below that we are unable to treat with Doob-Martin compactification methods.

We begin with a lemma that complements a result from [53] on exchanging the order of taking suprema and intersections of  $\sigma$ -fields.

**Lemma 10.1.** Suppose that on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  there is a collection of independent sub- $\sigma$ -fields  $\mathcal{H}_m$ ,  $m \in \mathbb{N}_0$ , and another collection of sub- $\sigma$ -fields  $\mathcal{G}_{m,n}$ ,  $m, n \in \mathbb{N}_0$ , with the properties

$$\begin{aligned} \mathcal{G}_{0,n} \subseteq \mathcal{G}_{1,n} \subseteq \dots, \quad \text{for all } n \in \mathbb{N}_0, \\ \mathcal{G}_{m,0} \supseteq \mathcal{G}_{m,1} \supseteq \dots, \quad \text{for all } m \in \mathbb{N}_0, \\ \mathcal{G}_{0,0} \subseteq \mathcal{H}_0, \end{aligned}$$

and

$$\mathcal{G}_{m+1,n} \subseteq \mathcal{G}_{m,n} \vee \mathcal{H}_{m+1}, \quad \text{for all } m, n \in \mathbb{N}_0.$$

Then, the two sub- $\sigma$ -fields  $\bigvee_{m \in \mathbb{N}_0} \bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{m,n}$  and  $\bigcap_{n \in \mathbb{N}_0} \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n}$  are equal up to null sets.

*Proof.* We first establish that

$$\bigvee_{m \in \mathbb{N}_0} \bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{m,n} \subseteq \bigcap_{n \in \mathbb{N}_0} \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n}.$$

It suffices to check for each  $M \in \mathbb{N}_0$  that

$$\bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{M,n} \subseteq \bigcap_{n \in \mathbb{N}_0} \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n},$$

but this follows from the observation that

$$\mathcal{G}_{M,n} \subseteq \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n}$$

for every  $n \in \mathbb{N}_0$ .

We now verify that

$$\bigvee_{m \in \mathbb{N}_0} \bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{m,n} \supseteq \bigcap_{n \in \mathbb{N}_0} \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n}$$

up to null sets. For this it suffices to show that any bounded random variable  $Z$  that is measurable with respect to  $\bigvee_{m \in \mathbb{N}_0} \bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{m,n}$ , satisfies the equality

$$\mathbb{E} \left[ Z \mid \bigcap_{n \in \mathbb{N}_0} \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n} \right] = Z \quad \text{a.s.}$$

By a monotone class argument, we may further suppose that  $Z$  is measurable with respect to  $\bigvee_{m=0}^M \bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{m,n} = \bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{M,n}$  for some  $M \in \mathbb{N}_0$ . Our assumptions guarantee that for all  $n \in \mathbb{N}_0$  and  $m > M$

$$\mathcal{G}_{m,n} \subseteq \mathcal{G}_{M,n} \vee \mathcal{H}_{M+1} \vee \dots \vee \mathcal{H}_m$$

and

$$\mathcal{G}_{M,n} \subseteq \mathcal{H}_0 \vee \dots \vee \mathcal{H}_M.$$

From these inclusions, the backwards and forwards martingale convergence theorems and the assumed independence of the  $\mathcal{H}_j$ ,  $j = 0, 1, \dots$  we see that

$$\begin{aligned} & \mathbb{E} \left[ Z \mid \bigcap_{n \in \mathbb{N}_0} \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n} \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[ Z \mid \bigvee_{m \in \mathbb{N}_0} \mathcal{G}_{m,n} \right] \\ &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \mathbb{E} [Z \mid \mathcal{G}_{m,n}] \\ &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \mathbb{E} [\mathbb{E} [Z \mid \mathcal{G}_{M,n} \vee \mathcal{H}_{M+1} \vee \dots \vee \mathcal{H}_m] \mid \mathcal{G}_{m,n}] \\ &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \mathbb{E} [\mathbb{E} [Z \mid \mathcal{G}_{M,n}] \mid \mathcal{G}_{m,n}] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} [Z \mid \mathcal{G}_{M,n}] \\ &= \mathbb{E} \left[ Z \mid \bigcap_{n \in \mathbb{N}_0} \mathcal{G}_{M,n} \right] = Z \quad \text{a.s.}, \end{aligned}$$

as required. □

By the assumptions of the trickle-down construction,  $((Y_n^u)^v)_{n \in \mathbb{N}_0}$  is nondecreasing  $\mathbb{Q}^{u,\xi}$ -almost surely for every  $u \in \mathbf{I}$ ,  $v \in \beta(u)$  and  $\xi \in \mathbf{S}^u$ . Therefore,  $(Y_\infty^u)^v := \lim_{n \rightarrow \infty} (Y_n^u)^v$  exists  $\mathbb{Q}^{u,\xi}$ -almost surely in the usual one-point compactification  $\mathbb{N}_0 \sqcup \{\infty\}$  of  $\mathbb{N}_0$ .

Recall for the Mallows tree and  $q$ -binomial tree processes that  $\mathbf{I} = \{0, 1\}^*$  and that the routing chains in both cases all had the property  $(Y_\infty^u)^{u^0} < \infty$  and  $(Y_\infty^u)^{u^1} = \infty$ ,  $\mathbb{Q}^{u,\xi}$ -almost surely. We see from the following result that it is straightforward to identify the tail  $\sigma$ -field for a trickle-down process if all of its routing chains exhibit this kind of behavior. Another example is the Catalan tree process defined in Section 11 below – see Proposition 11.1.

**Proposition 10.2.** *Suppose that  $\beta(u)$  is finite for all  $u \in \mathbf{I}$ . Fix  $x \in \mathbf{S}$ . Suppose that  $\#\{v \in \beta(u) : (Y_\infty^u)^v = \infty\} = 1$ ,  $\mathbb{Q}^{u,x^u}$ -a.s. for all  $u \in \mathbf{I}$ . Then, the tail  $\sigma$ -field*

$$\bigcap_{m \in \mathbb{N}_0} \sigma\{X_n : n \geq m\}$$

is generated by  $X_\infty := (X_\infty^u)_{u \in \mathbf{I}}$  up to  $\mathbb{P}^x$ -null sets.

*Proof.* By the standing hypotheses on  $\mathbf{I}$  and the assumption that  $\beta(u)$  is finite for all  $u \in \mathbf{I}$ , we can list  $\mathbf{I}$  as  $(u_p)_{p \in \mathbb{N}_0}$  in such a way that  $u_p \leq u_q$  implies  $p \leq q$  (that is, we can put a total order on  $\mathbf{I}$  that refines the partial order  $\leq$  in such a way that the resulting totally ordered set has the same order type as  $\mathbb{N}_0$ ). For each  $p \in \mathbb{N}_0$ , put  $\mathbf{J}_p := \{u_0, \dots, u_p\}$ . By Remark 2.6, each process  $((X_n^u)_{u \in \mathbf{J}_p})_{n \in \mathbb{N}_0}$  is a Markov chain.

Now,

$$\bigcap_{m \in \mathbb{N}_0} \sigma\{X_n : n \geq m\} = \bigcap_{m \in \mathbb{N}_0} \bigvee_{p \in \mathbb{N}_0} \sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\}.$$

By construction,

$$\sigma\{X_n^u : u \in \mathbf{J}_{p+1}, n \geq m\} \subseteq \sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\} \vee \sigma\{Y_n^{u_{p+1}} : n \in \mathbb{N}_0\}.$$

Thus, by Lemma 10.1,

$$\bigcap_{m \in \mathbb{N}_0} \sigma\{X_n : n \geq m\} = \bigvee_{p \in \mathbb{N}_0} \bigcap_{m \in \mathbb{N}_0} \sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\}$$

up to  $\mathbb{P}^x$ -null sets. To show the claimed assertion, it thus suffices to check that for all  $p \in \mathbb{N}_0$

$$\bigcap_{m \in \mathbb{N}_0} \sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\} = \sigma\{X_\infty^u : u \in \mathbf{J}_p\}.$$

We establish this via induction as follows.

For brevity we suppose that  $x^u = (0, 0, \dots)$  for all  $u \in \mathbf{I}$ . In this way we avoid the straightforward but somewhat tedious notational complications of the general case.

By assumption, there is a  $\mathbb{P}^x$ -a.s. unique random element  $V_0 \in \beta(u_0) = \beta(\hat{0})$  such that  $(X_\infty^{u_0})^{V_0} = \infty$ . With  $\mathbb{P}^x$ -probability one,

$$(X_n^{u_0})^v = \begin{cases} (X_\infty^{u_0})^v, & \text{if } v \neq V_0, \\ n - \sum_{w \neq V_0} (X_\infty^{u_0})^w, & \text{if } v = V_0. \end{cases}$$

for all  $v \in \beta(u_0)$  and  $n$  sufficiently large. Thus,  $\bigcap_{m \in \mathbb{N}_0} \sigma\{X_n^u : u \in \mathbf{J}_0, n \geq m\}$  is generated by  $(X_\infty^u)_{u \in \mathbf{J}_0} = X_\infty^{\hat{0}}$  up to  $\mathbb{P}^x$ -null sets.

Suppose we have shown for some  $p \in \mathbb{N}_0$  that  $\bigcap_{m \in \mathbb{N}_0} \sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\} = \sigma\{X_\infty^u : u \in \mathbf{J}_p\}$  up to  $\mathbb{P}^x$ -null sets.

Now,

$$A_n^{u_{p+1}} = \left( \sum_{u \in \alpha(u_{p+1})} (X_n^u)^{u_{p+1}} - 1 \right)_+.$$

Because  $\alpha(u_{p+1}) \subseteq \mathbf{J}_p$ , it follows from our inductive hypothesis that

$$\bigcap_{m \in \mathbb{N}_0} \sigma\{A_n^{u_{p+1}} : n \geq m\} \subseteq \bigcap_{m \in \mathbb{N}_0} \sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\} = \sigma\{X_\infty^u : u \in \mathbf{J}_p\}$$

up to  $\mathbb{P}^x$ -null sets. In particular, the  $\mathbb{N}_0 \sqcup \{\infty\}$ -valued random variable

$$A_\infty^{u_{p+1}} := \lim_{n \rightarrow \infty} A_n^{u_{p+1}}$$

is  $\sigma\{X_\infty^u : u \in \mathbf{J}_p\}$ -measurable up to  $\mathbb{P}^x$ -null sets.

On the event  $\{A_\infty^{u_{p+1}} = \infty\}$ , there is a unique random element  $V_{p+1} \in \beta(u_{p+1})$  such that  $(X_\infty^{u_{p+1}})^{V_{p+1}} = \infty$  and

$$(X_n^{u_{p+1}})^v = \begin{cases} (X_\infty^u)^v, & \text{if } v \neq V_{p+1}, \\ A_n^{u_{p+1}} - \sum_{w \neq V_{p+1}} (X_\infty^{u_{p+1}})^w, & \text{if } v = V_{p+1}, \end{cases}$$

for each  $v \in \beta(u_{p+1})$  and  $n$  sufficiently large. Note that

$$\{A_\infty^{u_{p+1}} = \infty, v = V_{p+1}\} = \{(X_\infty^{u_{p+1}})^v = \infty\}$$

for each  $v \in \beta(u_{p+1})$ . It follows that

$$\begin{aligned} & \bigcap_{m \in \mathbb{N}_0} [\sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\} \vee \sigma\{X_n^{u_{p+1}} \mathbf{1}\{A_\infty^{u_{p+1}} = \infty\} : n \geq m\}] \\ & \subseteq \sigma\{X_\infty^u : u \in \mathbf{J}_{p+1}\} \end{aligned}$$

up to  $\mathbb{P}^x$ -null sets.

Furthermore, on the event  $\{A_\infty^{u_{p+1}} < \infty\}$ ,  $X_n^{u_{p+1}} = X_\infty^{u_{p+1}}$  for all  $n$  sufficiently large, and so

$$\begin{aligned} & \bigcap_{m \in \mathbb{N}_0} [\sigma\{X_n^u : u \in \mathbf{J}_p, n \geq m\} \vee \sigma\{X_n^{u_{p+1}} \mathbf{1}\{A_\infty^{u_{p+1}} < \infty\} : n \geq m\}] \\ & \subseteq \sigma\{X_\infty^u : u \in \mathbf{J}_{p+1}\} \end{aligned}$$

up to  $\mathbb{P}^x$ -null sets. This completes the induction step, and thus the proof of the proposition.  $\square$

**Remark 10.3.** When  $\mathbf{I}$  is a tree and we are in the situation of Proposition 10.2, then  $X_\infty$  may be thought of as an infinite rooted subtree of  $\mathbf{I}$  with a single infinite directed path from the root  $\hat{0}$ . Regarding  $(X_n)_{n \in \mathbb{N}_0}$  as a tree-valued process, we have  $X_\infty = \bigcup_{n \in \mathbb{N}_0} X_n$ . Equivalently,  $X_\infty$  is the limit of the finite subsets  $X_n$  of  $\mathbf{I}$  if we identify the subsets of  $\mathbf{I}$  with the Cartesian product  $\{0, 1\}^{\mathbf{I}}$  in the usual way and equip the latter space with the product topology.

## 11 The Catalan tree process

Let  $\mathbf{S}_n$  denote the set of subtrees of the complete rooted binary tree  $\{0, 1\}^*$  that contain the root  $\emptyset$  and have  $n$  vertices. The set  $\mathbf{S}_n$  has cardinality  $C_n$ , where

$$C_n := \frac{1}{n+1} \binom{2n}{n}$$

is the  $n^{\text{th}}$  Catalan number. A special case of a construction in [37] gives a Markov chain  $(X_n)_{n \in \mathbb{N}_0}$  with state space the set of finite rooted subtrees of  $\{0, 1\}^*$  such that

$$\mathbb{P}^{\{\emptyset\}}\{X_n = \mathbf{t}\} = C_{n+1}^{-1}, \quad \mathbf{t} \in \mathbf{S}_n; \tag{11.1}$$

that is, if the chain begins in the trivial tree  $\{\emptyset\}$ , then its value at time  $n$  is uniformly distributed on  $\mathbf{S}_{n+1}$ . Moreover, the construction in [37] is an instance of the trickle-down construction in which  $\mathbf{I} = \{0, 1\}^*$  and all of the routing chains have the same dynamics.

For the sake of completeness, we reprise some of the development from [37]. Begin with the ansatz that there is indeed a trickle-down process  $(X_n)_{n \in \mathbb{N}_0}$  with  $\mathbf{I} = \{0, 1\}^*$  and identical routing chains such that (11.1) holds. Identify the state spaces of the routing chains with  $\mathbb{N}_0 \times \mathbb{N}_0$  and write  $Q$  for the common transition matrix. We have

$$\begin{aligned} Q^n((0, 0), (k, n - k)) &= \mathbb{P}^{\{\emptyset\}}\{\#X_n(0) = k, \#X_n(1) = n - k\} \\ &= \frac{C_k C_{n-k}}{C_{n+1}}, \quad n \in \mathbb{N}, \quad k = 0, \dots, n. \end{aligned} \tag{11.2}$$

Now,

$$Q((j, i), (j, i + 1)) = Q((i, j), (i + 1, j)) = 1 - Q((i, j), (i, j + 1))$$

by symmetry,

$$\begin{aligned} Q((0, j), (0, j + 1)) &= \frac{\mathbb{P}^{\{\emptyset\}}\{\#X_{j+1}(0) = 0, \#X_{j+1}(1) = j + 1\}}{\mathbb{P}^{\{\emptyset\}}\{\#X_j(0) = 0, \#X_j(1) = j\}} \\ &= \frac{C_0 C_{j+1}}{C_{j+2}} \bigg/ \frac{C_0 C_j}{C_{j+1}} = \frac{(j + 3)(2j + 1)}{(j + 2)(2j + 3)}, \end{aligned}$$

and

$$\begin{aligned} &\mathbb{P}^{\{\emptyset\}}\{\#X_{i+j}(0) = i, \#X_{i+j}(1) = j\} \\ &= \mathbb{P}^{\{\emptyset\}}\{\#X_{i+j-1}(0) = i - 1, \#X_{i+j-1}(1) = j\} Q((i - 1, j), (i, j)) \\ &\quad + \mathbb{P}^{\{\emptyset\}}\{\#X_{i+j-1}(0) = i, \#X_{i+j-1}(1) = j - 1\} Q((i, j - 1), (i, j)) \end{aligned}$$

where the appropriate probabilities on the right side are 0 if  $i = 0$  or  $j = 0$ , so that

$$Q(i, j - 1), (i, j)) = \frac{2j - 1}{j + 1} \left( \frac{i + j + 2}{2i + 2j + 1} - (1 - Q((i - 1, j), (i, j))) \frac{i + 1}{2i - 1} \right).$$

Combining these observations, we can calculate the entries of the transition matrix  $Q$  iteratively and, as observed in [37], the entries of  $Q$  are non-negative and the rows of  $Q$  sum to one. We refer to the resulting Markov chain as the Catalan urn process. Note that if the random tree  $T$  is uniformly distributed on  $\mathbf{S}_{n+1}$  then, conditional on the event  $\{\#T(0) = k, \#T(1) = n - k\}$ , the random trees  $\{u \in \{0, 1\}^* : 0u \in T\}$  and  $\{u \in \{0, 1\}^* : 1u \in T\}$  are independent and uniformly distributed on  $\mathbf{S}_k$  and  $\mathbf{S}_{n-k}$ , respectively. Thus, a trickle-down construction with each routing chain given by the Catalan urn process does indeed give a tree-valued chain satisfying (11.1).

Observe that

$$\lim_{n \rightarrow \infty} \frac{C_{n+1}}{C_n} = 4.$$

It follows from (11.2) that

$$\lim_{\ell \rightarrow \infty} Q^{k+\ell}((0, 0), (k, \ell)) = \lim_{\ell \rightarrow \infty} Q^{k+\ell}((0, 0), (\ell, k)) = 4^{-(k+1)} C_k$$

for all  $k \in \mathbb{N}_0$ . Moreover,  $2 \sum_{k \in \mathbb{N}_0} 4^{-(k+1)} C_k = 1$  from the well-known fact that the generating function of the Catalan numbers is

$$\sum_{k \in \mathbb{N}_0} C_k x^k = \frac{2}{1 + \sqrt{1 - 4x}}, \quad |x| < \frac{1}{4}.$$

Hence, if  $((Y'_n, Y''_n))_{n \in \mathbb{N}_0}$  is a Markov chain with transition matrix  $Q$  and laws  $\mathbb{Q}^{(y', y'')}$ , then

$$(Y'_\infty, Y''_\infty) := \left( \lim_{n \rightarrow \infty} Y'_n, \lim_{n \rightarrow \infty} Y''_n \right) \in (\mathbb{N}_0 \times \{\infty\}) \sqcup (\{\infty\} \times \mathbb{N}_0), \quad \mathbb{Q}^{(0,0)\text{-a.s.}},$$

with

$$\mathbb{Q}^{(0,0)}\{(Y'_\infty, Y''_\infty) = (k, \infty)\} = \mathbb{Q}^{(0,0)}\{(Y'_\infty, Y''_\infty) = (\infty, k)\} = 4^{-(k+1)} C_k, \quad k \in \mathbb{N}_0.$$

The following result is immediate from Proposition 10.2.

**Proposition 11.1.** *The tail  $\sigma$ -field of the Catalan tree process  $(X_n)_{n \in \mathbb{N}_0}$  is generated up to null sets by the infinite random tree  $X_\infty := \bigcup_{n \in \mathbb{N}_0} X_n$  under  $\mathbb{P}^{\{\emptyset\}}$ .*

As we noted in Remark 10.3, the tree  $X_\infty$  has a single infinite path from the root  $\emptyset$ . Denote this path by  $\emptyset = U_0 \rightarrow U_1 \rightarrow \dots$ . For  $n \in \mathbb{N}$ , define  $W_n \in \{0, 1\}$  by  $U_n = W_1 \dots W_n$ . It is apparent from the trickle-down construction and the discussion above that the sequence  $(W_n)_{n \in \mathbb{N}}$  is i.i.d. with  $\mathbb{P}\{W_n = 0\} = \mathbb{P}\{W_n = 1\} = \frac{1}{2}$ . Moreover, if we set  $\bar{W}_n = 1 - W_n$  and put

$$T_n := \{u \in \{0, 1\}^* : W_1 \dots W_{n-1} \bar{W}_n u \in X_\infty\},$$

so that  $T_n$  is either empty or a subtree of  $\{0, 1\}^*$  rooted at  $\emptyset$ , then the sequence  $(T_n)_{n \in \mathbb{N}}$  is i.i.d. and independent of  $(W_n)_{n \in \mathbb{N}}$  with

$$\mathbb{P}\{\#T_n = k\} = 2 \times 4^{-(k+1)} C_k, \quad k \in \mathbb{N}_0,$$

and

$$\mathbb{P}\{T_n = \mathbf{t} \mid \#T_n = k\} = \frac{1}{C_k}, \quad \mathbf{t} \in \mathbf{S}_k, k \in \mathbb{N}.$$

Note that if  $(S_n)_{n \in \mathbb{N}_0}$  is **any** sequence of random subtrees of  $\{0, 1\}^*$  such that  $S_n$  is uniformly distributed on  $\mathbf{S}_{n+1}$  for all  $n \in \mathbb{N}_0$ , then  $S_n$  converges in distribution to a random tree that has the same distribution as  $X_\infty$ , where the notion of convergence in distribution is the one that comes from thinking of subtrees of  $\{0, 1\}^*$  as elements of the Cartesian product  $\{0, 1\}^{\{0,1\}^*}$  equipped with the product topology — see Remark 10.3. The convergence in distribution of such a sequence  $(S_n)_{n \in \mathbb{N}_0}$  and the above description of the limit distribution have already been obtained in [29] using different methods. For a similar weak convergence result for uniform random trees, see [25] and the survey [2, Section 2.5]. Also, if we define rooted finite  $d$ -ary trees for  $d > 2$  as suitable subsets of  $\{0, 1, \dots, d-1\}^*$  in a manner analogous to the way we have defined rooted finite binary trees, then it is shown in [37] that it is possible to construct a Markov chain that grows by one vertex at each step and is uniformly distributed on the set of  $d$ -ary trees with  $n$  vertices at step  $n$  — in particular, there is an almost sure (and hence distributional) limit as  $n \rightarrow \infty$  in the same sense as we just observed for the uniform binary trees. We have not investigated whether this process is the result of a trickle-down construction. Lastly, we note that there are interesting ensembles of trees that can't be embedded into a trickle-down construction or, indeed, into any Markovian construction in which a single vertex is added at each step; for example, it is shown in [30] that this is not possible

for the ensemble obtained by taking a certain critical Galton-Watson tree with offspring distribution supported on  $\{0, 1, 2\}$  and conditioning the total number of vertices to be  $n \in \mathbb{N}$ .

**Acknowledgments.** The authors thank Heinrich von Weizsäcker for his role in bringing them together to work on the problems considered in the paper. They also thank Sasha Gnedin and the referee for a number of helpful comments and pointers to the literature.

## References

- [1] George E. Andrews, Richard Askey, and Ranjan Roy, *Special functions*, Cambridge University Press, Cambridge, 1999. MR-2000g:33001
- [2] David Aldous and J. Michael Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 1–72. MR-2023650
- [3] David M. Blei, Thomas L. Griffiths, and Michael I. Jordan, *The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies*, J. ACM **57** (2010), no. 2, Art. 7, 30. MR-2606082
- [4] David Blackwell and David Kendall, *The Martin boundary of Pólya's urn scheme, and an application to stochastic population growth*, J. Appl. Probability **1** (1964), 284–296. MR-0176518
- [5] David Blackwell and James B. MacQueen, *Ferguson distributions via Pólya urn schemes*, Ann. Statist. **1** (1973), 353–355. MR-0362614
- [6] Brigitte Chauvin, Michael Drmota, and Jean Jabbour-Hattab, *The profile of binary search trees*, Ann. Appl. Probab. **11** (2001), no. 4, 1042–1062. MR-1878289
- [7] Douglas E. Critchlow, Michael A. Fligner, and Joseph S. Verducci, *Probability models on rankings*, J. Math. Psych. **35** (1991), no. 3, 294–318. MR-1128236
- [8] Douglas E. Critchlow, *Metric methods for analyzing partially ranked data*, Lecture Notes in Statistics, vol. 34, Springer-Verlag, Berlin, 1985. MR-818986
- [9] Davide Crippa and Klaus Simon, *q-distributions and Markov processes*, Discrete Math. **170** (1997), no. 1-3, 81–98. MR-1452938
- [10] Luc Devroye, *Universal limit laws for depths in random trees*, SIAM J. Comput. **28** (1999), no. 2, 409–432 (electronic). MR-1634354
- [11] P. Diaconis and W. Fulton, *A growth model, a game, an algebra, Lagrange inversion, and characteristic classes*, Rend. Sem. Mat. Univ. Politec. Torino **49** (1991), no. 1, 95–119 (1993), Commutative algebra and algebraic geometry, II (Italian) (Turin, 1990). MR-1218674
- [12] Florian Dennert and Rudolf Grübel, *On the subtree size profile of binary search trees*, Combin. Probab. Comput. **19** (2010), 561–578.
- [13] Rui Dong, Christina Goldschmidt, and James B. Martin, *Coagulation-fragmentation duality, Poisson-Dirichlet distributions and random recursive trees*, Ann. Appl. Probab. **16** (2006), no. 4, 1733–1750. MR-2288702
- [14] Persi Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11, Institute of Mathematical Statistics, Hayward, CA, 1988. MR-964069
- [15] Michael Drmota, Svante Janson, and Ralph Neininger, *A functional limit theorem for the profile of search trees*, Ann. Appl. Probab. **18** (2008), no. 1, 288–333. MR-2380900
- [16] J. L. Doob, *Discrete potential theory and boundaries*, J. Math. Mech. **8** (1959), 433–458; erratum 993. MR-0107098
- [17] Michael Fuchs, *Subtree sizes in recursive trees and binary search trees: Berry-Esseen bounds and Poisson approximations*, Combin. Probab. Comput. **17** (2008), no. 5, 661–680. MR-2454562
- [18] M. A. Fligner and J. S. Verducci, *Distance based ranking models*, J. Roy. Statist. Soc. Ser. B **48** (1986), no. 3, 359–369. MR-876847

- [19] Frederick M. Goodman and Sergei V. Kerov, *The Martin boundary of the Young-Fibonacci lattice*, J. Algebraic Combin. **11** (2000), no. 1, 17–48. MR-1747061
- [20] Alexander Gnedin and Grigori Olshanski, *The boundary of the Eulerian number triangle*, Mosc. Math. J. **6** (2006), no. 3, 461–475. MR-2274860
- [21] ———, *Coherent permutations with descent statistic and the boundary problem for the graph of zigzag diagrams*, Int. Math. Res. Not. (2006), Art. ID 51968, 39. MR-2211157
- [22] ———, *A  $q$ -analogue of de Finetti's theorem*, Electron. J. Combin. **16** (2009), no. 1, Research Paper 78, 16. MR-2529787
- [23] ———,  *$q$ -exchangeability via quasi-invariance*, Ann. Probab. **38** (2010), no. 6, 2103–2135. MR-2683626
- [24] A. Gnedin and J. Pitman, *Exchangeable Gibbs partitions and Stirling triangles*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **325** (2005), no. Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 12, 83–102, 244–245. MR-2160320
- [25] G. R. Grimmett, *Random labelled trees and their branching networks*, J. Austral. Math. Soc. Ser. A **30** (1980/81), no. 2, 229–237. MR-607933
- [26] Rudolf Grübel, *On the median-of- $k$  version of Hoare's selection algorithm*, Theor. Inform. Appl. **33** (1999), no. 2, 177–192. MR-1707969
- [27] ———, *On the silhouette of binary search trees*, Ann. Appl. Probab. **19** (2009), no. 5, 1781–1802. MR-2569807
- [28] Paweł Hitczenko and Guy Louchard, *Distinctness of compositions of an integer: a probabilistic analysis*, Random Structures Algorithms **19** (2001), no. 3-4, 407–437, Analysis of algorithms (Krynica Morska, 2000). MR-1871561
- [29] Svante Janson, *Ideals in a forest, one-way infinite binary trees and the contraction method*, Mathematics and computer science, II (Versailles, 2002), Trends Math., Birkhäuser, Basel, 2002, pp. 393–414. MR-1940149
- [30] ———, *Conditioned Galton-Watson trees do not grow*, Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities, Discrete Math. Theor. Comput. Sci. Proc., AG, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2006, pp. 331–334. MR-2509643
- [31] Norman L. Johnson and Samuel Kotz, *Urn models and their application*, John Wiley & Sons, New York-London-Sydney, 1977, An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics. MR-0488211
- [32] Donald E. Knuth, *The art of computer programming. Vol. 1: Fundamental algorithms*, Second printing, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1969. MR-0286317
- [33] ———, *The art of computer programming. Volume 3*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973, Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing. MR-0445948
- [34] Sergei Kerov, Andrei Okounkov, and Grigori Olshanski, *The boundary of the Young graph with Jack edge multiplicities*, Internat. Math. Res. Notices (1998), no. 4, 173–199. MR-1609628
- [35] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp, *Denumerable Markov chains*, second ed., Springer-Verlag, New York, 1976, With a chapter on Markov random fields, by David Griffeath, Graduate Texts in Mathematics, No. 40. MR-0407981
- [36] Gregory F. Lawler, Maury Bramson, and David Griffeath, *Internal diffusion limited aggregation*, Ann. Probab. **20** (1992), no. 4, 2117–2140. MR-1188055
- [37] Malwina Luczak and Peter Winkler, *Building uniformly random subtrees*, Random Structures Algorithms **24** (2004), no. 4, 420–443. MR-2060629
- [38] Hosam M. Mahmoud, *Evolution of random search trees*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 1992, A Wiley-Interscience Publication. MR-1140708
- [39] C. L. Mallows, *Non-null ranking models. I*, Biometrika **44** (1957), 114–130. MR-0087267

- [40] John I. Marden, *Analyzing and modeling rank data*, Monographs on Statistics and Applied Probability, vol. 64, Chapman & Hall, London, 1995. MR-1346107
- [41] J. Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. MR-2245368
- [42] G. Pólya, *On the number of certain lattice polygons*, J. Combinatorial Theory **6** (1969), 102–105. MR-38
- [43] Massimo A. Picardello and Wolfgang Woess, *The full Martin boundary of the bi-tree*, Ann. Probab. **22** (1994), no. 4, 2203–2222. MR-1331221
- [44] Jim Pitman and Matthias Winkel, *Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions*, Ann. Probab. **37** (2009), no. 5, 1999–2041. MR-2561439
- [45] D. Revuz, *Markov chains*, North-Holland Publishing Co., Amsterdam, 1975, North-Holland Mathematical Library, Vol. 11. MR-0415773
- [46] L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. MR-1796539
- [47] Stanley A. Sawyer, *Martin boundaries and random walks*, Harmonic functions on trees and buildings (New York, 1995), Contemp. Math., vol. 206, Amer. Math. Soc., Providence, RI, 1997, pp. 17–44. MR-1463727
- [48] Marcel Paul Schützenberger, *Une interprétation de certaines solutions de l'équation fonctionnelle:  $F(x + y) = F(x)F(y)$* , C. R. Acad. Sci. Paris **236** (1953), 352–353. MR-14,768g
- [49] R. Sedgewick and Ph. Flajolet, *An introduction to the analysis of algorithms*, Addison-Wesley, Reading, 1996.
- [50] Robert T. Smythe and Hosam M. Mahmoud, *A survey of recursive trees*, Teor. Īmovĭr. Mat. Stat. (1994), no. 51, 1–29. MR-1445048
- [51] Richard P. Stanley, *Enumerative combinatorics. Vol. 1*, Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, Cambridge, 1997, With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original. MR-1442260
- [52] Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, and David M. Blei, *Hierarchical Dirichlet processes*, J. Amer. Statist. Assoc. **101** (2006), no. 476, 1566–1581. MR-2279480
- [53] Heinrich von Weizsäcker, *Exchanging the order of taking suprema and countable intersections of  $\sigma$ -algebras*, Ann. Inst. H. Poincaré Sect. B (N.S.) **19** (1983), no. 1, 91–100. MR-699981
- [54] Takesi Watanabe, *A probabilistic method in Hausdorff moment problem and Laplace-Stieltjes transform*, J. Math. Soc. Japan **12** (1960), 192–206. MR-0120683
- [55] Wolfgang Woess, *Random walks on infinite graphs and groups*, Cambridge Tracts in Mathematics, vol. 138, Cambridge University Press, Cambridge, 2000. MR-1743100