TREE-VALUED MARKOV CHAINS DERIVED FROM GALTON-WATSON PROCESSES. *

David Aldous and Jim Pitman

Technical Report No. 481

Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

February 1997

Abstract

Let $\mathcal{G}$ be a Galton-Watson tree, and for $0 \le u \le 1$ let $\mathcal{G}_u$ be the subtree of $\mathcal{G}$ obtained by retaining each edge with probability u. We study the tree-valued Markov process $(\mathcal{G}_u, 0 \le u \le 1)$ and an analogous process $(\mathcal{G}^*_u, 0 \le u \le 1)$ in which $\mathcal{G}^*_1$ is a critical or subcritical Galton-Watson tree conditioned to be infinite. Results simplify and are further developed in the special case of Poisson(μ) offspring distribution.

Running head. Tree-valued Markov chains.

Key words. Borel distribution, branching process, conditioning, Galton-Watson process, generalized Poisson distribution, h-transform, pruning, random tree, size-biasing, spinal decomposition, thinning.

AMS Subject classifications 05C80, 60C05, 60J27, 60J80

*Research supported in part by N.S.F. Grants DMS9404345 and 9622859

Contents

1 Introduction

This paper develops some theory for Galton-Watson trees $\mathcal{G}$ (i.e. family trees associated with Galton-Watson branching processes), starting from the following two known facts.
(i) [Lemma 10] For fixed $0 \leq u \leq 1$ let $\mathcal{G}_u$ be the "pruned" tree obtained by cutting edges of $\mathcal{G}$ and discarding the attached branch) independently with probability $1-u$. Then $\mathcal{G}_u$ is another Galton-Watson tree.
(ii) [Proposition 2] For critical or subcritical $\mathcal{G}$ one can define a tree $\mathcal{G}^\infty$,

interpretable as $\mathcal{G}$ *conditioned on non-extinction.* Qualitatively, $\mathcal{G}^\infty$ consists of a single infinite "spine" to which finite subtrees are attached.

We interpret (i) as defining a *pruning process* $(\mathcal{G}_u, 0 \leq u \leq 1)$, which is a tree-valued continuous-time inhomogeneous Markov chain such that $\mathcal{G}_0$ is the trivial tree consisting only of the root vertex, and $\mathcal{G}_1 = \mathcal{G}$. An analogous pruning process $(\mathcal{G}^*_u, 0 \leq u \leq 1)$ with $\mathcal{G}^*_1 = \mathcal{G}^\infty$ is constructed from the conditioned tree $\mathcal{G}^\infty$ of (ii). Section 3 gives a careful description of the transition rates and transition probabilities for these processes. The two processes are qualitatively different, in the following sense. If $\mathcal{G}$ is supercritical then on the event $\mathcal{G}$ is infinite there is a random *ascension time* A such that $\mathcal{G}_{A-}$ is finite but $\mathcal{G}_A$ is infinite: the chain "jumps to infinite size" at time A. In contrast, the process $(\mathcal{G}^*_u)$ "grows to infinity" at time 1, meaning that $\mathcal{G}^*_u$ is finite for $u < 1$ but $\mathcal{G}^*_{1-} = \mathcal{G}^*_1$ is infinite. A connection between the two processes is made (Section 3.4) by conditioning $(\mathcal{G}_u, 0 \leq u < A)$ on the event that A equals the critical time, i.e. the a for which $\mathcal{G}_a$ has mean offspring equal 1. By rescaling the time parameter we may take $a = 1$, and the conditioned process is then identified with $(\mathcal{G}^*_u, 0 \leq u < 1)$.

These results simplify, and further connections appear, in the special case of Poisson offspring distribution, which is the subject of Section 4. There we consider $(\mathcal{G}_\mu, 0 \leq \mu < \infty)$, where $\mathcal{G}_\mu$ is the family tree of the Galton-Watson branching process with Poisson(μ) offspring, and the associated pruned conditioned process $(\mathcal{G}^*_\mu, 0 \leq \mu \leq 1)$. To highlight four properties:

- The distribution of $\mathcal{G}^*_1$ has several different interpretations as a limit (Section 4.3).
- For fixed $\mu < 1$, the distribution of $\mathcal{G}^*_\mu$ is the distribution of $\mathcal{G}_\mu$, size-biased by the total size of $\mathcal{G}_\mu$ (Section 4.4).
- The process $(\mathcal{G}_\mu)$ run until its ascension time $A > 1$ has a representation in terms of $(\mathcal{G}^*_\mu)$ as (Section 4.5)

$$(\mathcal{G}_\lambda, 0 \leq \lambda < A) \stackrel{d}{=} (\mathcal{G}^*_{\lambda U}, 0 \leq \lambda < (-\log U)/(1-U))$$

  where U is uniform $(0,1)$, independent of $(\mathcal{G}^*_\mu, 0 \leq \mu \leq 1)$.
- In constructing $\mathcal{G}^*_\mu$ by pruning $\mathcal{G}^*_1$, a certain vertex becomes distinguished, i.e. the highest vertex of the spine of $\mathcal{G}^*_1$ retained in $\mathcal{G}^*_\mu$. This

vertex turns out to be distributed uniformly on $\mathcal{G}^*_\mu$, and a simple *spinal decomposition* of $\mathcal{G}^*_\mu$ into independent tree components is obtained by cutting the edges of $\mathcal{G}^*_\mu$ along the path from the root to the distinguished vertex (Section 4.6).

Other topics include consequent distributional indentities relating Borel and size-biased Borel distributions (Sections 4.5 and 4.7) and the interpretation of trees conditioned to be infinite as explicit Doob h-transforms, with the related identification of the Martin boundary of $(\mu, \mathcal{G}_\mu)$ (Section 4.4).

None of the individual results is especially hard; the length of the paper is due partly to our development of a precise formalism for writing rigorous proofs of such results. Section 2 contains this formalism and discussion of known results.

1.1 Related topics

Of course, branching processes form a classical part of probability theory. Various "probability on trees" topics of contemporary interest are treated in the forthcoming monograph by Lyons [30], which explores several aspects of Galton-Watson trees but touches only tangentially on the specific topics of this paper.

Our motivation came from the following considerations, which will be elaborated in a more wide-ranging but less detailed companion paper [7]. Suppose that for each N there is a Markov chain taking values in the set of forests on N vertices. Looking at the tree containing a given vertex gives a tree-valued process, and taking $N \to \infty$ limits may give a tree-valued Markov chain. The prototype example (not exactly forest-valued, of course) is the random graph process $(\mathrm{G}(N, P(\text{edge}) = \mu/N), 0 \leq \mu \leq N)$ for which the limit tree-valued Markov chain is our pruned Poisson-Galton-Watson process $(\mathcal{G}_\mu, 0 \leq \mu < \infty)$ [1]. The pruned conditioned Poisson-Galton-Watson process $(\mathcal{G}^*_\mu, 0 \leq \mu \leq 1)$ arises in a more subtle way as a limit of the Marcus-Lushnikov (discrete coalescent) process with additive kernel (see [6] for background on the general Marcus-Lushnikov process and [42, 43, 44, 38] for recent results on the additive case). More exotic variations of $(\mathcal{G}_\mu)$, e.g. a stationary Markov process in which branches grow and are cut down upon becoming infinite, arise as other $N \to \infty$ limits and are studied in [7]. Finally, we remark that the unconditioned and conditioned critical Poisson-Galton-

Watson distributions arise as $N \to \infty$ limits in several other contexts (as "fringes" in random tree models [3], in particular in random spanning trees [2, 37]; in the Wright-Fisher model) where there is no natural pruning structure.

2 Background and technical set-up

Here we set up our general notation for random trees, and present some background material about Galton-Watson trees.

2.1 Notation and terminology for trees

Except where otherwise indicated, by a *tree* $\mathbf{t}$ we mean a *rooted labeled tree*, that is a set $V = \mathrm{verts}(\mathbf{t})$, called the set of *vertices* or *labels* of $\mathbf{t}$, equipped with a *directed edge relation* $\xrightarrow{\mathbf{t}}$ such that for some (obviously unique) element $\mathrm{root}(\mathbf{t}) \in V$ there is for each vertex $v \in V$ a unique *path from the root to* v, that is a finite sequence of vertices $(v_0 = \mathrm{root}(\mathbf{t}), v_1, \ldots, v_h = v)$ such that $v_{i-1} \xrightarrow{\mathbf{t}} v_i$ for each $1 \le i \le h$. Then $h = h(v, \mathbf{t})$ is the *height* of vertex v in the tree $\mathbf{t}$. Formally, $\mathbf{t}$ is identified by its vertex set V and its set of *directed edges*, that is the set $\{(v, w) \in V \times V : v \xrightarrow{\mathbf{t}} w\}$. If a subset S of $\mathrm{verts}(\mathbf{t})$ is such that the restriction of the relation $\xrightarrow{\mathbf{t}}$ to $S \times S$ defines a tree $\mathbf{s}$ with $\mathrm{verts}(\mathbf{s}) = S$, then either S or $\mathbf{s}$ may be called called a *subtree* of $\mathbf{t}$. Let $\#V \in \{0, 1, 2, \ldots, \infty\}$ be the number of elements of a set V, and for a tree $\mathbf{t}$ let $\#\mathbf{t} = \#\mathrm{verts}(\mathbf{t})$. The number of edges of a tree $\mathbf{t}$ is $\#\mathbf{t} - 1$. For a tree $\mathbf{t}$ and $v \in \mathrm{verts}(\mathbf{t})$ let $\mathrm{children}(v, \mathbf{t}) := \{w \in \mathrm{verts}(\mathbf{t}) : v \xrightarrow{\mathbf{t}} w\}$ denote the set of *children* of v in $\mathbf{t}$, and let $c_v \mathbf{t} := \#\mathrm{children}(v, \mathbf{t})$, the number of children of v in $\mathbf{t}$. Each non-root vertex w of $\mathbf{t}$ is a child of some unique vertex v of $\mathbf{t}$, say $v = \mathrm{parent}(w, \mathbf{t})$. Let $\mathbf{s}$ and $\mathbf{t}$ be two trees. Call $\mathbf{s}$ a *relabeling of* $\mathbf{t}$ if there exists a bijection $\ell : \mathrm{verts}(\mathbf{s}) \to \mathrm{verts}(\mathbf{t})$ such that $v \xrightarrow{\mathbf{s}} w$ if and only if $\ell(v) \xrightarrow{\mathbf{t}} \ell(w)$. Then $\mathrm{root}(\mathbf{t}) = \ell(\mathrm{root}(\mathbf{s}))$ and $h(v, \mathbf{s}) = h(\ell(v), \mathbf{t})$.

In the discussion above there was no notion of "birth order" for children. To incorporate this notion, for $n \in \mathbb{N} := \{1, 2, \ldots\}$ let $\mathbf{T}_n$ denote the set of *family trees* (also called *rooted ordered trees* or *planted plane trees* [14, 46]) with n vertices. Figure 1 illustrates the 5 trees comprising $\mathbf{T}_4$.

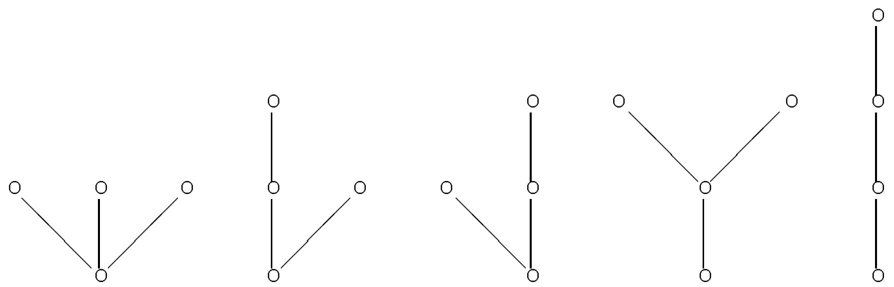

Figure 1

We interpret an element $\mathbf{t}$ of $\mathbf{T} := \cup_n \mathbf{T}_n$ as a finite family tree with the root representing a single *progenitor*, and each vertex of the tree representing an *individual* of the family. Then for $g = 0, 1, 2 \ldots$ each vertex at height g corresponds to an individual in the gth *generation* of the family. While the graphical representation of $\mathbf{t}$ in Figure 1 involves no explicit labeling of the vertices v of $\mathbf{t}$, we identify an individual in the gth generation of $\mathbf{t}$ as a sequence of g integers, for instance $(2, 7, 4)$ to indicate a third generation individual who is the 4th child of the 7th child of the 2nd child of the progenitor. Thus, following Harris[22, §VI.2] and Kesten [25], we identify each $\mathbf{t} \in \mathbf{T}_n$ as a rooted labeled tree with verts$(\mathbf{t}) \subset \mathcal{V}$, where $\mathcal{V} := \left(\cup_{g=1}^{\infty} \mathbb{N}^g\right) \cup \{0\}$ is the set of all finite sequences of positive integers, together with a root element denoted 0. Regard $\mathcal{V}$ as a rooted tree, with $v \stackrel{\mathcal{V}}{\to} w$ if and only if $w = (v, j)$ for some $j \in \mathbb{N}$, where $(v, j) \in \mathbb{N}^{g+1}$ is defined by appending j to $v \in \mathbb{N}^g$ for $g \geq 1$, and where $(0, j) = j \in \mathbb{N}$. If $w = (v, j)$ call w the jth child of v, call j the rank of w, and write $j = \mathrm{rank}(w)$. So for each non-root $w \in \mathcal{V}$, the positive integer $\mathrm{rank}(w) \in \mathbb{N}$ is the last component of the finite sequence w. A (finite or infinite) *family tree* is a subtree $\mathbf{t}$ of $\mathcal{V}$ such that $0 \in$ verts$(\mathbf{t})$, and for each $v \in$ verts$(\mathbf{t})$ the set of children of v is the set $\{(v, j) : 1 \leq j \leq c_v \mathbf{t}\}$. Let $\mathbf{T}^{(\infty)}$ denote the set of all such family trees $\mathbf{t}$. Each $\mathbf{t} \in \mathbf{T}^{(\infty)}$ is a subtree of $\mathcal{V}$, whose edge relation $\stackrel{\mathbf{t}}{\to}$ on verts$(\mathbf{t})$ is defined by restriction of the edge relation $\stackrel{\mathcal{V}}{\to}$ on $\mathcal{V}$. A family tree $\mathbf{t}$ is therefore uniquely identified by its vertex set verts$(\mathbf{t}) \subset \mathcal{V}$, and it is convenient to identify $\mathbf{t}$ with verts$(\mathbf{t})$. Thus $\mathbf{T}^{(\infty)}$ is identified as a collection of subsets of $\mathcal{V}$ subject to certain constraints indicated above, and we may write for instance $v \in \mathbf{t}$ instead of $v \in$ verts$(\mathbf{t})$

to indicate that v is a vertex of $\mathbf{t}$. ¿From the definitions above

$$\mathbf{T}_n := \{\mathbf{t} \in \mathbf{T}^{(\infty)} : \#\mathbf{t} = n\}; \quad \mathbf{T} := \{\mathbf{t} \in \mathbf{T}^{(\infty)} : \#\mathbf{t} < \infty\} = \bigcup_{n=1}^{\infty} \mathbf{T}_n.$$

The *height* of a finite tree is the maximum height of all vertices in the tree. There is a natural *restriction map* $r_h : \mathbf{T}^{(\infty)} \to \mathbf{T}^{(h)}$ where $\mathbf{T}^{(h)}$ is the set of finite family trees of height at most h. For $\mathbf{t}$ identified as a subset of $\mathcal{V}$, $r_h\mathbf{t}$ is the tree formed by all vertices of $\mathbf{t}$ of height at most h. A tree $\mathbf{t} \in \mathbf{T}^{(\infty)}$ is identified by the sequence $(r_h\mathbf{t}, h \geq 0)$. Note that the $r_h\mathbf{t} \in \mathbf{T}^{(h)}$ are subject only to the consistency condition that $r_h\mathbf{t} = r_h(r_{h+1}\mathbf{t})$. The set $\mathbf{T}^{(\infty)}$ is now identified as a subset of an infinite product of countable sets

$$\mathbf{T}^{(\infty)} \subset \mathbf{T}^{(0)} \times \mathbf{T}^{(1)} \times \mathbf{T}^{(2)} \times \cdots.$$

We give $\mathbf{T}^{(\infty)}$ the topology derived by this identification from the product of discrete topologies on the $\mathbf{T}^{(h)}$. So a sequence of trees $\mathbf{t}_n$ has a limit $\lim_n \mathbf{t}_n = \mathbf{t} \in \mathbf{T}^{(\infty)}$ iff for every h there exists $\mathbf{t}^{(h)} \in \mathbf{T}^{(h)}$ and $n(h)$ such that $r_h\mathbf{t}_n = \mathbf{t}^{(h)}$ for all $n \geq n(h)$; the limit is then the unique $\mathbf{t} \in \mathbf{T}^{(\infty)}$ with $r_h\mathbf{t} = \mathbf{t}^{(h)}$. In particular, for each $\mathbf{t} \in \mathbf{T}^{(\infty)}$, the sequence $r_n\mathbf{t}$ has limit $\mathbf{t}$ as $n \to \infty$.

Let $\mathbf{s}$ be a tree whose vertex set V is a subset of some set S equipped with a total ordering. In particular, we have in mind the cases $S = \mathbb{N}$ with the usual ordering, and $S = \mathcal{V}$ with lexicographical ordering and 0 as least element. Suppose each $v \in \mathrm{verts}(\mathbf{s})$ has only a finite number of children. Then there is a natural relabeling ℓ of the vertices of $\mathbf{s}$ by $\mathcal{V}$ which defines a family tree $\mathbf{t}$ associated with $\mathbf{s}$, say $\mathbf{t} = \mathrm{fam}(\mathbf{s})$. The relabeling $\ell : \mathrm{verts}(\mathbf{s}) \to \mathcal{V}$ is defined as follows. Let $\ell(\mathrm{root}(\mathbf{s})) = 0$. For each non-root vertex v of $\mathbf{s}$ let $\mathrm{rank}(v)$ be the rank of v amongst its siblings, that is the number of $w \in \mathrm{verts}(\mathbf{s})$ such that w has the same parent as v and $w \leq v$. For v with height $h \geq 1$ in $\mathbf{s}$ let $(\mathrm{root}(\mathbf{s}), v_1, \ldots, v_h = v)$ be the path from $\mathrm{root}(\mathbf{s})$ to v in $\mathbf{s}$, and define

$$\ell(v) := (\mathrm{rank}(v_1), \ldots, \mathrm{rank}(v_h)) \in \mathbb{N}^h.$$

The tree $\mathbf{t} = \mathrm{fam}(\mathbf{s}) \in \mathbf{T}^{(\infty)}$ is the subtree of $\mathcal{V}$ whose vertex set is $\ell(\mathrm{verts}(\mathbf{s}))$, the range of the relabeling map $\ell : \mathrm{verts}(\mathbf{s}) \to \mathcal{V}$.

For $\mathbf{t} \in \mathbf{T}^{(\infty)}$ and $g \geq 0$, let $\mathrm{gen}(g, \mathbf{t})$ be the *gth generation of individuals in* $\mathbf{t}$, in other words the set of vertices of $\mathbf{t}$ of height g. To illustrate notation, there are the following identities between subsets of the set $\mathcal{V}$: for all $\mathbf{t}$, for all $h = 0, 1, \ldots$:

$$\operatorname{gen}(h+1, \mathbf{t}) = \bigcup_{v \in \operatorname{gen}(h, \mathbf{t})} \operatorname{children}(v, \mathbf{t}); \ r_h \mathbf{t} = \bigcup_{g=0}^{h} \operatorname{gen}(g, \mathbf{t}).$$

Let $Z_h \mathbf{t} := \#\operatorname{gen}(h, \mathbf{t})$, the size of generation h of $\mathbf{t}$. The above identities imply

$$Z_{h+1}\mathbf{t} = \sum_{v \in \operatorname{gen}(h, \mathbf{t})} c_v \mathbf{t}; \ \#r_h \mathbf{t} = \sum_{n=0}^{h} Z_n \mathbf{t}; \ \#\mathbf{t} = \sum_{n=0}^{\infty} Z_n \mathbf{t}.$$

The abbreviation

$$Z\mathbf{t} := Z_1 \mathbf{t} = c_0 \mathbf{t}$$

makes a convenient notation for the number in individuals in the first generation of $\mathbf{t}$, that is the number of children of the root 0 of $\mathbf{t}$. Denote the trivial family tree with the single vertex 0 by $\bullet$. Starting from $r_0 \mathbf{t} = \bullet$, a family tree $\mathbf{t}$ is conveniently specified as the unique tree $\mathbf{t}$ such that $r_h \mathbf{t} = \mathbf{t}^{(h)}$ for all h for some sequence of trees $\mathbf{t}^{(h)} \in \mathbf{T}^{(h)}$ determined recursively as follows. Given that $\mathbf{t}^{(h)} \in \mathbf{T}^{(h)}$ has been defined, the set of vertices $\operatorname{gen}(h, \mathbf{t}) = \operatorname{gen}(h, \mathbf{t}^{(h)}) = r_h \mathbf{t} - r_{h-1}\mathbf{t}$ is determined, hence so is the size $Z_h \mathbf{t} = Z_h \mathbf{t}^{(h)}$ of this set; for each possible choice of $Z_h \mathbf{t}$ non-negative integers $(c_v, v \in \operatorname{gen}(h, \mathbf{t}))$, there is a unique $\mathbf{t}^{(h+1)} \in \mathbf{T}^{(h+1)}$ such that $r_h \mathbf{t}^{(h+1)} = \mathbf{t}^{(h)}$ and $c_v \mathbf{t}^{(h+1)} = c_v$ for all $v \in \operatorname{gen}(h, \mathbf{t})$. So a unique tree $\mathbf{t} \in \mathbf{T}^{(\infty)}$ is determined by specifying for each $h \geq 0$ the way in which these $Z_h \mathbf{t}$ non-negative integers are chosen given that $r_h \mathbf{t} = \mathbf{t}^{(h)}$ for some $\mathbf{t}^{(h)} \in \mathbf{T}^{(h)}$.

A *random family tree* is a random element of $\mathbf{T}^{(\infty)}$, formally specified by its sequence of height restrictions, say $\mathcal{T} = (r_h \mathcal{T}, h = 0, 1, \ldots)$, where each $r_h \mathcal{T}$ is a random variable with values in the countable set $\mathbf{T}^{(h)}$, and $r_h \mathcal{T} = r_h(r_{h+1}\mathcal{T})$ for all h. The *distribution of* $\mathcal{T}$, denoted $\operatorname{dist}(\mathcal{T})$, is then determined by the sequence of distributions of $r_h \mathcal{T}$ for $h \geq 0$. Such a distribution is determined by a specification for each $h \geq 0$ of the joint conditional distribution given $r_h \mathcal{T}$ of the numbers of children $c_v \mathcal{T}$ as v ranges over $\operatorname{gen}(h, r_h \mathcal{T})$.

Define convergence of distributions on $\mathbf{T}^{(\infty)}$ by weak convergence relative to the product of discrete topologies on $\mathbf{T}^{(h)}$. That is, for random family trees $\mathcal{T}_n, n = 1, 2, \ldots$ and $\mathcal{T}$, we say that $\mathcal{T}_n$ converges in distribution to $\mathcal{T}$, and write either $\mathcal{T}_n \xrightarrow{d} \mathcal{T}$, or $\mathrm{dist}(\mathcal{T}_n) \to \mathrm{dist}(\mathcal{T})$, or $\lim_n \mathrm{dist}(\mathcal{T}_n) = \mathrm{dist}(\mathcal{T})$ if

$$P(r_h\mathcal{T}_n = \mathbf{t}) \to P(r_h\mathcal{T} = \mathbf{t}) \;\; \forall h \geq 0, \mathbf{t} \in \mathbf{T}. \tag{1}$$

2.2 Galton-Watson trees

Let $p(\cdot) = (p(0), p(1), \ldots)$ be a probability distribution on the non-negative integers with $p(1) < 1$. Following [22, 25, 34, 35], call a random family tree $\mathcal{G}$ a *Galton-Watson (GW) tree with offspring distribution* $p(\cdot)$ if the number of children $Z\mathcal{G}$ of the root has distribution $p(\cdot)$:

$$P(Z\mathcal{G} = n) = p(n) \qquad \forall\; n \geq 0$$

and for each $h = 1, 2, \ldots$, conditionally given $r_h\mathcal{G} = \mathbf{t}^{(h)}$, the numbers of children $c_v\mathcal{G}, v \in \mathrm{gen}(h, \mathbf{t}^{(h)})$, are i.i.d. according to $p(\cdot)$. Equivalently, for each $h \geq 1$ the distribution of $r_h\mathcal{G}$ is determined by the formula

$$P(r_h\mathcal{G} = \mathbf{t}) = \prod_{v \in r_{h-1}\mathbf{t}} p(c_v\mathbf{t}) \qquad \forall\; \mathbf{t} \in \mathbf{T}^{(h)} \tag{2}$$

where the product is over all vertices v of $\mathbf{t}$ of height at most $h-1$. The restriction of the distribution of $\mathcal{G}$ on $\mathbf{T}^{(\infty)}$ to the set $\mathbf{T}$ of finite family trees is then given by the formula

$$P(\mathcal{G} = \mathbf{t}) = \prod_{v \in \mathbf{t}} p(c_v\mathbf{t}) \qquad \forall\; \mathbf{t} \in \mathbf{T} \tag{3}$$

where the product is over all vertices v of $\mathbf{t}$.

Denote the mean of the offspring distribution of $\mathcal{G}$ by μ:

$$\mu := E(Z\mathcal{G}) = \sum_n np(n).$$

It is well known that $P(\#\mathcal{G} < \infty) = 1$, or equivalently $P(Z_h\mathcal{G} > 0) \to 0$ as $h \to \infty$, if and only if $\mu \leq 1$. So for $\mu \leq 1$ the distribution of $\mathcal{G}$ is completely determined by formula (3). Let

$$S_n = \sum_{i=1}^{n} X_i \text{ where the } X_i \text{ are i.i.d. copies of} \tag{4}$$

Whatever μ, it is known [15] that the distribution of the *total progeny* $\#\mathcal{G}$ on the event $(\#\mathcal{G} < \infty)$ is given by the formula

$$P(\#\mathcal{G} = n) = \frac{1}{n} P(S_n = n-1) \quad \forall\ n = 1, 2, \ldots \tag{5}$$

Let $k \geq 1$. Given $Z\mathcal{G} = k$, for $1 \leq i \leq k$ let $\mathcal{G}(i)$ be the subtree of $\mathcal{G}$ formed by the ith child of the root and all its descendants, and observe that the associated family trees $\mathrm{fam}(\mathcal{G}(i))$ for $1 \leq i \leq k$ are i.i.d. copies of $\mathcal{G}$. So

$$\mathrm{dist}(\#\mathcal{G} \mid Z\mathcal{G} = k) = \mathrm{dist}\left(1 + \sum_{i=1}^{k} \#_i\right) \tag{6}$$

where $\#_1, \#_2, \ldots$ are i.i.d. copies of $\#\mathcal{G}$. For all $k \geq 1$ and $n \geq 0$

$$P(\#\mathcal{G} = n+1 \mid Z\mathcal{G} = k) = P\left(\sum_{i=1}^{k} \#_i = n\right) = \frac{k}{n} P(S_n = n-k) \tag{7}$$

for S_n as in (4), where the first equality spells out (6), and the second equality generalizes (5). See [15, 27, 39, 47, 48] for derivations of this formula and other interpretations of the distribution of $\#\mathcal{G}$ involving random walks and queues.

2.3 Poisson-Galton-Watson trees

For $\mu \geq 0$ let $\mathcal{G}_\mu$ be a GW tree with the *Poisson*(μ) offspring distribution $p_\mu(n) := e^{-\mu}\mu^n/n!$. Denote the distribution of $\mathcal{G}_\mu$ on $\mathbf{T}^{(\infty)}$ by $\mathrm{PGW}(\mu)$. ¿From (3) and (2)

$$P(\mathcal{G}_\mu = \mathbf{t}) = e^{-\mu\#\mathbf{t}} \mu^{\#\mathbf{t}-1} \prod_{v \in \mathbf{t}} \frac{1}{(c_v \mathbf{t})!} \qquad \forall\ \mathbf{t} \in \mathbf{T} \tag{8}$$

$$P(r_h \mathcal{G}_\mu = \mathbf{t}) = P(\mathcal{G}_\mu = \mathbf{t}) \exp(\mu Z_h(\mathbf{t})) \quad \forall\ \mathbf{t} \in \mathbf{T}^{(h)}. \tag{9}$$

In this case, S_n in (4) and (5) has Poisson$(n\mu)$ distribution. So from (5) the total progeny of a $\mathrm{PGW}(\mu)$ tree has probability distribution P_μ on $\{1, 2, \cdots, \infty\}$ which is known as the *Borel* (μ) *distribution* [12, 35, 50]:

$$P(\#\mathcal{G}_\mu = n) = P_\mu(n) := \frac{(\mu n)^{n-1}}{n!} e^{-\mu n} \qquad \forall\ n = 1, 2, \ldots \tag{10}$$

¿From (7), the sum of k independent random variables $N_\mu(i)$, each with the Borel(μ) distribution (10), has distribution on $\{1, 2, \cdots, \infty\}$ specified by

$$P\left(\sum_{i=1}^{k} N_\mu(i) = n\right) = \frac{k}{n} \frac{(\mu n)^{n-k}}{(n-k)!} e^{-\mu n} \qquad \forall\, n = k, k+1, \ldots \tag{11}$$

This is the *Borel-Tanner distribution* [21, 49, 50] with parameters k and μ.

2.4 Uniform Random Trees

Let $\mathbf{R}_{[n]}$ be the set of all rooted trees labeled by $[n] := \{1, 2, \ldots, n\}$. For a rooted labeled tree $\mathbf{t}$ with n vertices, let $\tilde{\mathbf{t}}$ denote the corresponding rooted unlabeled tree. Formally, define $\tilde{\mathbf{t}}$ to be the set of all trees $\mathbf{s} \in \mathbf{R}_{[n]}$ obtained by some relabeling of the vertices of $\mathbf{t}$ by $[n]$. So $\tilde{\mathbf{t}} \in \tilde{\mathbf{R}}_{[n]}$ where $\tilde{\mathbf{R}}_{[n]}$ is a set of equivalence classes of elements of $\mathbf{R}_{[n]}$. Let $\mathcal{U}_n$ be a random tree with uniform distribution on $\mathbf{R}_{[n]}$. Aldous [4] observed that for $\mathcal{G}_\mu$ a PGW(μ) tree

$$(\tilde{\mathcal{G}_\mu} \,|\, \#\mathcal{G}_\mu = n) \stackrel{d}{=} \tilde{\mathcal{U}_n}. \tag{12}$$

That is, ($\mathcal{G}_\mu$ given $\#\mathcal{G}_\mu = n$) and $\mathcal{U}_n$ induce identical distributions on $\tilde{\mathbf{R}}_{[n]}$ when unlabeled. To put this another way, fix $\mu > 0$ and generate a PGW family tree $\mathcal{G}_\mu$. Given that verts($\mathcal{G}_\mu$) $= V$ for some set V with $\#V = n$, let $\mathcal{G}_\mu^\dagger \in \mathbf{R}_{[n]}$ be $\mathcal{G}_\mu$ relabeled by a uniform random permutation $\sigma : V \to [n]$. Then (12) amounts to:

$$\mathrm{dist}(\mathcal{G}_\mu^\dagger \,|\, \#\mathcal{G}_\mu = n) = \mathrm{dist}(\mathcal{U}_n). \tag{13}$$

Call a function Ψ of a rooted labeled tree $\mathbf{t}$ an *invariant* if $\Psi(\mathbf{t}) = \Psi(\mathbf{s})$ whenever $\mathbf{s}$ is a relabeling of $\mathbf{t}$. For example, the number $Z_h\mathbf{t}$ of vertices of $\mathbf{t}$ at height h is an invariant. So is the matrix $M(\mathbf{t}) = (M_{h,c}(\mathbf{t}), h \geq 0, c \geq 0)$ where $M_{h,c}(\mathbf{t})$ is the number of individuals in generation h of $\mathbf{t}$ that have c children. The identity (12) can be restated as

$$(\Psi(\mathcal{G}_\mu) \,|\, \#\mathcal{G}_\mu = n) \stackrel{d}{=} \Psi(\mathcal{U}_n) \qquad \forall \text{ invariant } \Psi \tag{14}$$

This identity for $\Psi = M$ and $\Psi = (Z_1, \ldots, Z_n)$ was discovered earlier and exploited by Kolchin [26, 27]. The following proposition records a sharper result, implicit in the discussion of [5] and explicit in [39], which obviously implies all of these identities (12)-(14). We call formula (15) the $\mathcal{U}_n$-*representation* of $\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n)$.

Proposition 1 *For each $n = 1, 2, \ldots$ the conditional distribution of a PGW(μ) tree $\mathcal{G}_\mu$ given $\#\mathcal{G}_\mu = n$ is the same for all $\mu > 0$, and identical to the distribution of* $\mathrm{fam}(\mathcal{U}_n)$. *This common distribution is given by*

$$P(\mathcal{G}_\mu = \mathbf{t} \,|\, \#\mathcal{G}_\mu = n) = P(\mathrm{fam}(\mathcal{U}_n) = \mathbf{t}) = \frac{1}{n^{n-1}} \frac{n!}{\prod_{v \in \mathbf{t}} (c_v \mathbf{t})!} \tag{15}$$

for all $\mathbf{t} \in \mathbf{T}$ *with* $\#\mathbf{t} = n$.

2.5 Conditioning on non-extinction

An infinite random tree $\mathcal{G}^\infty$, which we call $\mathcal{G}$ *conditioned on non-extinction* is derived in the following proposition from a critical or subcritical GW tree $\mathcal{G}$. The probabilistic description of $\mathcal{G}^\infty$ involves the *size-biased* distribution p^* associated with a probability distribution $p(\cdot)$ on the non-negative integers with mean $\mu \in (0, \infty)$:

$$p^*(n) := \mu^{-1} n p(n) \qquad \forall\, n \geq 0. \tag{16}$$

Here is an exact statement of "known result (ii)" in the Introduction.

Proposition 2 (Kesten [25]). *Let $Z_n\mathcal{G} := \#\mathrm{gen}(n, \mathcal{G})$ be the number of individuals in the nth generation of a GW tree $\mathcal{G}$ with offspring distribution $p(\cdot)$ such that $p(0) < 1$ and $\mu \leq 1$. Then*

(i)

$$\mathrm{dist}(\mathcal{G} \,|\, Z_n\mathcal{G} > 0) \to \mathrm{dist}(\mathcal{G}^\infty) \qquad \text{as } n \to \infty \tag{17}$$

where $\mathrm{dist}(\mathcal{G}^\infty)$ *is the distribution of a random family tree $\mathcal{G}^\infty$ specified by*

$$P(r_h\mathcal{G}^\infty = \mathbf{t}) = \mu^{-h}(Z_h\mathbf{t})P(r_h\mathcal{G} = \mathbf{t}) \quad \forall\ \mathbf{t} \in \mathbf{T}^{(h)}, h \geq 0. \tag{18}$$

(ii) Almost surely $\mathcal{G}^\infty$ contains a unique infinite path $(\mathrm{root} = V_0, V_1, V_2, \ldots)$ *such that V_{h+1} is a child of V_h for every $h = 0, 1, 2, \ldots$.*

(iii) For each h the joint distribution of $r_h\mathcal{G}^\infty$ and V_h is given by

$$P(r_h\mathcal{G}^\infty = \mathbf{t}, V_h = v) = \mu^{-h} P(r_h\mathcal{G} = \mathbf{t})\ \forall\ \mathbf{t} \in \mathbf{T}^{(h)}, v \in \mathrm{gen}(h, \mathbf{t}) \tag{19}$$

(iv) The joint distribution of $(V_0, V_1, V_2, \ldots)$ and $\mathcal{G}^\infty$ is determined recursively as follows: for each $h = 0, 1, 2, \ldots$, given $(V_0, V_1, \ldots, V_h)$ and $r_h\mathcal{G}^\infty$, the numbers of children $c_v\mathcal{G}^\infty$ are independent as v ranges over $\mathrm{gen}(h, \mathcal{G}^\infty)$, *with distribution $p(\cdot)$ for $v \neq V_h$, and with the size-biased distribution $p^*(\cdot)$ for $v = V_h$; given also the numbers of children $c_v\mathcal{G}^\infty$ for $v \in \mathrm{gen}(h, r_h\mathcal{G}^\infty)$, the vertex V_{h+1} has uniform distribution on the set of $c_{V_h}\mathcal{G}^\infty$ children of V_h.*

That (18) defines a probability distribution for an infinite family tree $\mathcal{G}^\infty$ follows from the well known fact that $(\mu^{-n} Z_n \mathcal{G}, n = 0, 1, \ldots)$ is a non-negative martingale with expectation 1. The sequence of height restrictions $(r_n \mathcal{G}^\infty, n = 0, 1, \cdots)$ which determines $\mathcal{G}^\infty$ is a Markov chain with state space $\mathbf{T}$ obtained as the *Doob h-transform* of the Markov chain $(r_n \mathcal{G}, n = 0, 1, \cdots)$ via the space-time harmonic function $h(n, \mathbf{t}) := \mu^{-n} Z_n \mathbf{t}$. See [41] for background and other applications of h-transforms, and [31] for an elegant treatment of the recursive construction (iv) of $\mathcal{G}^\infty$ and the infinite path (V_h), which we call the *spine* of $\mathcal{G}^\infty$. As observed in [31], the construction (iv) of V_h and $\mathcal{G}^\infty$ such that (18) and (19) hold can also be carried out in the supercritical case $1 < \mu < \infty$. While our focus in this paper will be on the case $0 < \mu \leq 1$, we note in passing that for $\mu > 1$ the path (V_h) is almost surely not the only infinite path from the root in $\mathcal{G}^\infty$; rather, there are uncountably many such paths almost surely. Also, the conditional limit theorem (17) does not hold for $\mu > 1$ with $\mathcal{G}^\infty$ constructed via (18). Rather, there is the elementary result that

$$\lim_{n\to\infty} \text{dist}(\mathcal{G} \,|\, Z_n \mathcal{G} > 0) = \text{dist}(\mathcal{G} \,|\, \#\mathcal{G} = \infty) \quad \text{for } \mu > 1$$

where the right side does not have the same distribution as $\mathcal{G}^\infty$ defined by (18), except when $p(\cdot)$ is degenerate.

In particular, Proposition 2 contains the classical result ([9] sec. I.8) that for the usual integer-valued GW process $(Z_h \mathcal{G}, h = 0, 1, \ldots)$ started at $Z_0 = 1$, for $\mu \leq 1$ there is the conditioned limit theorem

$$\lim_{n\to\infty} \text{dist}(Z_1 \mathcal{G}, \ldots, Z_h \mathcal{G} \,|\, Z_n \mathcal{G} > 0) = \text{dist}(Z_1 \mathcal{G}^\infty, \ldots, Z_h \mathcal{G}^\infty) \quad \forall\, h = 1, 2, \ldots \tag{20}$$

where $(Z_h \mathcal{G}^\infty, h = 0, 1, \ldots)$ is a Markov chain with state space $\mathbb{N}$ and homogeneous transition probabilities defined as follows. Given $Z_h \mathcal{G}^\infty = m$ say, the number $Z_{h+1} \mathcal{G}^\infty$ of individuals in the next generation of $\mathcal{G}^\infty$ is distributed as the sum of m independent random variables, with $m - 1$ of these variables distributed according to the offspring distribution $p(\cdot)$ of $Z\mathcal{G}$, and one variable distributed according to the size-biased offspring distribution $p^*(\cdot)$. In other words, the process $(Z_h \mathcal{G}^\infty - 1, h = 0, 1, \ldots)$ is a branching process with immigration, starting from an initial population of zero, with offspring distribution $p(\cdot)$ and immigration distribution $p^{*-}(\cdot)$, where $p^{*-}(\cdot)$ is the distribution of $Z^* - 1$ for Z^* with the size-biased offspring distribution $p^*(\cdot)$, that is

$$p^{*-}(n) := \mu^{-1}(n+1)p(n+1) \qquad \forall\, n = 0, 1, \ldots \tag{21}$$

It is elementary and well known that

$$p^{*-}(\cdot) = p(\cdot) \text{ iff } p(\cdot) \text{ is Poisson}(\mu) \text{ for some } \mu \geq 0.$$

The following corollary is an easy consequence of this fact combined with the previous Proposition.

Corollary 3 (Spinal Decomposition of $\mathcal{G}^\infty$). *In the setting of Proposition 2, with (V_h) the infinite spine of $\mathcal{G}^\infty$ derived by conditioning $\mathcal{G}$ on non-extinction, for $i = 0, 1, \ldots$ let $\mathcal{G}^{(i)}$ be the family tree derived from the subtree of $\mathcal{G}^\infty$ with root V_i in the random forest obtained from $\mathcal{G}^\infty$ by deleting each edge along the spine, and let V_{i+1} be the J_ith child of V_i. Then*

(i) the trees $\mathcal{G}^{(i)}, i = 0, 1, \ldots$ are independent and almost surely finite, with identical distribution

$$P(\mathcal{G}^{(i)} = \mathbf{t}) = \sum_{m=0}^{\infty} P(\mathcal{G} = \mathbf{t} \,|\, Z\mathcal{G} = m)\, p^{*-}(m) \qquad \forall\, \mathbf{t} \in \mathbf{T}; \tag{22}$$

(ii) the trees $\mathcal{G}^{(i)}$ have the same distribution as $\mathcal{G}$ iff $p(\cdot)$ is Poisson(μ);

(iii) conditionally given $(\mathcal{G}^{(i)}, i = 0, 1, \ldots)$ the ranks J_i are independent and J_i has uniform $[Z\mathcal{G}^{(i)} + 1]$ distribution, where $Z\mathcal{G}^{(i)} = c_{V_i}\mathcal{G}^\infty - 1$.

The common distribution of the $\mathcal{G}^{(i)}$ described by (22) is that of a modified GW tree, in which the number of the first generation individuals has distribution $p^{*-}(\cdot)$, while these and all subsequent individuals have offspring distribution $p(\cdot)$. Note that $V_h = (J_0, \ldots, J_{h-1}) \in \mathbb{N}^h$ for all $h \geq 1$. So the spinal decomposition specifies the joint distribution of the spine of $\mathcal{G}^\infty$ and the sequence of finite family trees derived by cutting all edges of the spine. This determines the distribution of $\mathcal{G}^\infty$, for it is clear that $\mathcal{G}^\infty$ is a measurable function of (V_h) and the $\mathcal{G}^{(i)}$. We record now for later use the following consequence of Proposition 2.

Lemma 4 *In the setting of Proposition 2, with (V_h) the infinite spine of $\mathcal{G}^\infty$ derived by conditioning $\mathcal{G}$ on non-extinction, let H be a non-negative integer random variable independent of $\mathcal{G}^\infty$, and let $\mathcal{G}_{(H)}$ be the family tree derived from the finite subtree of $\mathcal{G}^\infty$ that contains the root after $\mathcal{G}^\infty$ is cut into two* *subtrees by deletion of the edge (V_H, V_{H+1}). Then for each $\mathbf{t} \in \mathbf{T}$ and each vertex v of $\mathbf{t}$ at height h*

$$P(\mathcal{G}_{(H)} = \mathbf{t}, V_H = v) = P(H = h)\mu^{-h}\rho^{*-}(c_v\mathbf{t})P(\mathcal{G} = \mathbf{t}) \tag{23}$$

where

$$\rho^{*-}(n) := \frac{p^{*-}(n)}{p(n)} = \frac{(n+1)p(n+1)}{\mu p(n)} \quad \forall\, n = 0, 1, \ldots \tag{24}$$

for $p(\cdot)$ the offspring distribution of $\mathcal{G}$ and $\mu := \sum_n np(n) \leq 1$.

Proof. By conditioning on H it suffices to prove (23) for a constant H, say $H = h$. Let $\mathbf{t}_h = r_h\mathbf{t}$. Then for each v at height h in $\mathbf{t}$ the left side of (23) equals

$$P(r_h\mathcal{G}_{(h)} = \mathbf{t}_h, V_h = v)\, P(\mathcal{G}_{(h)} = \mathbf{t} \,|\, r_h\mathcal{G}_{(h)} = \mathbf{t}_h, V_h = v) \tag{25}$$

But for $H = h$ fixed, $r_h\mathcal{G}_{(h)} = r_h\mathcal{G}^\infty$ by construction, so (18) gives

$$P(r_h\mathcal{G}_{(h)} = \mathbf{t}_h, V_h = v) = \mu^{-h}P(r_h\mathcal{G} = \mathbf{t}_h). \tag{26}$$

Also, given $r_h\mathcal{G}_{(h)} = \mathbf{t}_h$ and $V_h = v$, according to part (iv) of Proposition 2, $\mathcal{G}_{(h)}$ develops over generations $h+1, h+2, \ldots$ much like $\mathcal{G}$, with individuals having independent numbers of offspring, except that in $\mathcal{G}_{(h)}$ each individual except v has offspring distribution $p(\cdot)$, whereas v has offspring distribution $p^{*-}(\cdot)$. By consideration of the product formulae (2) and (3), it follows that for any particular tree $\mathbf{t}$ in which v has n offspring,

$$P(\mathcal{G}_{(h)} = \mathbf{t} \,|\, r_h\mathcal{G}_{(h)} = \mathbf{t}_h, V_h = v) = \rho^{*-}(n)P(\mathcal{G} = \mathbf{t} \,|\, r_h\mathcal{G} = \mathbf{t}_h) \tag{27}$$

for $\rho^{*-}(n)$ as in (24). Combine (26) and (27) in (25) to obtain (23) for a fixed $H = h$. □

Conditioning on the total progeny. Kennedy [24] obtained an analog of the conditioned limit theorem (20) as $n \to \infty$ with conditioning on $\#\mathcal{G} = n$ instead of $Z_n\mathcal{G} > 0$, where $\#\mathcal{G} = \sum_n Z_n\mathcal{G}$ is the total progeny. His assumption on the offspring distribution $p(\cdot)$ is that the generating function $g(s) := \sum_n p(n)s^n$ satisfies

$$\exists\ a > 0 \text{ with } g(a) = ag'(a) < \infty, \text{ and } g''(a) < \infty. \tag{28}$$

Reinterpreting his argument in terms of family trees gives the convergence assertion in (29): the equality follows easily from the product formula (3).

Proposition 5 *Let $\mathcal{G}$ be a GW tree whose offspring generating function g satisfies (28). Let $\overline{\mathcal{G}}$ be the critical GW tree whose offspring generating function is $g(a)^{-1}g(as) = \sum_n g(a)^{-1}a^n p_n s^n$ for a as in (28). Then*

$$\mathrm{dist}(\mathcal{G} \,|\, \#\mathcal{G} = n) = \mathrm{dist}(\overline{\mathcal{G}} \,|\, \#\overline{\mathcal{G}} = n) \to \mathrm{dist}(\overline{\mathcal{G}}^{\infty}) \quad \text{as } n \to \infty \tag{29}$$

where $\overline{\mathcal{G}}^{\infty}$ is $\overline{\mathcal{G}}$ conditioned on non-extinction.

In terms of the offspring mean μ, condition (28) is always satisfied if $\mu > 1$ and $p(\cdot)$ is nondegenerate, with $a < 1$. If $\mu = 1$ then (28) holds if and only if $\sum_n n^2 p(n) < \infty$, in which case $a = 1$; then $\overline{\mathcal{G}} = \mathcal{G}$ and $\overline{\mathcal{G}}^{\infty} = \mathcal{G}^{\infty}$. If $\mu < 1$ then (28) requires $p(n)$ to decay exponentially, and $a > 1$; then the distributions of $\mathcal{G}^{\infty}$ and $\overline{\mathcal{G}}^{\infty}$ are different.

Assume for this paragraph that (28) holds, and consider what happens if we condition on $(\#\mathcal{G} \geq n)$ instead of $(\#\mathcal{G} = n)$ and then let $n \to \infty$:

$$\lim_{n\to\infty} \mathrm{dist}(\mathcal{G} \,|\, \#\mathcal{G} \geq n) = \begin{cases} \mathrm{dist}(\mathcal{G} \,|\, \#\mathcal{G} = \infty) & \text{if } \mu > 1 \\ \mathrm{dist}(\mathcal{G}^{\infty}) & \text{if } \mu = 1 \\ \mathrm{dist}(\overline{\mathcal{G}}^{\infty}) & \text{if } \mu < 1 \end{cases}$$

The first case is elementary, and the second two cases follow easily from Proposition 5. Note the paradoxical fact that while

$$\bigcap_n (Z_n\mathcal{G} > 0) = \bigcap_n (\#\mathcal{G} \geq n) = (\#\mathcal{G} = \infty)$$

and both intersections involve decreasing sequences of events,

$$\lim_n \mathrm{dist}(\mathcal{G} \,|\, Z_n\mathcal{G} > 0) = \lim_n \mathrm{dist}(\mathcal{G} \,|\, \#\mathcal{G} \geq n) \;\text{ only if } \mu \geq 1. \tag{30}$$

For $\mu > 1$ both limits in (30) are the naively defined $\mathrm{dist}(\mathcal{G} \,|\, \#\mathcal{G} = \infty)$. For $\mu = 1$ both limits yield $\mathrm{dist}(\mathcal{G}^{\infty})$, which suggests the intuitive interpretation of $\mathcal{G}^{\infty}$ as $\mathrm{dist}(\mathcal{G} \,|\, \#\mathcal{G} = \infty)$. However, such interpretations are potentially slippery, as shown by the fact that for $0 < \mu < 1$

$$\lim_n \mathrm{dist}(\mathcal{G} \,|\, Z_n\mathcal{G} > 0) = \mathrm{dist}(\mathcal{G}^{\infty}) \neq \mathrm{dist}(\overline{\mathcal{G}}^{\infty}) = \lim_n \mathrm{dist}(\mathcal{G} \,|\, \#\mathcal{G} \geq n). \tag{31}$$

3 Pruning Random Trees

Let $\mathcal{T}$ be a random family tree. Call a $\mathbf{T}^{(\infty)}$-valued process $(\mathcal{T}_u, 0 \le u \le 1)$ a *uniform pruning* of $\mathcal{T}$ if

$$\mathcal{T}_1 = \mathcal{T} \text{ almost surely and } (\mathcal{T}_u, 0 \le u \le 1) \stackrel{d}{=} (\mathcal{T}(u), 0 \le u \le 1) \tag{32}$$

where $(\mathcal{T}(u), 0 \le u \le 1)$ is constructed as follows from some $\mathcal{T}(1)$ with $\mathcal{T}(1) \stackrel{d}{=} \mathcal{T}$. Here $\stackrel{d}{=}$ denotes equality in distribution, meaning equality of finite dimensional distributions in a display such as (32). Suppose that given $\mathcal{T}(1)$ there are independent uniform$(0,1)$ random variables ξ_e attached to the edges e of $\mathcal{T}(1)$; let $\mathcal{T}^\dagger(u)$ be the component containing the root in the subgraph of $\mathcal{T}(1)$ consisting of those edges e with $\xi_e < u$, and let $\mathcal{T}(u) = \operatorname{fam}(\mathcal{T}^\dagger(u))$ be $\mathcal{T}^\dagger(u)$ relabeled as a family tree.

Let I be an interval of the form either $[0, \zeta]$ for some $0 < \zeta < \infty$ or $[0, \zeta)$ for some $0 < \zeta \le \infty$. Call a $\mathbf{T}^{(\infty)}$-valued process $(\mathcal{T}_t, t \in I)$ a *uniform pruning process* if

$$(\mathcal{T}_{ut}, 0 \le u \le 1) \text{ is a uniform pruning of } \mathcal{T}_t \text{ for all } t \in I. \tag{33}$$

If $(\mathcal{T}_u, 0 \le u \le 1)$ is a uniform pruning of $\mathcal{T}_1$ then $(\mathcal{T}_u, 0 \le u \le 1)$ is a uniform pruning process, and almost surely

$$\mathcal{T}_0 = \bullet, \text{ the single-vertex tree, and } \lim_{u \uparrow 1} \mathcal{T}_u = \mathcal{T}_1. \tag{34}$$

The second equality means that for almost every ω in the basic probability space, for each h there exists a $u(h, \omega) < 1$ such that $r_h \mathcal{T}_u(\omega) = r_h \mathcal{T}_1(\omega)$ for all $u(h, \omega) < u \le 1$. It is easily seen that every uniform pruning process is an inhomogeneous Markov process. All uniform pruning processes $(\mathcal{T}_t, t \in I)$ share the same co-transition probabilities, which have an obvious invariance under scaling: for $0 < u \le 1$ and $z \in I$ with $z > 0$,

$$\operatorname{dist}(\mathcal{T}_{uz} \,|\, \mathcal{T}_z = \mathbf{t}) = \operatorname{dist}(\mathcal{T}_u^{(\mathbf{t})}) \ \forall\, \mathbf{t} \in \mathbf{T}^{(\infty)} \tag{35}$$

where $(\mathcal{T}_u^{(\mathbf{t})}, 0 \le u \le 1)$ is a uniform pruning of the fixed tree $\mathbf{t}$. For a finite tree $\mathbf{t} \in \mathbf{T}$ these transition probabilities are described by the following formula, which is derived by conditioning on which subtree $\mathbf{s}$ of $\mathbf{t}$ remains containing the root after cutting each edge of $\mathbf{t}$ with probability $1 - u$:

$$P(\mathcal{T}_{uz} = \mathbf{r} \,|\, \mathcal{T}_z = \mathbf{t}) = P(\mathcal{T}_u^{(\mathbf{t})} = \mathbf{r}) = \sum_{\mathbf{s}} u^{\#\mathbf{s}-1}(1-u)^{n(\mathbf{s},\mathbf{t})} \tag{36}$$

where the sum ranges over all subtrees $\mathbf{s}$ of $\mathbf{t}$ with $0 \in \mathbf{s}$ and $\mathrm{fam}(\mathbf{s}) = \mathbf{r}$, and $n(\mathbf{s}, \mathbf{t})$ is the number of edges (v, w) of $\mathbf{t}$ such that $v \in \mathbf{s}$ and $w \in \mathbf{t} - \mathbf{s}$. This formula determines the co-transition probabilities of every uniform pruning process, for it is easily seen that a $\mathbf{T}^{(\infty)}$-valued process $(\mathcal{T}_t)$ is a uniform pruning process iff for each $h \geq 0$ the height restricted $\mathbf{T}$-valued process $(r_h \mathcal{T}_t)$ is a uniform pruning process.

If $(\mathcal{T}_u, 0 \leq u \leq 1)$ is derived from $\mathcal{T}_1$ by uniform pruning, the size $Z\mathcal{T}_u$ of the first generation of $\mathcal{T}_u$ is distributed as the sum of $Z\mathcal{T}_1$ independent Bernoulli(u) variables. In terms of probability generating functions [16]

$$g_u(s) := \sum_{n=0}^{\infty} s^n P(Z\mathcal{T}_u = n) \text{ is given by } g_u(s) = g_1(1 - u + us). \tag{37}$$

In particular, if $\mathrm{dist}(Z\mathcal{T}_1) = \mathrm{Poisson}(\mu)$ for some $\mu \geq 0$ then $\mathrm{dist}(Z\mathcal{T}_u) = \mathrm{Poisson}(u\mu)$.

The following two lemmas record some more technical properties of uniform pruning processes for ease of later reference.

Lemma 6 *Let $(\mathcal{T}_u^{(n)}, u \in I_n)$ be a sequence of uniform pruning processs and let $t_n \in I_n$ be such that $t_n \to 1$ and $\mathcal{T}_{t_n}^{(n)} \xrightarrow{d} \mathcal{T}$ as $n \to \infty$ for some random family tree $\mathcal{T}$. Then*

(i) for each $\epsilon > 0$

$$\mathrm{dist}(\mathcal{T}_u^{(n)}, 0 \leq u \leq 1 - \epsilon) \to \mathrm{dist}(\mathcal{T}_u, 0 \leq u \leq 1 - \epsilon) \tag{38}$$

where $(\mathcal{T}_u, 0 \leq u \leq 1)$ is a uniform pruning of $\mathcal{T}_1$ with $\mathcal{T}_1 \stackrel{d}{=} \mathcal{T}$;

(ii) if $t_n \geq 1$ for all n then (38) holds also for $\epsilon = 0$.

Proof. The convergence in (38) should be understood as convergence of finite dimensional distributions of height restricted processes for each finite height h. The probability that a height restricted uniform pruning process $(\mathcal{T}_u)$ passes through some sequence of trees $\mathbf{t}_i, 1 \leq i \leq k$ at times $0 < u_0 < \cdots < u_k$ is the product of $P(\mathcal{T}_{u_k} = \mathbf{t}_k)$ and a sequence of co-transition probabilities of the form (36). The conclusions of the lemma follow from the fact that the function of $(u, \mathbf{r}, \mathbf{t})$ displayed in (36) is continuous in u for all $\mathbf{r}, \mathbf{t} \in \mathbf{T}$. □

Lemma 7 *If $(\mathcal{T}_u, 0 \leq u < 1)$ is a uniform pruning process then $\mathcal{T}_1 := \lim_{u \uparrow 1} \mathcal{T}_u$ exists almost surely, and $(\mathcal{T}_u, 0 \leq u \leq 1)$ is a uniform pruning of $\mathcal{T}_1$.*

Proof. The process $(Z\mathcal{T}_u, 0 \le u < 1)$ is almost surely increasing, hence has an almost sure limit, say $Z_1 \le \infty$. Moreover, $Z_1 < \infty$ almost surely, as the following argument by contradiction shows. If $P(Z_1 = \infty) = 3\delta > 0$, than for all $\epsilon < \delta$ there exists $u(\epsilon) \in (1/2, 1)$ with $P(Z\mathcal{T}_{u(\epsilon)} > 3/\epsilon) > 2\delta$. But from the pruning property, conditionally given $Z\mathcal{T}_{u(\epsilon)} > 3/\epsilon$ the number $Z\mathcal{T}_{1/2}$ exceeds the number of successes in $3/\epsilon$ independent trials with success probability $(2u(\epsilon))^{-1} > 1/2$. By the law of large numbers, $P(Z\mathcal{T}_{1/2} > 1/\epsilon \,|\, \mathcal{T}_{u(\epsilon)} > 3/\epsilon) \to 1$ as $\epsilon \to 0$. Therefore, for all sufficiently small ϵ this conditional probability exceeds $1/2$, and for such small ϵ we deduce that $P(Z\mathcal{T}_{1/2} > 1/\epsilon) > \delta$. Let $\epsilon \to 0$ to deduce that $P(Z\mathcal{T}_{1/2} = \infty) > \delta$, in contradiction to the fact that $P(Z\mathcal{T}_{1/2} = \infty) = 0$ because $\mathcal{T}_{1/2} \in \mathbf{T}^{(\infty)}$ and $Z\mathbf{t} < \infty$ for all $\mathbf{t} \in \mathbf{T}^{(\infty)}$. Thus $P(Z_1 < \infty) = 1$. Therefore, $\lim_{u \uparrow 1} r_1 \mathcal{T}_u$ exists almost surely and equals $\mathcal{T}_1^{(1)}$ say, where $\mathcal{T}_1^{(1)}$ is the tree of height at most one in which the root has Z_1 children. Now proceed inductively. Suppose for some h that $\lim_{u \uparrow 1} r_h \mathcal{T}_u$ exists almost surely and equals say $\mathcal{T}_1^{(h)} \in \mathbf{T}^{(h)}$. Let $U_h = \inf\{u : r_h \mathcal{T}_u = \mathcal{T}_1^{(h)}\}$. Then $P(U_h < 1) = 1$ by inductive hypothesis, and for each v in generation h of $\mathcal{T}_1^{(h)}$ the number $c_v \mathcal{T}_u$ of children of v in the next generation of $\mathcal{T}_u$ is increasing for $U_h < u < 1$. Therefore, $c_v \mathcal{T}_u$ has an almost sure limit $c_v(1)$ say as $u \uparrow 1$. That the $c_v(1)$ are a.s. finite for all v in generation h of $\mathcal{T}_1^{(h)}$ can be shown by a reprise of the previous argument by contradiction for $h = 0$ after conditioning on $\mathcal{T}_1^{(h)}$. It follows that $\lim_{u \uparrow 1} r_{h+1} \mathcal{T}_u = \mathcal{T}_1^{(h+1)}$ almost surely where $\mathcal{T}_1^{(h+1)} \in \mathbf{T}^{(h+1)}$ is such that $r_h \mathcal{T}_1^{(h+1)} = \mathcal{T}_1^{(h)}$ and each v in generation h of $\mathcal{T}_1^{(h)}$ has $c_v(1)$ children in $\mathcal{T}_1^{(h+1)}$. So by induction, these limits $\mathcal{T}_1^{(h)}$ exist for all h almost surely, and hence almost surely $\lim_{u \uparrow 1} \mathcal{T}_u = \mathcal{T}_1$ where $\mathcal{T}_1 \in \mathbf{T}^{(\infty)}$ is defined by $r_h \mathcal{T}_1 = \mathcal{T}_1^{(h)}$ for all h. □

3.1 Transition rates

The transition rates for a uniform pruning process will be given in Lemma 8. Note that in this paper, Markov processes are constructed in fairly explicit fashion, in contrast to the usual way of specifying a Markov process by stating its transition rates.

For $\mathcal{T}_1$ with $\#\mathcal{T}_1 < \infty$, a uniform pruning process $(\mathcal{T}_u, 0 \le u \le 1)$ is an inhomogeneous Markov chain with step function paths of jump-hold type on the countable state space $\mathbf{T}$. This chain is determined by its co-transition probabilities (36), or by its co-transition rates, which are much simpler and can be described as follows. For $\mathbf{t}$ in $\mathbf{T}^{(\infty)}$ and w a non-root vertex of $\mathbf{t}$, let $v = \text{parent}(w)$. Deleting (v, w) from the set of edges of $\mathbf{t}$ defines a directed graph on verts($\mathbf{t}$) with two component subtrees, say $\mathbf{t}_w$ and $\mathbf{t}^w$, with $\text{root}(\mathbf{t}_w) = \text{root}(\mathbf{t}) = 0$, and $\text{root}(\mathbf{t}^w) = w$. Call $\mathbf{t}_w$ the *remaining tree* and $\mathbf{t}^w$ the *pruned branch* derived by *pruning* $\mathbf{t}$ *below* w, or by *cutting the edge* (v, w) *of* $\mathbf{t}$. ¿From the construction of a uniform pruning process $(\mathcal{T}_u, 0 \leq u \leq 1)$ with independent uniform times, there is the following formula: for all finite family trees $\mathbf{t}$ and $\mathbf{r}$ the co-transition rate $\hat{q}_u(\mathbf{t} \to \mathbf{r})$ from $\mathbf{t}$ to $\mathbf{r}$ at time $0 < u \leq 1$ given $\mathcal{T}_u = \mathbf{t}$ is given by

$$\hat{q}_u(\mathbf{t} \to \mathbf{r}) = u^{-1} \# V(\mathbf{r}, \mathbf{t}) \tag{39}$$

where

$$V(\mathbf{r}, \mathbf{t}) := \{w \in \text{verts}(\mathbf{t}) - \{0\} : \text{fam}(\mathbf{t}_w) = \mathbf{r}\} \tag{40}$$

In words, $\#V(\mathbf{r}, \mathbf{t})$ is the number of ways to choose a non-root vertex w of $\mathbf{t}$ such that if $\mathbf{t}$ is pruned below w, and the remaining tree $\mathbf{t}_w$ is relabeled as a family tree, the result is $\mathbf{r}$. Lemma 9 below gives a more explicit description of how $V(\mathbf{r}, \mathbf{t})$ and $\#V(\mathbf{r}, \mathbf{t})$ are determined by $\mathbf{r}$ and $\mathbf{t}$.

Consider now the forwards transition rates of a uniform pruning process $(\mathcal{T}_u, 0 \leq u \leq 1)$. For two finite family trees $\mathbf{r}$ and $\mathbf{t}$ and $0 < u < 1$ let $q_u(\mathbf{r} \to \mathbf{t})$ be the rate of forwards transitions of from $\mathbf{r}$ to $\mathbf{t}$ at time u, given $\mathcal{T}_u = \mathbf{r}$. Combining (39) and the obvious identity of unconditional rates

$$P(\mathcal{T}_u = \mathbf{r}) q_u(\mathbf{r} \to \mathbf{t}) = P(\mathcal{T}_u = \mathbf{t}) \hat{q}_u(\mathbf{t} \to \mathbf{r}) \qquad \forall\ \mathbf{r}, \mathbf{t} \in \mathbf{T}$$

gives the following formula.

Lemma 8 *For a uniform pruning process $(\mathcal{T}_u, 0 \leq u \leq 1)$ with $\#\mathcal{T}_1 < \infty$, the forwards transition rate from $\mathbf{r}$ to $\mathbf{t}$ at time $0 < u < 1$ is*

$$q_u(\mathbf{r} \to \mathbf{t}) = \frac{\#V(\mathbf{r}, \mathbf{t})}{u} \frac{P(\mathcal{T}_u = \mathbf{t})}{P(\mathcal{T}_u = \mathbf{r})} \tag{41}$$

The meaning of the combinatorial factor $\#V(\mathbf{r}, \mathbf{t})$ can be clarified in terms of the following operation on family trees. Suppose that $\mathbf{r}, \mathbf{s} \in \mathbf{T}^{(\infty)}$, that v is a vertex of $\mathbf{r}$, and that $j \in [c_v \mathbf{r} + 1]$. Let $\mathbf{t}(\mathbf{r}, v, j, \mathbf{s})$ denote the family tree obtained by *attaching the root of* $\mathbf{s}$ *to* $\mathbf{r}$ *as the* j*th child of* v. This is the unique family tree $\mathbf{t}$ whose vertices contain v and a child w of v of rank j such that $\mathrm{fam}(\mathbf{t}_w) = \mathbf{r}$ and $\mathrm{fam}(\mathbf{t}^w) = \mathbf{s}$. Note that $c_v\mathbf{t} = c_v\mathbf{r} + 1$, and that $\#\mathbf{t} = \#\mathbf{r} + \#\mathbf{s}$. The notation is illustrated by Figure 2, in which $\mathbf{t} = \mathbf{t}(\mathbf{r}, v, 2, \mathbf{s}) = \mathbf{t}(\mathbf{r}, v, 3, \mathbf{s})$ so that $\#V(\mathbf{r}, \mathbf{t}) = 2$.

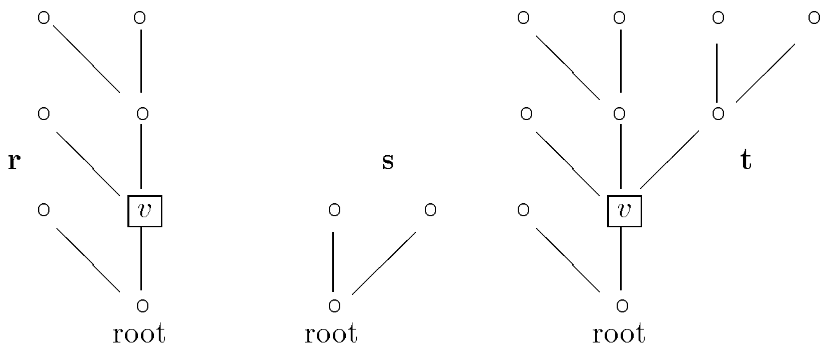

Figure 2

The following lemma is intuitively clear from pictures like Figure 2; we leave its proof to the dedicated reader!

Lemma 9 *For* $\mathbf{r}, \mathbf{t} \in \mathbf{T}$*, let* $V(\mathbf{r}, \mathbf{t}) := \{w \in \mathrm{verts}(\mathbf{t}) - \{0\} : \mathrm{fam}(\mathbf{t}_w) = \mathbf{r}\}$*.*

(i) If $V(\mathbf{r}, \mathbf{t})$ *is not empty, then this set is of the form* $\{w_1, w_2, \cdots, w_k\}$ *where* $k = \#V(\mathbf{r}, \mathbf{t})$ *and the* w_i *are consecutive siblings. That is, the* w_i *have a common parent* $v \in \mathbf{t}$*, and* w_i *is the* $(m + i)$*th child of* v *for some* $m \geq 0$ *and* $1 \leq i \leq k$*.*

(ii) For any particular $w \in V(\mathbf{r}, \mathbf{t})$*, the set* $V(\mathbf{r}, \mathbf{t})$ *is the maximal set* $\{w_1, w_2, \cdots, w_k\}$ *of consecutive children of* $\mathrm{parent}(w)$ *such that* $w = w_i$ *for some* i *and* $\mathrm{fam}(\mathbf{t}^{w_i}) = \mathrm{fam}(\mathbf{t}^w)$ *for all* i*.*

(iii) The number $\#V(\mathbf{r}, \mathbf{t})$ *equals the number of representations of* $\mathbf{t}$ *as* $\mathbf{t} = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s})$*. That is,* $\#V(\mathbf{r}, \mathbf{t})$ *is the number of triples* $(v, j, \mathbf{s})$ *with* $v \in \mathbf{r}, j \in [c_v\mathbf{r} + 1], \mathbf{s} \in \mathbf{T}$ *such that* $\mathbf{t}(\mathbf{r}, v, j, \mathbf{s}) = \mathbf{t}$*.*

(iv) For $\mathbf{r}$ *and* $\mathbf{t}$ *such that* $\#V(\mathbf{r}, \mathbf{t}) \geq 1$*, there is a unique vertex* v *of* $\mathbf{r}$ *and a unique* $\mathbf{s} \in \mathbf{T}$ *such that* $\mathbf{t}(\mathbf{r}, v, j, \mathbf{s}) = \mathbf{t}$ *for some* $j \in [c_v\mathbf{r} + 1]$*.*

(v) For given $\mathbf{r}, v$ *and* $\mathbf{s}$ *let* $I = I(\mathbf{r}, v, \mathbf{s})$ *be the set of* i *such that the descendants in* $\mathbf{r}$ *of the ith child of* v *form a family tree identical to* $\mathbf{s}$*. As* j *ranges over* $i_1 \leq j \leq i_m + 1$ *where* $\{i_1, \ldots, i_m\}$ *is a maximal sequence of consecutive elements of* I*, the tree* $\mathbf{t} = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s})$ *is the same, and such that* $\#V(\mathbf{r}, \mathbf{t}) = m + 1$*, whereas every other choice of* $j \in [c_v\mathbf{r} + 1]$ *defines a tree* $\mathbf{t}' = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s})$ *with* $\#V(\mathbf{r}, \mathbf{t}') = 1$*.*

Given a family tree $\mathbf{r}$, a vertex v of $\mathbf{r}$, and a random family tree $\mathcal{S}$, say that a random family tree $\mathcal{T}$ is constructed by *random attachment of $\mathcal{S}$ to $\mathbf{r}$ at v* if $\mathcal{T} = \mathbf{t}(\mathbf{r}, v, J, \mathcal{S})$ where J has uniform $[c_v\mathbf{r} + 1]$ distribution independently of $\mathcal{S}$. Part (iii) of the above lemma shows that the distribution of $\mathcal{T}$ is then determined by the following formula: for all $v \in \mathbf{r}$ and $j \in [c_v\mathbf{r} + 1]$

$$P(\mathcal{T} = \mathbf{t}) = \frac{\#V(\mathbf{r}, \mathbf{t})P(\mathcal{S} = \mathbf{s})}{c_v\mathbf{r} + 1} \text{ for } \mathbf{t} = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s}). \tag{42}$$

3.2 Pruning a Galton-Watson tree

Throughout this section let $(\mathcal{G}_u, 0 \leq u \leq 1)$ be the $\mathbf{T}^{(\infty)}$-valued process derived by uniform pruning of a GW tree $\mathcal{G}_1$ with offspring distribution $p_1(\cdot)$. We shall describe the forwards transition rates ((45) and Proposition 12) and transition probabilities (Proposition 14) for this process.

Repeated application of the argument which justified (37) yields "known fact (i)" from the Introduction.

Lemma 10 (Lyons [29]) *The tree $\mathcal{G}_u$ is a GW tree whose offspring distribution $p_u(n) := P(Z\mathcal{G}_u = n)$ is determined for all $n = 0, 1, 2, \ldots$ by the formula*

$$g_u(s) := \sum_n p_u(n)s^n = g_1(1 - u + us) \qquad \forall\, u \in [0, 1]. \tag{43}$$

In particular, if $\mathcal{G}_1$ is a $PGW(\mu_1)$ tree then $\mathcal{G}_u$ is a $PGW(u\mu_1)$ tree.

Lyons [29, 30] and Haase [20] give applications of Lemma 10 to the theory of percolation on trees. Note the case $g_1(s) = s^k$ for some positive integer k, when $(\mathcal{G}_u, 0 \leq u \leq 1)$ is a uniform pruning of the deterministic tree in which each vertex has k children. Let

$$\mu_1 := E(Z\mathcal{G}_1) = \sum_n np_1(n). \tag{44}$$

Note that for all $0 \leq u \leq 1$

$$E(Z\mathcal{G}_u) = \sum_n np_u(n) = u\mu_1.$$

The forwards transition rates. Assuming $\mu_1 \leq 1$, so $\#\mathcal{G}_1 < \infty$ almost surely, we find from (3) and (41) that for $\mathbf{r}, \mathbf{s} \in \mathbf{T}$, for each $v \in \mathbf{r}$ and $j \in [c_v \mathbf{r} + 1]$ the forwards transition rate from $\mathbf{r}$ to $\mathbf{t} = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s})$ at time u given $\mathcal{G}_u = \mathbf{r}$ is

$$q_u(\mathbf{r} \to \mathbf{t}) = \frac{\#V(\mathbf{r}, \mathbf{t})\mu_1 \rho_u^{*-}(c_v \mathbf{r})}{c_v \mathbf{r} + 1} P(\mathcal{G}_u = \mathbf{s}) \quad \text{for each } \mathbf{t} = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s}) \quad (45)$$

where

$$\rho_u^{*-}(n) := \frac{p_u^{*-}(n)}{p_u(n)} \text{ for } p_u^{*-}(n) = (u\mu_1)^{-1}(n+1)p_u(n+1). \qquad (46)$$

Note that $p_u^{*-}(\cdot)$ is the distribution of $(Z\mathcal{G}_u)^* - 1$ for $(Z\mathcal{G}_u)^*$ with the size-biased distribution derived from the distribution $p_u(\cdot)$ of $Z\mathcal{G}_u$. The following lemma, which is the key to a later calculation, shows that $p_u^{*-}(\cdot)$ is also the distribution of $Z\mathcal{G}_u^{*-}$ where $(\mathcal{G}_u^{*-}, 0 \leq u \leq 1)$ is a uniform pruning of $\mathcal{G}_1^{*-}$ defined as a GW process with $\mathrm{dist}(Z\mathcal{G}_1^{*-}) = \mathrm{dist}((Z\mathcal{G}_1)^* - 1)$.

Lemma 11 *For a non-negative integer random variable Z with mean $\mu \in (0, \infty)$, let Z^* have the size-biased distribution $P(Z^* = n) = nP(Z = n)/\mu$, let $Z^{*-} = Z^* - 1$, and let Z_u denote the sum of Z independent Bernoulli(u) variables. Then for each $0 \leq u \leq 1$.*

$$(Z^{*-})_u \stackrel{d}{=} (Z_u)^{*-}. \qquad (47)$$

Proof. Let $g(s) := \sum_n P(Z = n)s^n$. Then Z^{*-} has generating function $g'(s)/\mu$ where $\mu = g'(1)$, and Z_u has generating function $g(1 - u + us)$ and mean $u\mu$. So (47) amounts to

$$\frac{g'(1 - u + us)}{\mu} = \frac{d}{ds}\frac{g(1 - u + us)}{u\mu}$$

which is true by the chain rule. □

In terms of operators on the set of probability measures on the non-negative integers with finite non-zero mean, the lemma states that the operator $\mathrm{dist}(Z) \to \mathrm{dist}(Z^{*-})$ commutes with the operator $\mathrm{dist}(Z) \to \mathrm{dist}(Z_u)$.

Formula (45) only displays transition rates between finite trees, assuming $\mu_1 \leq 1$. However, by comparison with (42), and by consideration of similar formulae for height restricted processes, we obtain the following intuitive description of the forwards evolution of $(\mathcal{G}_u)$ which is valid even without assuming $\mu_1 \leq 1$. Let $\mathrm{GW}(u)$ denote the distribution of the GW tree $\mathcal{G}_u$.

Proposition 12 *The distribution of the process* $(\mathcal{G}_u, 0 \le u \le 1)$ *with* $\mathcal{G}_0 = \bullet$ *is uniquely determined by the following transition rates: at each time* $u \in (0,1)$*, given* $\mathcal{G}_u = \mathbf{r}$*, each vertex* v *of* $\mathbf{r}$ *with* c *children runs a risk of appending a new branch at rate* $u^{-1}(c+1)p_u(c+1)/p_u(c)$ *where* $p_u(n) = P(Z\mathcal{G}_u = n)$ *is determined by (43); given that a branch is appended to* $v \in \mathcal{G}_{u-}$ *at time* u *the new branch is appended as if by random attachment of a branch with distribution* $GW(u)$*.*

Let $k = \sup\{n : p_1(n) > 0\}$. Provided $c < k+1$ it is elementary that $p_u(c)$ is a strictly positive and continuous function of $u \in (0,1)$, hence so is the rate $u^{-1}(c+1)p_u(c+1)/p_u(c)$ appearing above. So these rates are determined for all $0 \le c < k+1$ and $0 < u < 1$ as appropriate limits as $\epsilon \to 0$ of naively defined conditional probabilities from the joint distribution of $\mathcal{G}_u$ and $\mathcal{G}_{u+\epsilon}$.

As a corollary of Proposition 12 there is the following characterization of a PGW tree in terms of the evolution of its uniform pruning process. Part (i) of the corollary sharpens a similar description of uniform pruning for an unlabeled PGW tree in Aldous [1].

Corollary 13 *(i) If* $\mathcal{G}_1$ *is a* $PGW(\mu_1)$ *tree then* $\mathcal{G}_u$ *is a* $PGW(u\mu_1)$ *tree. The rate of attachment of branches to* $v \in \mathcal{G}_u$ *at time* u *is then identically equal to* μ_1*, for all* $0 < u < 1$ *and* $v \in \mathcal{G}_u$*, and given that at time* u *a branch is appended to* $v \in \mathcal{G}_{u-}$ *the new branch is appended as if by random attachment of a branch with distribution* $PGW(u\mu_1)$*.*

(ii) Conversely, if a uniform pruning $(\mathcal{G}_u)$ *of a GW tree* $\mathcal{G}_1$ *is such that for some* $0 < u < 1$*, and some vertex* v *such that* $P(v \in \mathcal{G}_u) > 0$*, the rate of attachment of branches to* v *given* $\mathcal{G}_u$ *does not depend on the number of children of* v *in* $\mathcal{G}_u$*, then* $\mathcal{G}_1$ *is a* $PGW(\mu_1)$ *tree for some* $0 \le \mu_1 < \infty$*.*

Proof. Part (i) is the specialization of Proposition 12 to a PGW tree. The assumption in (ii) forces $u^{-1}(c+1)p_u(c+1)/p_u(c) = g(u)$ for some $g(u)$ not depending on c. It follows that $p_u(\cdot)$ is Poisson$(u\mu_1)$ for some $\mu_1 < \infty$, and hence that $g_1(1-u+us) = g_u(s) = \exp(-u\mu_1(1-s))$. Since this determines $g_1(z) = \exp(-\mu_1(1-z))$ for $z \in (1-u, 1)$, and a probability generating function $g_1(z)$ is an analytic function of z for $|z| < 1$, it follows that $g_1(z) = \exp(-\mu_1(1-z))$ for all $|z| < 1$, and hence that dist$(Z\mathcal{G}_1) =$ Poisson(μ_1). □

Forwards transition probabilities By extension of the earlier notion of attaching the root of one family tree $\mathbf{s}$ as the jth child of some vertex v of another family tree $\mathbf{r}$ to form a new tree $\mathbf{t}(\mathbf{r}, v, j, \mathbf{s})$, we can make sense of attaching various trees as variously ranked children of various vertices of $\mathbf{r}$. Given non-negative integers $k(v), v \in \mathbf{r}$ and for each $v \in \mathbf{r}$ an increasing sequence of $k(v)$ integers

$$1 \leq j_1(v) < \ldots < j_{k(v)}(v) \leq c_v\mathbf{r} + k(v) \tag{48}$$

and a sequence of $k(v)$ trees $\mathbf{s}_1(v), \ldots, \mathbf{s}_{k(v)}(v)$, we can construct a new family tree by *attaching the root of* $\mathbf{s}_i(v)$ *to* $\mathbf{r}$ *as the* $j_i(v)$*th child of* v *for each* $i \in [k(v)]$ *and each* $v \in \mathbf{r}$. Given $\mathbf{r}$ and $k(v), v \in \mathbf{r}$, and some distribution for a random family tree $\mathcal{S}$, say that a random tree $\mathcal{T}$ is *constructed by random attachment of* $k(v)$ *independent copies of* $\mathcal{S}$ *to* v *for each* $v \in \mathbf{r}$ if the distribution of $\mathcal{T}$ is that induced by making the above construction with a uniform random choice of $j_i(v), i \in [k(v)]$ subject to (48) and random choice of $\mathbf{s}_i(v)$ according to the distribution of $\mathcal{S}$, independently for each (v, i) with $v \in \mathbf{r}, i \in [k(v)]$. Assuming for simplicity that $\mathbf{r}$ is a finite tree and $\mathcal{S}$ is an almost surely finite tree, the probability of making the construction with any particular choice of $j_i(v)$ and $\mathbf{s}_i(v)$ is then

$$\prod_{v \in \mathbf{r}} \binom{c_v\mathbf{r} + k(v)}{k(v)}^{-1} \prod_{i=1}^{k(v)} P(\mathcal{S} = \mathbf{s}_i(v)). \tag{49}$$

As for random attachment of one copy of $\mathcal{S}$ to one vertex v of $\mathbf{r}$ discussed earlier, various choices of $j_i(v)$ and $\mathbf{s}_i(v)$ may result in construction of the same family tree $\mathbf{t}$. So for $\mathcal{T}$ constructed by random attachment of $k(v)$ independent copies of $\mathcal{S}$ to v for each $v \in \mathbf{r}$, the probability $P(\mathcal{T} = \mathbf{t})$ is obtained by summing the expression (49) over all such choices.

The following proposition describes the joint distribution of $\mathcal{G}_u$ and $\mathcal{G}_1$ for each fixed $0 \leq u < 1$ in terms of random attachment of branches. The joint distribution of $\mathcal{G}_u$ and $\mathcal{G}_z$ for arbitrary $0 \leq u < z \leq 1$ is then determined by rescaling. For $c \geq 0$ let $\bar{P}_u(c, \cdot)$ denote the conditional distribution of $Z\mathcal{G}_1 - Z\mathcal{G}_u$ given $Z\mathcal{G}_u = c$. That is, for $m \geq 0$

$$\bar{P}_u(c, m) = P(Z\mathcal{G}_1 - Z\mathcal{G}_u = m \,|\, Z\mathcal{G}_u = c) = \frac{p_1(m+c)}{p_u(c)} \binom{m+c}{c} u^c (1-u)^m. \tag{50}$$

It is known [40] that $\bar{P}_u(c,\cdot)$ does not depend on c, meaning that $Z\mathcal{G}_1 - Z\mathcal{G}_u$ and $Z\mathcal{G}_u$ are independent, iff $p_1(\cdot)$ is Poisson.

Proposition 14 *Fix $0 \le u < 1$. Given $\mathcal{G}_u$, let $K_u(v), v \in \mathrm{verts}(\mathcal{G}_u)$ be independent with distributions $\bar{P}_u(c_v\mathcal{G}_u,\cdot)$ and given the $K_u(v)$ for all $v \in \mathcal{G}_u$ let $\hat{\mathcal{G}}_1$ be defined by random attachment of $K_u(v)$ independent copies of $\mathcal{G}_1$ to v for each $v \in \mathcal{G}_u$. Then $(\mathcal{G}_u, \mathcal{G}_1) \stackrel{d}{=} (\mathcal{G}_u, \hat{\mathcal{G}}_1)$.*

Proof. ¿From the uniform pruning construction, $(\mathcal{G}_u, \mathcal{G}_1) \stackrel{d}{=} (\mathrm{fam}(\mathcal{G}_u^\dagger), \mathcal{G}_1)$ where $\mathcal{G}_u^\dagger$ is the subtree of $\mathcal{G}_1$ containing the root after removing each edge of $\mathcal{G}_1$ independently with probability $1-u$. It is easily seen that conditionally given $\mathcal{G}_u^\dagger$, independently as v ranges over $\mathcal{G}_u^\dagger$ the number of children w of v in $\mathcal{G}_1$ that are not children of v in $\mathcal{G}_u$ has distribution $\bar{P}_u(c_v\mathcal{G}_u,\cdot)$. Moreover, each one of these children $w \in \mathcal{G}_1 - \mathcal{G}_u^\dagger$ is the root of a subtree of $\mathcal{G}_1$ which when identified as a family tree is an independent copy of $\mathcal{G}_1$. The identity $(\mathcal{G}_u, \mathcal{G}_1) \stackrel{d}{=} (\mathcal{G}_u, \hat{\mathcal{G}}_1)$ now follows after appropriate relabeling, using the two following consequences of the uniform pruning construction:
(i) for each $v \in \mathcal{V}$, given $v \in \mathcal{G}_u^\dagger$ and both the number c of children of v in $\mathcal{G}_u^\dagger$ and the number $c+k$ of children of v in $\mathcal{G}_1$, the set of c ranks of the children in $\mathcal{G}_u^\dagger$ is a uniform random subset of $[c+k]$ of size c;
(ii) for each $h \ge 0$, conditionally given $r_h\mathcal{G}_1$, $r_h\mathcal{G}_u^\dagger$, and the $c(v)$ and $c(v)+k(v)$ as v ranges over vertices of $\mathcal{G}_u^\dagger$ of height h, these random subsets of sizes $c(v)$ picked from $[c(v)+k(v)]$ are independent.

□

3.3 Pruning a GW tree conditioned on non-extinction

In this section, let $\mathcal{G}_1$ be a GW tree which is critical or subcritical, with non-degenerate offspring distribution, so

$$\mu_1 := E(Z\mathcal{G}_1) \in (0,1] \text{ and } \#\mathcal{G}_1 < \infty \text{ almost surely.}$$

As in the previous section, let

$$(\mathcal{G}_u, 0 \le u \le 1) \text{ be a uniform pruning of } \mathcal{G}_1.$$

Let $\mathcal{G}_1^\infty$ be $\mathcal{G}_1$ conditioned on non-extinction, as in Proposition 2, and let

$$(\mathcal{G}_u^*, 0 \le u \le 1) \text{ be a uniform pruning of } \mathcal{G}_1^\infty.$$

Intuitively,

$$(\mathcal{G}^*_u, 0 \le u \le 1) \text{ is } (\mathcal{G}_u, 0 \le u \le 1) \text{ conditioned on } \#\mathcal{G}_1 = \infty \tag{51}$$

but as indicated around (31), this interpretation is hazardous for $\mu < 1$. Our descriptions of $(\mathcal{G}^*_u, 0 \le \mu \le 1)$ parallel similar descriptions, in terms of weak limits or h-transforms, of bridges and excursions of Markov processes such as Brownian motion [17, 41].

To be careful about (51), it follows from Proposition 2 and Lemma 6 that

$$\operatorname{dist}(\mathcal{G}^*_u, 0 \le u \le 1) = \lim_{h \uparrow \infty} \operatorname{dist}(\mathcal{G}_u, 0 \le u \le 1 \,|\, Z_h \mathcal{G}_1 > 0) \tag{52}$$

and from Proposition 5 and Lemma 6 that *provided* $\mu_1 = 1$

$$\operatorname{dist}(\mathcal{G}^*_u, 0 \le u \le 1) = \lim_{n \uparrow \infty} \operatorname{dist}(\mathcal{G}_u, 0 \le u \le 1 \,|\, \#\mathcal{G}_1 = n). \tag{53}$$

Here (52) and (53) refer to convergence of finite dimensional distributions, which extends easily to convergence of distributions on suitable path spaces. Note that (53) is false for $\mu_1 < 1$. With the conditions and notation of Proposition 5, the limit in distribution on the right side of (53) is rather the distribution of the uniform pruning $(\bar{\mathcal{G}}^*_u, 0 \le u \le 1)$ of $\bar{\mathcal{G}}^\infty_1$ obtained by conditioning $\bar{\mathcal{G}}$ on non-extinction, where $\bar{\mathcal{G}}$ is the critical GW tree such that

$$\operatorname{dist}(\bar{\mathcal{G}} \,|\, \#\bar{\mathcal{G}} = n) = \operatorname{dist}(\mathcal{G} \,|\, \#\mathcal{G} = n) \quad \forall\, n \ge 1.$$

Since $\mathcal{G}^\infty_1$ is a tree with only one infinite path, it is obvious that $\#\mathcal{G}^*_u < \infty$ almost surely for all $u < 1$. And, from (34),

$$\mathcal{G}^*_{1-} = \mathcal{G}^*_1 = \mathcal{G}^\infty_1 \text{ almost surely.} \tag{54}$$

The following proposition provides an explicit formula for the distribution of $\mathcal{G}^*_u$ via its density relative to the distribution of $\mathcal{G}_u$. Recall that the distribution of $\mathcal{G}_u$ is given by the product formula (3) with $p_u(\cdot)$ in place of $p(\cdot)$.

Proposition 15 *The distribution of the uniform pruning process $(\mathcal{G}^*_u, 0 \le u < 1)$ with countable state space $\mathbf{T}$ is determined by the co-transition probabilities (36) of any uniform pruning process and the following distribution of $\mathcal{G}^*_u$ for each $0 \le u < 1$:*

$$P(\mathcal{G}^*_u = \mathbf{t}) = h^*(u, \mathbf{t}) P(\mathcal{G}_u = \mathbf{t}) \qquad \forall\, \mathbf{t} \in \mathbf{T} \tag{55}$$

where $h^*(0,\bullet) = 1$ *and for* $0 < u < 1$

$$h^*(u,\mathbf{t}) := (1-u)\sum_{v\in\mathbf{t}} \mu_1^{-h(v)} \rho_u^{*-}(c_v\mathbf{t}) \tag{56}$$

with $h(v)$ *the height of* v*, and* $c_v\mathbf{t}$ *the number of children of* v *in* $\mathbf{t}$*, and* $\rho_u^{*-}(n) := (n+1)p_u(n+1)(u\mu_1 p_u(n))^{-1}$ *for* $p_u(\cdot)$ *the offspring distribution of* $\mathcal{G}_u$ *as in (43).*

After proving this proposition, we point out some reformulations of it.
Proof. It suffices to derive (55) for $(\mathcal{G}_u^*)$ constructed as $\mathcal{G}_u^* = \mathrm{fam}(\mathcal{G}_u^\dagger)$ where $\mathcal{G}_u^\dagger$ is the component containing the root in the subgraph of $\mathcal{G}_1^\infty$ consisting of those edges e with $\xi_e \le u$ where the ξ_e are independent uniform$(0,1)$ random variables. Let (V_h) be the infinite spine of $\mathcal{G}_1^\infty$, let

$$H_u := \sup\{h : V_h \in \mathcal{G}_u^\dagger\}$$

that is the height of the highest spinal vertex of $\mathcal{G}_1^\infty$ that is a vertex of $\mathcal{G}_u^\dagger$. Formula (55) follows from formula (57) of the next lemma by summation over $v \in \mathbf{t}$. □

Lemma 16 *For* $0 \le u < 1$ *let* V_u^* *be the vertex of* $\mathcal{G}_u^*$ *at height* H_u *which is the image of* $V_{H_u} \in \mathcal{G}_u^\dagger$ *via the relabeling map from* $\mathcal{G}_u^\dagger \to \mathcal{G}_u^*$*. Then for all* $\mathbf{t} \in \mathbf{T}$ *and* $v \in \mathbf{t}$*,*

$$P(\mathcal{G}_u^* = \mathbf{t}, V_u^* = v) = (1-u)\mu_1^{-h(v)} \rho_u^{*-}(c_v\mathbf{t})\, P(\mathcal{G}_u = \mathbf{t}). \tag{57}$$

Proof. The next lemma allows this formula to be read from (23) by application of Lemma 4 with $\mathcal{G} = \mathcal{G}_u$, $\mathcal{G}^\infty = \mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ for $\mathcal{G}_u^{\infty\dagger}$ as defined below, and $H = H_u$. □

Lemma 17 *Fix* $0 < u < 1$*. Let* $\mathcal{G}_u^{\infty\dagger}$ *be the subtree of* $\mathcal{G}_1^\infty$ *which is the component containing 0 in the random graph defined by deletion of each edge* e *of* $\mathcal{G}_1^\infty$ *not in its infinite spine and such that* $\xi_e > u$*, where the* ξ_e *are the independent uniform* $(0,1)$ *variables used to construct* $\mathcal{G}_u^*$*. Then*

(i) The tree $\mathcal{G}_u^*$ *is the family tree derived from the finite subtree of* $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ *which remains after cutting the spine of* $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ *between heights* H_u *and* $H_u + 1$*.*

(ii) Let $\mathcal{G}_u^\infty$ be $\mathcal{G}_u$ conditioned on non-extinction. Then

$$\mathrm{fam}(\mathcal{G}_u^{\infty\dagger}) \stackrel{d}{=} \mathcal{G}_u^\infty.$$

(iii) The height H_u is independent of $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ *with the geometric*$(1-u)$ *distribution:*

$$P(H_u = n) = (1-u)u^n \qquad \forall\ n = 0, 1, \ldots$$

Proof. Part (i) is evident from the definitions, and (iii) is obvious because H_u is the least n such that the edge $e := (V_n, V_{n+1})$ has $\xi_e > u$. To prove (ii), observe that from the description of $\mathcal{G}_1^\infty$ in Proposition (2) (iv), the definition of $\mathcal{G}_u^{\infty\dagger}$, and (43), at each level h of $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$, the vertices of $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ have independent numbers of offspring with the offspring distribution $p_u(\cdot)$ of $\mathcal{G}_u$ for non-spinal vertices, and a modified offspring distribution for spinal vertices. According to Proposition (2) (iv) applied to $\mathcal{G} = \mathcal{G}_u$, a similar statement applies to $\mathcal{G}_u^\infty$. So to show $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger}) \stackrel{d}{=} \mathcal{G}_u^\infty$, it suffices to check that

(a) the offspring distribution of each spinal vertex of $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ is identical to the offspring distribution of each spinal vertex of $\mathcal{G}_u^\infty$;

(b) given that a spinal vertex of $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ has c children, the rank of the next spinal vertex of $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ is uniform on $[c]$.

But, in the notation of Lemma 11, by Proposition 2(iv) applied to $\mathcal{G}_1^\infty$ and the construction of $\mathcal{G}_u^{\infty\dagger}$, each spinal vertex of $\mathrm{fam}(\mathcal{G}_u^{\infty\dagger})$ has offspring according to the distribution of $(Z\mathcal{G}_1^\infty - 1)_u + 1 = ((Z\mathcal{G}_1)^{*-})_u + 1$. On the other hand, by Proposition 2 (iv) applied to $\mathcal{G}_u^\infty$, each spinal vertex of $\mathcal{G}_u^\infty$ has offspring according to the distribution of $(Z\mathcal{G}_u)^* = (Z\mathcal{G}_u)^{*-} + 1$. But

$$(Z\mathcal{G}_u)^{*-} + 1 \stackrel{d}{=} ((Z\mathcal{G}_1)_u)^{*-} + 1 \stackrel{d}{=} ((Z\mathcal{G}_1)^{*-})_u + 1 \tag{58}$$

where the first equality is due to (43) and the second is the commutation rule of Lemma 11. This proves (a), and (b) is quite easy. $\square$

Reformulation of Proposition 15. Since $(\mathcal{G}_u, 0 \le u < 1)$ and $(\mathcal{G}_u^*, 0 \le u < 1)$ are two uniform pruning processes with the same co-transition probabilities (36), for all $\mathbf{t} \in \mathbf{T}$ and $0 < u < 1$

$$\mathrm{dist}(\mathcal{G}_t^*, 0 \le t \le u \,|\, \mathcal{G}_u^* = \mathbf{t}) = \mathrm{dist}(\mathcal{G}_t, 0 \le t \le u \,|\, \mathcal{G}_u = \mathbf{t}). \tag{59}$$

We interpret this as an equality of distributions on the path space $\Omega[0,u]$, where for $u > 0$ we let $\Omega[0,u]$ be the space of all right continuous step function paths $\omega : [0,u] \to \mathbf{T}$ with at most a finite number of jumps, equipped with the σ-field generated by the maps $t \to \omega_t$ where $\omega = (\omega_t, 0 \le t \le u)$. Then (59) combined with (55) amounts to the following formula: for each non-negative measurable function f defined on $\Omega[0,u]$,

$$E[f(\mathcal{G}_t^*, 0 \le t \le u)] = E[h(u,\mathcal{G}_u) f(\mathcal{G}_t, 0 \le t \le u)]. \tag{60}$$

Implicit in Proposition 15 is the consistency of this prescription of distributions of $(\mathcal{G}_t^*, 0 \le t \le u)$ as u varies. By a standard argument, this consistency amounts to:

Corollary 18 *Relative to the filtration generated by* $(\mathcal{G}_u, 0 \le u < 1)$,

the process $(h^*(u,\mathcal{G}_u), 0 \le u < 1)$ *is a non-negative martingale, with*

$$Eh^*(u,\mathcal{G}_u) = 1 \qquad \forall\ 0 \le u < 1.$$

In other words, $h^*(u,\mathbf{t})$ defined by (56) is a space-time harmonic function for the chain $(\mathcal{G}_u, 0 \le u < 1)$, and the Doob h^*-transform of $(\mathcal{G}_u, 0 \le u < 1)$ is $(\mathcal{G}_u^*, 0 \le u < 1)$. See the end of Section 4.4 for identification of the corresponding Martin boundary.

3.4 The supercritical case

In the critical case, the GW tree conditioned to be infinite (Proposition 2) has another interpretation as a limit of supercritical GW trees conditioned on non-extinction. This fact, and its consequence for pruning processes, are spelled out in Proposition 19.

It is convenient here to rescale the time parameter to make it identical to the offspring mean. So suppose $(\mathcal{G}_\mu, \mu \in I)$ is a uniform pruning process parameterized by an interval I with $[0,1] \subset I \subseteq [0,\infty)$, such that $\mathcal{G}_\mu$ is a GW tree with $E(Z\mathcal{G}_\mu) = \mu$ for all $\mu \in I$. For example, take $I = [0,k]$ for some $k = 2,3,\ldots$, and let $(\mathcal{G}_{uk}, 0 \le u \le 1)$ be a uniform pruning of the deterministic regular tree in which each vertex has k children. Or, as in the next section, take $\mathcal{G}_\mu$ to be a PGW(μ) tree.

Proposition 19

$$\lim_{\mu\downarrow 1} \mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = \infty) = \lim_{\mu\downarrow 1} \mathrm{dist}(\mathcal{G}_1 \,|\, \#\mathcal{G}_\mu = \infty) = \mathrm{dist}(\mathcal{G}_1^\infty) \tag{61}$$

$$\lim_{\mu\downarrow 1} \mathrm{dist}(\mathcal{G}_u, 0 \le u \le 1 \,|\, \#\mathcal{G}_\mu = \infty) = \mathrm{dist}(\mathcal{G}_u^*, 0 \le u \le 1) \tag{62}$$

Proof. By application of Lemma 6, it suffices to show that

$$\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = \infty) \to \mathrm{dist}(\mathcal{G}_1^\infty) \quad \text{as } \mu \downarrow 1. \tag{63}$$

Let

$$F(\mu) := P(\#\mathcal{G}_\mu < \infty); \qquad \bar{F}(\mu) := 1 - F(\mu) = P(\#\mathcal{G}_\mu = \infty). \tag{64}$$

Let g_μ be the generating function of $Z\mathcal{G}_\mu$. It is well known that the *extinction probability* $F(\mu)$ is the least non-negative root of the equation

$$F(\mu) = g_\mu(F(\mu)) \tag{65}$$

and that $\bar{F}(\mu)$ is strictly positive iff $\mu > 1$. Fix $\nu > 1$ with $\nu \in I$. It is assumed that $(\mathcal{G}_{u\nu}, 0 \le u \le 1)$ is a uniform pruning of $\mathcal{G}_\nu$, and hence

$$g_\mu(s) = g_\nu(1 - (\mu/\nu) + (\mu/\nu)s) \tag{66}$$

It is easy to see from this formula and the strict convexity of g_ν that $F(\mu)$ is a continuous decreasing function of μ. Hence $\bar{F}(\mu) \downarrow 0$ as $\mu \downarrow 1$, Now fix $h \ge 0$ and $\mathbf{t} \in \mathbf{T}$ and compute

$$\begin{aligned} P(r_h\mathcal{G}_\mu = \mathbf{t} \,|\, \#\mathcal{G}_\mu = \infty) &= P(r_h\mathcal{G}_\mu = \mathbf{t})\, \frac{1 - (1 - \bar{F}(\mu))^{Z_h\mathbf{t}}}{\bar{F}(\mu)} \\ &\to P(r_h(\mathcal{G}_1) = \mathbf{t})\,(Z_h\mathbf{t}) = P(r_h\mathcal{G}_1^\infty = \mathbf{t}) \text{ as } \mu \downarrow 1 \end{aligned}$$

where the continuity of $P(r_h(\mathcal{G}_\mu) = \mathbf{t})$ is evident from (2) and the explicit formula for $p_\mu(\cdot)$ in terms of $p_\nu(\cdot)$ which can be read from (66), and the last equality is (18) for $\mu = 1$. □

The ascension time. In terms of the uniform pruning process $(\mathcal{G}_\mu, \mu \in I)$ we define the *ascension time* $A := \inf\{\mu : \#\mathcal{G}_\mu = \infty\}$. So $A > 1$ a.s. The events $(A \leq \mu)$ and $(\#\mathcal{G}_\mu = \infty)$ are identical, so

$$P(A \leq \mu) = P(\#\mathcal{G}_\mu = \infty) = \bar{F}(\mu) \qquad \forall\ \mu \in I \tag{67}$$

and we interpret (62) intuitively as

$$\mathrm{dist}(\mathcal{G}_u, 0 \leq u \leq 1 \,|\, A = 1) = \mathrm{dist}(\mathcal{G}^*_u, 0 \leq u \leq 1). \tag{68}$$

In contrast, Proposition 12 implies that $\#\mathcal{G}_{A-} < \infty$ a.s. and implies an explicit formula for the joint law of $\mathcal{G}_{A-}$ and A, which can be used to obtain a continuously varying conditional distribution of $\mathcal{G}_{A-}$ given $A = a$ for $a > 1$. But this distribution is concentrated on finite trees rather than on infinite ones. In Section 4.2 we study the Poisson case where there is much simplification.

4 The PGW pruning process

It follows easily from Lemma 10 by Kolmogorov's extension theorem that there exists a unique distribution for a $\mathbf{T}^{(\infty)}$-valued inhomogeneous Markov process $(\mathcal{G}_\mu, 0 \leq \mu < \infty)$ such that

$$\mathrm{dist}(\mathcal{G}_\mu) = \mathrm{PGW}(\mu) \quad \forall\ 0 \leq \mu < \infty. \tag{69}$$

and $(\mathcal{G}_{u\mu}, 0 \leq u \leq 1)$ is a uniform pruning of $\mathcal{G}_\mu$ for each $\mu > 0$. Essentially the same Markov process, with states identified as rooted unlabeled trees rather than family trees, featured in Aldous [1]. We shall show how the results of section 3 may be simplified and extended in this PGW setting. We sometimes give both "proof by specialization" from the general GW result in section 3, and an "autonomous proof" directly exploiting Poisson structure.

4.1 The joint law of $(\mathcal{G}_\lambda, \mathcal{G}_\mu)$

For each fixed pair of times (λ, μ) with $0 \leq \lambda < \mu < \infty$ this joint law is determined by the $\mathrm{PGW}(\mu)$ distribution of $\mathcal{G}_\mu$ and the conditional distribution of $\mathcal{G}_\lambda$ given $\mathcal{G}_\mu$, as given by (36). The following specialization of Proposition 14 describes the conditional distribution of $\mathcal{G}_\mu$ given $\mathcal{G}_\lambda$:

Proposition 20 *Fix λ and μ with $0 \leq \lambda < \mu < \infty$. Given $\mathcal{G}_\lambda$, let $N_{\lambda,\mu}(v), v \in \mathcal{G}_\lambda$ be independent with Poisson($\mu - \lambda$) distribution, and given the $N_{\lambda,\mu}(v), v \in \mathcal{G}_\lambda$ let $\hat{\mathcal{G}}_\mu$ be defined by random attachment of $N_{\lambda,\mu}(v)$ independent copies of $\mathcal{G}_\mu$ to v for each $v \in \mathcal{G}_\lambda$. Then $(\mathcal{G}_\lambda, \mathcal{G}_\mu) \stackrel{d}{=} (\mathcal{G}_\lambda, \hat{\mathcal{G}}_\mu)$. In particular*

$$(\mathcal{G}_\lambda, \#\mathcal{G}_\mu) \stackrel{d}{=} \left(\mathcal{G}_\lambda, \#\mathcal{G}_\lambda + \sum_{i=1}^{N_{\lambda,\mu}} \#\mathcal{G}_\mu(i)\right) \tag{70}$$

where $N_{\lambda,\mu}$ is the sum of $N_{\lambda,\mu}(v)$ over all vertices v of $\mathcal{G}_\lambda$, so given $\mathcal{G}_\lambda$ with $\#\mathcal{G}_\lambda = \ell$ the distribution of $N_{\lambda,\mu}$ is Poisson with mean $\ell(\mu - \lambda)$, and given $\mathcal{G}_\lambda$ and $N_{\lambda,\mu}$ the $\mathcal{G}_\mu(i)$ are i.i.d. copies of $\mathcal{G}_\mu$.

For general offspring distribution, the process $(\#\mathcal{G}_u)$ is not necessarily Markov, but in the Poisson setting it is. We suspect that either of the results of the following corollary can be used to characterize the Poisson case.

Corollary 21 *The process $(\#\mathcal{G}_\mu, \mu \geq 0)$ with state space $\{1, 2, \ldots, \infty\}$ has the Markov property, and*

$$\textit{the process } ((1-\mu)\#\mathcal{G}_\mu, 0 \leq \mu < 1) \textit{ is a martingale} \tag{71}$$

relative to the filtration generated by $(\mathcal{G}_\mu, 0 \leq \mu < 1)$.

Proof by specialization. The forwards Markov property of $(\#\mathcal{G}_\mu, \mu \geq 0)$ follows easily from (70) and the Markov property of $(\mathcal{G}_\mu, \mu \geq 0)$. The martingale property is the particular case of Corollary 18 for $\mathcal{G}_1$ a PGW(1) tree.

Autonomous proof. Result (71) can be checked directly from (70) as follows. Since

$$E(\#\mathcal{G}_\mu) = E\left(\sum_{h=0}^{\infty} Z_h \mathcal{G}_\mu\right) = \sum_{h=0}^{\infty} \mu^h = (1-\mu)^{-1} \tag{72}$$

formula (70) implies that for $0 \leq \lambda \leq \mu < 1$

$$E(\#\mathcal{G}_\mu \,|\, \mathcal{G}_\lambda) = \#\mathcal{G}_\lambda + \#\mathcal{G}_\lambda(\mu - \lambda)(1-\mu)^{-1} = \left(\frac{1-\lambda}{1-\mu}\right)\#\mathcal{G}_\lambda.$$

With the Markov property of $(\mathcal{G}_\mu, 0 \leq \mu < 1)$ this gives

$$E((1-\mu)\#\mathcal{G}_\mu \,|\, \mathcal{G}_\eta, 0 \leq \eta \leq \lambda) = E((1-\mu)\#\mathcal{G}_\mu \,|\, \mathcal{G}_\lambda) = (1-\lambda)\#\mathcal{G}_\lambda$$

which is (71). $\square$

Formulae for the forwards transition probabilities of the Markov chain $(\#\mathcal{G}_\mu, \mu \geq 0)$ can be obtained as follows from the representation (70). Consider for $\theta > 0$ and $0 \leq \mu \leq 1$, the distribution of

$$X_{\theta,\mu} := \sum_{i=1}^{N_\theta} X_{\mu,i}$$

where N_θ has Poisson (θ) distribution, and given $N_\theta = n$ the $X_{\mu,i}$ for $1 \leq i \leq n$ are i.i.d. with the Borel(μ) distribution $P_\mu(\cdot)$. The distribution of $X_{\theta,\mu}$ is known [13] as the *generalized Poisson distribution (GPD) on* $\{0, 1, \ldots, \infty\}$ *with parameters* (θ, μ), and given by the formula

$$P(X_{\theta,\mu} = k) = \frac{1}{k!}\theta(\theta + k\mu)^{k-1}\, e^{-\theta - k\mu} \qquad \forall k = 0, 1, \ldots \tag{73}$$

which follows easily from (11). According to (70), for $0 \leq \lambda \leq \mu < \infty$ the conditional distribution of $\#\mathcal{G}_\mu - \#\mathcal{G}_\lambda$ given $\#\mathcal{G}_\lambda = \ell$ is GPD(θ, μ) for $\theta = \ell(\mu - \lambda)$. That is to say

$$P(\#\mathcal{G}_\mu = \ell + k | \#\mathcal{G}_\lambda = \ell) = P(X_{\theta,\lambda} = k) \tag{74}$$

as in (73) for $\theta = \ell(\mu - \lambda)$. Take $k = m - \ell$ in (74) to obtain the following formula for $0 \leq \lambda < \mu$ and $1 \leq \ell \leq m < \infty$:

$$P(\#\mathcal{G}_\mu = m \,|\, \#\mathcal{G}_\lambda = \ell) = \frac{\ell(\mu - \lambda)}{(m - \ell)!}(m\mu - \ell\lambda)^{m-\ell-1} e^{-(m\mu - \ell\lambda)}. \tag{75}$$

Using Bayes' rule and the Borel distributions (10) of $\#\mathcal{G}_\lambda$ and $\#\mathcal{G}_\mu$, this formula can be inverted to obtain an expression for the co-transition probability $P(\#\mathcal{G}_\lambda = \ell \,|\, \#\mathcal{G}_\mu = m)$. Due to the $\mathcal{U}_n$-representation (15) of dist$(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n)$ and the scaling property of a uniform pruning process, this co-transition probability depends on (λ, μ) only through the ratio λ/μ, and has a simple combinatorial interpretation in terms of pruning a uniform random tree. See Section 4.8 for further discussion and references to earlier appearances of the same co-transition probabilities.

4.2 Transition rates and the ascension process

For each $h = 1, 2, \ldots$ the height restricted process $(r_h \mathcal{G}_\mu, 0 \leq \mu < \infty)$ is an inhomogeneous Markov chain with countable state space $\mathbf{T}^{(h)}$ and step-function paths, whose co-transition rates and co-transition probabilities can be read from the general formulae of section 3.

The process $(\mathcal{G}_\mu, 0 \leq \mu < \infty)$ develops with time running forwards by a process of attachment of trees which can be described as follows, due to Corollary 13. Starting from $\mathcal{G}_0 = \bullet$, the single-vertex tree, at each time $\mu \geq 0$ and at each vertex v of $\mathcal{G}_\mu$, attachments to v are made at rate 1; given $\mathcal{G}_\mu = \mathbf{r}$ and that an attachment is made to $v \in \mathbf{r}$ at time μ, the tree to be attached has PGW(μ) distribution, and the root of this tree is attached to $\mathbf{r}$ as the jth child of v for a j chosen independently and uniformly from $[c_v\mathbf{r} + 1]$. Thus for each $v \in \mathbf{r}$ and $j \in [c_v\mathbf{r} + 1]$ the forwards transition rate from $\mathbf{r}$ to $\mathbf{t} = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s})$ at time u given $\mathcal{G}_\mu = \mathbf{r}$ is

$$q_\mu(\mathbf{r} \to \mathbf{t}) = \frac{\# V(\mathbf{r}, \mathbf{t}) P(\mathcal{G}_\mu = \mathbf{s})}{c_v\mathbf{r} + 1} \quad \text{for each } \mathbf{t} = \mathbf{t}(\mathbf{r}, v, j, \mathbf{s}). \tag{76}$$

The rates (76) for $\mathbf{r}, \mathbf{t} \in \mathbf{T}$ determine the evolution of $(\mathcal{G}_\mu, \mu \geq 0)$ only up to the *ascension time* $A := \inf\{\lambda : \#\mathcal{G}_\lambda = \infty\}$. As noted in Aldous [1], what happens at time A is that the process attaches an infinite branch to some vertex of $\mathcal{G}_{A-}$, which is an almost surely finite random tree, to form an infinite tree $\mathcal{G}_A$.

Consider now the *ascension process* $(\mathcal{G}_\mu, 0 \leq \mu < A)$ in which the state at time μ is $\mathcal{G}_\mu$ if $0 \leq \mu < A$ and $\boldsymbol{\infty}$ if $A \leq \mu$, where $\boldsymbol{\infty}$ is a state representing any infinite tree. The ascension process is a Markov chain with countable state-space $\mathbf{T} \cup \{\boldsymbol{\infty}\}$, where $\boldsymbol{\infty}$ is an absorbing state. The distribution of the ascension process is determined by the initial state $\mathcal{G}_0 = \bullet$, the transition rates (76), and the *ascension rate*

$$q_\mu(\mathbf{r} \to \boldsymbol{\infty}) = (\#\mathbf{r}) P(\#\mathcal{G}_\mu = \infty) =: (\#\mathbf{r}) \bar{F}(\mu). \tag{77}$$

Note that the total rate of transitions out of each state $\mathbf{r} \in \mathbf{T}$ is $\#\mathbf{r}$, which is the sum of the combined rate $(\#\mathbf{r})F(\mu)$ of transitions to all other finite trees, and the rate $(\#\mathbf{r})\bar{F}(\mu)$ for transitions to $\boldsymbol{\infty}$, where $F(\mu) + \bar{F}(\mu) = 1$.

We recall now some known features of the function $\bar{F} : [1, \infty) \to (0, 1]$ and the conditional distribution of $\mathcal{G}_\mu$ given that $\mathcal{G}_\mu < \infty$. See Alon-Spencer

[8, §6.4 and §6.5], or Aldous [1] for further discussion. For the Poisson(μ) generating function $g_\mu(s) = \exp(-\mu(1-s))$ the equation (65) shows that for $\mu > 1$ the non-extinction probability $\bar{F}(\mu)$ is the strictly positive solution of $1 - \bar{F}(\mu) = \exp(-\mu\bar{F}(\mu))$. It follows that the inverse function $\bar{F}^{-1} : (0,1] \to [1,\infty)$ is

$$\bar{F}^{-1}(u) = \frac{-\log(1-u)}{u} = 1 + \frac{u}{2} + \frac{u^2}{3} + \cdots \tag{78}$$

For $\mu \geq 1$ define the *conjugate* $\hat{\mu} \leq 1$ by $\hat{\mu} = \mu F(\mu)$, where $F(\mu) = 1 - \bar{F}(\mu)$ is the extinction probability. Then for $\mu \geq 1$ it follows from (8) that

$$P(\mathcal{G}_\mu = \mathbf{t}) = F(\mu)P(\mathcal{G}_{\hat{\mu}} = \mathbf{t}) \qquad \forall\, \mathbf{t} \in \mathbf{T} \tag{79}$$

and hence for $\mu \geq 1$

$$\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu < \infty) = \mathrm{dist}(\mathcal{G}_{\hat{\mu}}) = \mathrm{PGW}(\hat{\mu}). \tag{80}$$

The formula (77) for the ascension rate combined with these results gives the following formulae for the joint distribution of A and $\mathcal{G}_{A-}$. Formula (81) and the variant of (82) for an unlabeled $(\mathcal{G}_\mu)$ appear as formulae (13) and (15) in [1].

Lemma 22 *For $\mu \geq 1$ and $t \in \mathbf{T}$,*

$$\begin{aligned} P(A \leq \mu) &= \bar{F}(\mu) && (81)\\ P(\mathcal{G}_{A-} = \mathbf{t}, A \in d\mu) &= \bar{F}(\mu)(\#\mathbf{t})P(\mathcal{G}_\mu = \mathbf{t})\, d\mu && (82)\\ P(\mathcal{G}_{A-} = \mathbf{t} \,|\, A = \mu) &= (1-\hat{\mu})(\#\mathbf{t})P(\mathcal{G}_{\hat{\mu}} = \mathbf{t}) && (83) \end{aligned}$$

for $\bar{F}(\mu)$ and the conjugate $\hat{\mu}$ as defined above. Let U have uniform$(0,1)$ distribution. Then

$$\left(A, \frac{\hat{A}}{A}\right) := (A, F(A)) \stackrel{d}{=} \left(\frac{-\log U}{1-U}, U\right). \tag{84}$$

Proof. First, $P(A \leq \mu) = P(\#\mathcal{G}_\mu = \infty) = \bar{F}(\mu)$, giving (81). Next, given $\mathcal{G}_\mu = \mathbf{t}$, each of the $\#\mathbf{t}$ vertices has chance $\bar{F}(\mu)d\mu$ of acquiring an infinite branch during time $d\mu$, giving (82). Next, for fixed $\mu \geq 1$ the conditional probabilities $P(\mathcal{G}_{A-} = \mathbf{t} \,|\, A = \mu)$ are by (82) proportional to $(\#\mathbf{t})P(\mathcal{G}_\mu = \mathbf{t})$, and thus by (80) proportional to $(\#\mathbf{t})P(\mathcal{G}_{\hat{\mu}} = \mathbf{t})$. ¿From (72), the mean of the Borel($\hat{\mu}$) distribution is $1/(1-\hat{\mu})$, and (83) follows. Finally, the fact that A has distribution function $\bar{F}$ means that $A \stackrel{d}{=} \bar{F}^{-1}(U) \stackrel{d}{=} \bar{F}^{-1}(1-U) = (-\log U)/(1-U)$. So

$$(A, \bar{F}(A)) \stackrel{d}{=} \left(\frac{-\log U}{1-U}, 1-U \right).$$

and hence

$$(A, F(A)) = (A, 1-\bar{F}(A)) \stackrel{d}{=} \left(\frac{-\log U}{1-U}, U \right)$$

which is (84). □

4.3 The PGW$^\infty$(1) distribution

For $0 < \mu \leq 1$ write PGW$^\infty(\mu)$ for the distribution of the infinite tree obtained by conditioning PGW(μ) to be infinite (Proposition 2). In other words, PGW$^\infty(\mu)$ is the distribution of the tree constructed by attaching a sequence of i.i.d. PGW(μ) family trees to a single infinite spine, as described in Corollary 3. The particular case PGW$^\infty$(1) plays a fundamental role in the sequel. The following proposition summarizes some of the many ways this process arises as a weak limit. Here $\mathcal{G}_\mu$ is a PGW(μ) family tree, and $\mathrm{fam}(\mathcal{U}_n)$ is the family tree derived from $\mathcal{U}_n$ with uniform distribution on the set $\mathbf{R}_{[n]}$ of n^{n-1} rooted trees labeled by $[n]$.

Lemma 23 *(i)* (Grimmett [19])

$$\mathrm{fam}(\mathcal{U}_n) \xrightarrow{d} PGW^\infty(1) \text{ as } n \to \infty$$

(ii) For each $\mu \in (0, \infty)$,

$$\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n) \xrightarrow{d} PGW^\infty(1) \text{ as } n \uparrow \infty.$$

(iii) For each $\mu \in (0, 1]$,

$$\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu \geq n) \xrightarrow{d} PGW^\infty(1) \text{ as } n \uparrow \infty.$$

(iv) (Kesten [25]) *For each $\mu \in (0, 1]$*

$$\mathrm{dist}(\mathcal{G}_\mu \,|\, Z_h\mathcal{G}_\mu > 0) \xrightarrow{d} PGW^\infty(\mu) \text{ as } h \uparrow \infty.$$

(v)

$$\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = \infty) \xrightarrow{d} PGW^\infty(1) \text{ as } \mu \downarrow 1.$$

Proof. Grimmett [19] presented the variant of (i) for unlabeled trees, but his argument yields the sharper result for family trees. Part (ii) is the particular case of Proposition 5 for a PGW tree. Either of (i) and (ii) follow from the other due to the $\mathcal{U}_n$-representation (15) of $\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n)$. Part (iii) follows easily from (ii). Part (iv) is Proposition 2 for PGW(μ). Part (v) can be read from (61) for the Poisson family.

4.4 The process $(\mathcal{G}^*_\mu, 0 \leq \mu \leq 1)$

Let $(\mathcal{G}^*_\mu, 0 \leq \mu \leq 1)$ be a uniform pruning of $\mathcal{G}_1^\infty$ with PGW$^\infty$(1) distribution. As explained in the more general setting of Section 3.3, this process should be understood intuitively as $(\mathcal{G}_\mu, 0 \leq \mu \leq 1)$ conditioned on $\#\mathcal{G}_1 = \infty$ where $(\mathcal{G}_\mu, 0 \leq \mu \leq 1)$ is a uniform pruning of $\mathcal{G}_1$ with PGW(1) distribution. Note that $\mathcal{G}^*_{1-} = \mathcal{G}^*_1 = \mathcal{G}_1^\infty$ almost surely, and that for all $\mathbf{t} \in \mathbf{T}, 0 < \mu < 1$ and $0 < \lambda < \infty$

$$\mathrm{dist}(\mathcal{G}^*_{t\mu}, 0 \leq t \leq 1 \,|\, \mathcal{G}^*_\mu = \mathbf{t}) = \mathrm{dist}(\mathcal{G}_{t\lambda}, 0 \leq t \leq 1 \,|\, \mathcal{G}_\lambda = \mathbf{t}). \tag{85}$$

According the following proposition, for each fixed μ with $0 \leq \mu < 1$, the distribution of $\mathcal{G}^*_\mu$ is the PGW(μ) distribution size-biased by total population size. We denote this probability distribution on finite family trees by PGW*(μ). Sheth [42, 43] studied various features of the PGW*(μ) distribution in connection with a model for a coalescent process, and the following corollary of Proposition 15 is closely related to Sheth's results.

Corollary 24 *The process $(\mathcal{G}^*_\mu, 0 \leq \mu < 1)$ is an inhomogeneous Markov chain with countable state space $\mathbf{T}$, whose distribution is uniquely determined by (85) and the following formula:*

$$P(\mathcal{G}^*_\mu = \mathbf{t}) = (1-\mu)(\#\mathbf{t})P(\mathcal{G}_\mu = \mathbf{t}) \qquad \forall \mu \in [0,1), \mathbf{t} \in \mathbf{T}. \tag{86}$$

Proof by specialization. Apply Proposition 15 for $\mathcal{G}_1$ a PGW(1) tree. Then $\mu_1 = 1$, and $\rho^{*-}_u(n) = 1$ for all $0 < u < 1$ and $n = 0, 1, \ldots$, so (55) simplifies to (86).

Autonomous proof. Formula (85) combined with (86) amounts to the following formula: for each non-negative measurable function f defined on the path space $\Omega[0,\mu]$ defined above (60),

$$E[f(\mathcal{G}_t^*, 0 \leq t \leq \mu)] = (1-\mu)E[(\#\mathcal{G}_\mu)f(\mathcal{G}_t, 0 \leq t \leq \mu)].$$

That this is a consistent prescription of $\mathrm{dist}(\mathcal{G}_t^*, 0 \leq t \leq \mu)$ as μ varies amounts to the martingale property of $((1-\mu)\#\mathcal{G}_\mu, 0 \leq \mu < 1)$ obtained in Corollary 21. The existence and uniqueness in distribution of a process $(\mathcal{G}_\mu^*, 0 \leq \mu < 1)$ satisfying (86) and (85) are now clear by Kolmogorov's extension theorem. Lemma 7 implies the existence of $\mathcal{G}_1^* := \mathcal{G}_{1-}^* \in \mathbf{T}^{(\infty)}$ as an almost sure limit, and implies that $(\mathcal{G}_\mu^*, 0 \leq \mu \leq 1)$ is a uniform pruning of $\mathcal{G}_1^*$. To finish the argument it just has to be shown that $\mathcal{G}_1^* \stackrel{d}{=} \mathcal{G}_1^\infty$. That is, for each $h \geq 0$ and $\mathbf{t} \in \mathbf{T}$

$$P(r_h\mathcal{G}_\mu^* = \mathbf{t}) \to P(r_h\mathcal{G}_1^\infty = \mathbf{t}) \text{ as } \mu \uparrow 1. \tag{87}$$

But from (85), for $\mu \in (0,1]$ we can compute

$$P(r_h\mathcal{G}_\mu^* = \mathbf{t}) = \sum_{n=0}^{\infty} P(r_h\mathcal{G}_\mu = \mathbf{t} \,|\, \#\mathcal{G}_\mu = n)P(\#\mathcal{G}_\mu^* = n) \tag{88}$$

where we know by the $\mathcal{U}_n$-representation (15) of $\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n)$ and Lemma 23 that

$$P(r_h\mathcal{G}_\mu = \mathbf{t} \,|\, \#\mathcal{G}_\mu = n) = P(r_h\mathcal{G}_1 = \mathbf{t} \,|\, \#\mathcal{G}_1 = n) \to P(r_h\mathcal{G}_1^\infty = \mathbf{t})$$

as $n \to \infty$. But for each fixed n

$$P(\#\mathcal{G}_\mu^* = n) = (1-\mu)nP(\#\mathcal{G}_\mu = n) \to 0 \text{ as } \mu \uparrow 1$$

by inspection of the Borel formula (10), and (87) now follows easily from (88). $\square$

The corollary above identifies the process $(\mathcal{G}_\mu^*, 0 \leq \mu < 1)$ as the Doob h^*-transform of $(\mathcal{G}_\mu, 0 \leq \mu < 1)$ associated with the space-time harmonic function $h^*(\mu, \mathbf{t}) := (1-\mu)(\#\mathbf{t})$ for the inhomogeneous Markov chain $(\mathcal{G}_\mu, 0 \leq \mu < 1)$ with state-space $\mathbf{T}$. The autonomous proof yields also the following corollary, where the explicit formula (89) is obtained from (75). The limit relation (90) is evident from this argument without calculation, but it can also be checked from (89) and (10) using Stirling's formula.

Corollary 25 *Every non-negative function $h(\mu, n)$ such that $h(0,1) = 1$ and the process $(h(\mu, \#\mathcal{G}_\mu), 0 \leq \mu < 1)$ is a martingale relative to the filtration generated by $(\mathcal{G}_\mu, 0 \leq \mu < 1)$ admits a unique representation as*

$$h(\mu, n) = \sum_{m \in \{1,2,\ldots,\infty\}} P(m) h_m(\mu, n)$$

for some probability distribution $P(\cdot)$ on $\{1, 2, \ldots, \infty\}$ where for $m = 1, 2, \ldots$

$$h_m(\mu, n) = \frac{P(\#\mathcal{G}_1 = m \mid \#\mathcal{G}_\mu = n)}{P(\#\mathcal{G}_1 = m)} = (1-\mu) n e^{n\mu} \frac{m!(m - n\mu)^{m-n-1}}{(m-n)! m^{m-1}} \tag{89}$$

and for $m = \infty$

$$h_\infty(\mu, n) = \lim_{m \to \infty} h_m(\mu, n) = (1 - \mu) n. \tag{90}$$

In particular, the only h such that $h(\mu, \#\mathcal{G}_\mu) \to 0$ almost surely as $\mu \uparrow 1$ is $h(\mu, n) = h_\infty(\mu, n)$, so that $h(\mu, \#\mathcal{G}_\mu) = h^*(\mu, \mathcal{G}_\mu)$ as above.

Essentially, this is an identification of the Martin boundary of the space-time process $((\mu, \#\mathcal{G}_\mu,), 0 \leq \mu < 1)$ with the set $\{1, 2, \ldots, \infty\}$. Similarly, the Martin boundary of the space-time process $((\mu, \mathcal{G}_\mu,), 0 \leq \mu < 1)$ can be identified with the subset $\mathbf{T}_0^{(\infty)}$ of $\mathbf{T}^{(\infty)}$ comprising those trees $\mathbf{t}$ such that if $(\mathcal{T}_\mu, 0 \leq \mu \leq 1)$ is a uniform pruning of $\mathbf{t}$ then $\#\mathcal{T}_\mu < \infty$ almost surely for all $0 \leq \mu < 1$. The extreme harmonic function h corresponding to such a $\mathbf{t}$ is

$$h(\mu, \mathbf{t}) = P(\mathcal{T}_\mu = \mathbf{t}) / P(\mathcal{G}_\mu = \mathbf{t}).$$

4.5 A representation of the ascension process

Consider again the ascension time $A := \inf\{\lambda : \#\mathcal{G}_\lambda = \infty\}$ for the PGW pruning process, studied in Section 4.2. Combining Corollary 24 with the formulae of Lemma 22 leads to the following rather surprising representation of the ascension process $(\mathcal{G}_\lambda, 0 \leq \lambda < A)$.

Proposition 26

$$(\mathcal{G}_\lambda, 0 \leq \lambda < A) \stackrel{d}{=} \left(\mathcal{G}^*_{\lambda U}, 0 \leq \lambda < \frac{-\log U}{1 - U} \right) \tag{91}$$

where U is uniform $(0,1)$, independent of $(\mathcal{G}^*_\mu, 0 \leq \mu \leq 1)$.

Proof. Take $(\mathcal{G}^*_\mu, 0 \le \mu \le 1)$ independent of A. Then

$$P(\mathcal{G}^*_{\hat{A}} = \mathbf{t} | A = a) = P(\mathcal{G}^*_{\hat{a}} = \mathbf{t}) = (1 - \hat{a})(\#\mathbf{t}) P(\mathcal{G}_{\hat{a}} = \mathbf{t})$$

by (86). Comparing with (83), we see $(A, \mathcal{G}_{A-}) \stackrel{d}{=} (A, \mathcal{G}^*_{\hat{A}})$. By conditioning on these terminal values and reversing time, (85) implies

$$(\mathcal{G}_\lambda, 0 \le \lambda < A) \stackrel{d}{=} (\mathcal{G}^*_{\hat{A}\lambda/A}, 0 \le \lambda < A).$$

Use (84) to rewrite the right side in terms of U and (91) follows. □

Since $P(A > 1) = 1$, the identity in distribution (91) implies in particular that

$$\mathcal{G}_\lambda \stackrel{d}{=} \mathcal{G}^*_{\lambda U} \qquad \forall\, 0 \le \lambda \le 1 \tag{92}$$

where λU has uniform distribution on $(0, \lambda)$ independent of $(\mathcal{G}^*_\mu, 0 \le \mu \le 1)$. We spell out the meaning of (92) in the following corollary:

Corollary 27 *Fix $0 < \lambda \le 1$. Let $\mathcal{G}^\infty_1$ have $PGW^\infty(1)$ distribution, and independently of $\mathcal{G}^\infty_1$ let U_λ have uniform distribution on $(1 - \lambda, 1)$. Given $\mathcal{G}^\infty_1$ and U_λ, construct a random forest $\mathcal{F}_\lambda$ by cutting each edge of $\mathcal{G}^\infty_1$ independently with probability U_λ. Then the family tree derived from the component of $\mathcal{F}_\lambda$ that contains the root of $\mathcal{G}^\infty_1$ has distribution $PGW(\lambda)$.*

The identity (92) can also be checked as follows. By Corollary 24 and the $\mathcal{U}_n$-representation (15) of $\operatorname{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n)$, the random trees $\mathcal{G}_\lambda$ and $\mathcal{G}^*_{\lambda U}$ share a common conditional distribution given their total size. So (92) amounts to $\#\mathcal{G}_\lambda \stackrel{d}{=} \#\mathcal{G}^*_{\lambda U}$ for all $0 \le \lambda \le 1$, that is

$$P_\lambda(n) = \frac{1}{\lambda} \int_0^\lambda P^*_\mu(n) d\mu \qquad \forall\, 0 < \lambda \le 1, n = 1, 2, \ldots \tag{93}$$

where $P_\lambda(\cdot)$ is the Borel(λ) distribution of $\#\mathcal{G}_\lambda$ displayed in (10), and P^*_μ is the distribution of $\mathcal{G}^*_\mu$, which by Corollary 24 is the *size-biased Borel*(μ) *distribution*

$$P^*_\mu(n) := P(\#\mathcal{G}^*_\mu = n) = (1 - \mu) n P_\mu(n). \tag{94}$$

But (93) in turn amounts to

$$\frac{d}{d\lambda} [\lambda P_\lambda(n)] = P^*_\lambda(n) \qquad \forall\, 0 < \lambda < 1, n = 1, 2, \ldots \tag{95}$$

which is easily checked by calculus using the formula (10) for $P_\lambda(n)$. We restate (93) in the following corollary:

Corollary 28 *For each $0 < \lambda \leq 1$ the Borel(λ) distribution is the uniform mixture of size-biased Borel(μ) distributions over $0 < \mu < \lambda$.*

Let N_μ denote a random variable with Borel(μ) distribution $P_\mu(\cdot)$, and N_μ^* a random variable with size-biased Borel(μ) distribution $P_\mu^*(\cdot)$. ¿From (94) and (95), for $0 < \mu < 1$

$$E[f(N_\mu^*)] = (1-\mu)E[N_\mu f(N_\mu)] = E[f(N_\mu)] + \mu \frac{d}{d\mu} E[f(N_\mu)] \tag{96}$$

where the first equality holds for every non-negative function f, and the second equality holds at μ for any f such that the derivative exists, as can be seen by differentiation of the next formula (97) with $nf(n)$ instead of $f(n)$. Apply (93) and the first equality of (96) with $f(n)/n$ instead of $f(n)$ to deduce that for all non-negative function f and $0 \leq \mu \leq 1$

$$E\left(\frac{f(N_\mu)}{N_\mu}\right) = \frac{1}{\mu}\int_0^\mu (1-\lambda)E[f(N_\lambda)]\,d\lambda. \tag{97}$$

These formulae imply numerous identities involving moments of N_μ and N_μ^*. For example, (96) for $f(n) = n$ gives easily

$$E(N_\mu^*) = (1-\mu)E(N_\mu^2) = (1-\mu)^{-2}. \tag{98}$$

Take $f = 1$ in (97) to obtain

$$E(1/N_\mu) = \frac{1}{\mu}\int_0^\mu (1-\lambda)\,d\lambda = 1 - \mu/2. \tag{99}$$

For $d_k(n) := (n)_k/n^k$, where $(n)_k := n!/(n-k)!$ the formula $d_{k+1}(n) = d_k(n) - d_k(n)/n$ combined with (97) and an easy induction shows that

$$E(d_k(N_\mu)) = \mu^{k-1}/k, \qquad \forall\ k \geq 1,\ \mu \in [0,1]. \tag{100}$$

A similar calculation yields

$$E(d_k(N_\mu^*)) = \mu^{k-1}, \qquad \forall\ k \geq 1,\ \mu \in [0,1). \tag{101}$$

Since $d_k(n)$ is the probability of no repeats in a sequence of k independent uniform random picks from a set of n elements, we deduce the following curious result. See also section 4.7 for related results.

Corollary 29 *Let $\mathcal{G}_\mu$ have PGW(μ) distribution for some $0 < \mu \leq 1$. Given $\mathcal{G}_\mu$ let $V_1, \ldots, V_k$ be k vertices of $\mathcal{G}_\mu$ picked independently and uniformly at random. Then the unconditional probability that these k vertices of $\mathcal{G}_\mu$ are all distinct is μ^{k-1}/k. For $\mathcal{G}^*_\mu$ instead of $\mathcal{G}_\mu$ the corresponding probability is μ^{k-1} for all $0 < \mu < 1$.*

4.6 The spinal decomposition of $\mathcal{G}^*_\mu$

Fix $0 < \mu \leq 1$. A random tree $\mathcal{G}^*_\mu$ with PGW$^*(\mu)$ distribution has a number of remarkable properties as a consequence of the results in Section 3.3. Following the notation of that section, suppose that $\mathcal{G}^*_\mu$ has been constructed as $\mathcal{G}^*_\mu = \mathrm{fam}(\mathcal{G}^\dagger_\mu)$ where $\mathcal{G}^\dagger_\mu$ is the component containing the root in the subgraph of $\mathcal{G}^\infty_1$ consisting of those edges e with $\xi_e \leq \mu$ where the ξ_e are independent uniform$(0,1)$ random variables. Let (V_h) be the infinite spine of $\mathcal{G}^\infty_1$, and let $H_\mu := \sup\{n : V_n \in \mathcal{G}^{\infty\dagger}_\mu\}$, so H_μ has the geometric$(1-\mu)$ distribution

$$P(H_\mu = n) = (1-\mu)\mu^n \qquad \forall\ n = 0, 1, \ldots \tag{102}$$

Let $V^*_\mu \in \mathcal{G}^*_\mu$ be the vertex of $\mathcal{G}^*_\mu$ at height H_μ which corresponds to $V_{H_\mu} \in \mathcal{G}^\dagger_\mu$ via the relabeling map from $\mathcal{G}^\dagger_\mu \to \mathcal{G}^*_\mu$. According to formula (57) specialized to the case at hand, there is the following refinement of (86). For $0 \leq \mu < 1$ and $\mathcal{G}^*_\mu$ and $V^*_\mu \in \mathcal{G}^*_\mu$ defined as above,

$$P(\mathcal{G}^*_\mu = \mathbf{t}, V^*_\mu = v) = (1-\mu)P(\mathcal{G}_\mu = \mathbf{t})\ \forall \mathbf{t} \in \mathbf{T}, v \in \mathbf{t}.$$

That is, the vertex V^*_μ of $\mathcal{G}^*_\mu$ at height H_μ is a *uniform random vertex* of $\mathcal{G}^*_\mu$, meaning that the conditional distribution of V_μ given $\mathcal{G}^*_\mu$ is uniform on $\mathcal{G}^*_\mu$. By further examination of the argument in Section 3.3 we deduce:

Corollary 30 (Spinal decomposition of $\mathcal{G}^*_\mu$). *Fix $0 < \mu \leq 1$. For a random tree $\mathcal{G}^*_\mu$ with PGW$^*(\mu)$ distribution, and V^* a uniform random vertex of $\mathcal{G}^*_\mu$, let H_μ be the height of V^*, and let* $(\mathrm{root} = V^*_0, \ldots, V^*_{H_\mu} = V^*)$ *be the path in $\mathcal{G}^*_\mu$ from the root to V^*, call it a* spine *of $\mathcal{G}^*_\mu$. For $0 \leq i \leq H_\mu$ let $\mathcal{G}(i)$ be the family tree derived from the subtree of $\mathcal{G}^*_\mu$ with root V^*_i in the forest obtained by deleting all edges of $\mathcal{G}^*_\mu$ on the path from the root to V^*. Then*

(i) H_μ has the geometric$(1-\mu)$ distribution (102).

(ii) given $H_\mu = h$ the $\mathcal{G}(i)$ for $0 \leq i \leq h$ are independent with PGW(μ) distribution, and

(iii) given $H_\mu = h \geq 1$ and these family trees $\mathcal{G}(i)$ for $0 \leq i \leq h$, the path from the root to V^* is defined by $V_i^* = (J_1, \cdots, J_i)$ for $1 \leq i \leq h$ where the J_i are independent and J_i has uniform $[Z\mathcal{G}(i) + 1]$ distribution.

Proof. Lemma 17 combined with the spinal decomposition of $\mathcal{G}_\mu^\infty$ given in Corollary 3 show that this result holds for the particular construction of $\mathcal{G}_\mu^*$ and H_μ used to obtain (57), with $V^* = V_\mu^*$. But by change of variables the result must also be true as stated for any triple $(\mathcal{G}_\mu^*, V^*, H_\mu)$ with the same joint distribution as this particular triple $(\mathcal{G}_\mu^*, V_\mu^*, H_\mu)$. $\square$

We remark that our spinal decomposition is the probabilistic analog of similar combinatorial decompositions of Joyal [23] and Labelle [28], which were partly anticipated by Meir and Moon [32].

Note that for a fixed $0 < \mu < 1$, it only makes sense to speak of *a* spine of a $\mathrm{PGW}^*(\mu)$ distributed tree $\mathcal{G}_\mu^*$ rather than *the* spine of $\mathcal{G}_\mu^*$, because the construction of the spine involves the extra randomization of picking a uniform random vertex V^* of $\mathcal{G}_\mu^*$. But according to the above discussion, if $(\mathcal{G}_\mu^*, 0 \leq \mu < 1)$ is a uniform pruning process such that $\mathcal{G}_\mu^*$ has $\mathrm{PGW}^*(\mu)$ distribution for each $0 \leq \mu < 1$, then with probability one this process grows a unique infinite spine as $\mu \uparrow 1$. This is the spine of the tree $\mathcal{G}_{1-}^*$ with $\mathrm{PGW}^\infty(1)$ distribution, and this infinite spine induces a finite spine in $\mathcal{G}_\mu^*$ for each $0 \leq \mu < 1$ in such a way that the length $H_\mu + 1$ of this spine is increasing as μ increases. In this construction each of the non-negative integer-valued processes $(H_\mu, 0 \leq \mu < 1)$ and $(\#\mathcal{G}_\mu^*, 0 \leq \mu < 1)$ is increasing and inhomogeneous Markov. The transition probabilities of (H_μ) are quite obvious. Those of $(\#\mathcal{G}_\mu^*)$ can be read from formula (75) and the fact that $(\#\mathcal{G}_\mu^*)$ is the Doob h^∞-transform of $(\#\mathcal{G}_\mu)$ for $h^\infty(\mu, n) = (1 - \mu)n$.

Since in the setting of Corollary 30 the entire family tree $\mathcal{G}_\mu^*$ can be reconstructed from the random elements whose joint law is described by (i)-(iii), the spinal decomposition implies the the following recursive construction of a $\mathrm{PGW}^*(\mu)$ tree:

Corollary 31 *Let random elements*

$$(H_\mu;\ \mathcal{G}(i), 0 \leq i \leq H_\mu;\ J_i, 1 \leq i \leq H_\mu)$$

have the joint distribution described in (i)-(iii) of Corollary 30. Recursively define trees $\mathcal{G}^*(i), 0 \leq i \leq H_\mu$ and vertices $V_i^*, 0 \leq i \leq H_\mu$, as follows. Let $\mathcal{G}^*(0) = \mathcal{G}(0), V_0^* = 0$; for $1 \leq i \leq H_\mu$ let $\mathcal{G}^*(i)$ be the family tree *obtained by attachment of $\mathcal{G}(i)$ to $\mathcal{G}^*(i-1)$ as the J_ith child of V^*_{i-1}, and let $V^*_i = (V^*_{i-1}, J_i)$. Then $\mathcal{G}^*(H_\mu)$ has $PGW^*(\mu)$ distribution, and the vertex $V^*_{H_\mu}$ at height H_μ is a random vertex of $\mathcal{G}^*(H_\mu)$.*

4.7 Some distributional identities

Distributional relationships between random trees imply distributional relationships between the integer-valued random variables which record the sizes of trees. In this section we spell out several such relationships.

Borel and size-biased Borel distributions The spinal decomposition (Corollary 30) expresses $\mathcal{G}^*_\mu$ as the union of $H_\mu + 1$ subtrees which can be relabeled as independent PGW(μ) trees. Since we know that $\#\mathcal{G}^*_\mu$ has a size-biased Borel (μ) distribution (94), we deduce:

Corollary 32 *Let $N_\mu(1), N_\mu(2), \dots$ be independent with the Borel(μ) distribution (10), independent also of H_μ with the geometric($1-\mu$) distribution (102). Then*

$$N^*_\mu := N_\mu(1) + \cdots + N_\mu(H_\mu + 1) \text{ has size-biased Borel}(\mu) \text{ distribution.} \tag{103}$$

An elementary proof of Corollary 32 can be given as follows. By conditioning on $H_\mu = h$ and using the Borel-Tanner formula (11) for the distribution of the sum of $h+1$ independent Borel(μ) variables, for N^*_μ defined by the sum in (103), and $P^*_\mu(\cdot)$ the size-biased Borel distribution, we obtain for all $0 \le h \le n-1$

$$P(H_\mu = h, N^*_\mu = n) = \frac{(h+1)(n-1)!}{n^{h+1}(n-h-1)!} P^*_\mu(n). \tag{104}$$

The right side is easily summed over $0 \le h \le n-1$ to confirm that $P(N^*_\mu = n) = P^*_\mu(n)$.

The decomposition (103) should be compared to the obvious consequence of (6) that for Z_μ with Poisson (μ) distribution independent of i.i.d. Borel(μ) variables $N_\mu(i)$

$$N_\mu := N_\mu(1) + \cdots + N_\mu(Z_\mu) + 1 \text{ has Borel}(\mu) \text{ distribution.} \tag{105}$$

We do not know of any reference to the representation (103) of the size-biased Borel distribution in terms of the more elementary Borel and geometric distributions, or to the companion representation (93) of the Borel distribution as a mixture of size-biased Borel distributions. But both the Borel and size-biased Borel distributions have found applications in various contexts [11, 10, 21, 42, 43, 44, 50] where these representations might prove useful. See [13, 36, 45] for study of the general class of Lagrangian distributions, which includes both the Borel and size-biased Borel distributions as particular cases.

It is a well known consequence of (105) that the Borel(μ) distribution P_μ of N_μ is infinitely divisible. Representation (103) gives $P^*_\mu(\cdot)$ as a convolution of P_μ and $\mathrm{dist}(N_\mu(1) + \ldots + N_\mu(H_\mu))$, and the latter inherits the infinite divisibility property of H_μ. We deduce

Corollary 33 *The size-biased Borel(μ) distribution $P^*_\mu(\cdot)$ is infinitely divisible.*

More on the height of a random vertex. For any finite rooted random tree $\mathcal{T}$, let $H\mathcal{T}$ denote the height of a uniform random vertex V of $\mathcal{T}$. Let $\mathcal{U}_n$ have uniform distribution on the n^{n-1} rooted trees labeled by $[n]$. ¿From (86) and the $\mathcal{U}_n$-representation (15) of $\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n)$ we have for $0 \leq \mu < 1$ that

$$P(H\mathcal{G}^*_\mu = h \,|\, \#\mathcal{G}^*_\mu = n) = P(H\mathcal{G}_\mu = h \,|\, \#\mathcal{G}_\mu = n) = P(H\mathcal{U}_n = h). \tag{106}$$

On the other hand, by the spinal decomposition of $\mathcal{G}^*_\mu$ (Corollary 30), there is the identity $(H\mathcal{G}^*_\mu, \#\mathcal{G}^*_\mu) \stackrel{d}{=} (H_\mu, N^*_\mu)$ for (H_μ, N^*_μ) with the joint distribution (104). It follows that

$$P(H\mathcal{U}_n = h) = \frac{(h+1)(n-1)!}{n^{h+1}(n-h-1)!} \quad \forall \;\; 0 \leq h \leq n-1. \tag{107}$$

For $n \geq 2$ let D_n be the number of vertices on the path from 1 to 2 in $\mathcal{U}_n$. By symmetry, the conditional distribution of $H\mathcal{U}_n$, given that the random vertex V of $\mathcal{U}_n$ is not the root, is identical to the unconditional distribution of $D_n - 1$. So

$$P(H\mathcal{U}_n = h) = \frac{1}{n} 1(h = 0) + \frac{(n-1)}{n} P(D_n = h+1) 1(h \geq 1)$$

and (107) amounts to the result of Meir and Moon [32] that

$$P(D_n = k) = \frac{k\,(n-2)!}{n^{k-1}\,(n-k)!} \quad \forall\, 2 \le k \le n. \tag{108}$$

As a variation, for $\mathcal{G}_\mu$ with PGW(μ) distribution, we can apply Corollary 92 to compute for $h = 0, 1, 2, \ldots$ and $\mu \in (0, 1]$

$$P(H\mathcal{G}_\mu \ge h) = \frac{1}{\mu}\int_0^\mu P(H\mathcal{G}^*_\lambda \ge h)d\lambda = \frac{1}{\mu}\int_0^\mu \lambda^h d\lambda = \frac{\mu^h}{h+1}. \tag{109}$$

Compare with the consequence of Corollary 30 (i) that for $\mu \in (0, 1)$

$$P(H\mathcal{G}^*_\mu \ge h) = \mu^h. \tag{110}$$

These formulae can also be checked by conditioning on $\#\mathcal{G}_\mu$ or $\#\mathcal{G}^*_\mu$ to reduce to (107) and then using (100) and (101).

Some asymptotic distributions. It is known [32] that as $n \to \infty$ the asymptotic distribution of $(H\mathcal{U}_n)/\sqrt{n}$ is that of a random variable R with the *Rayleigh density* $re^{-r^2/2}$ for each $r > 0$. Also, as $\mu \uparrow 1$ the asymptotic distribution of $(1-\mu)^2 N^*_\mu$ is that of Z^2 where Z now denotes a normal variable with $E(Z) = 0$ and $E(Z^2) = 1$. Both these assertions are easily checked by asymptotic density calculations, which establish corresponding local limit theorems. Since from (86) and (15) we have that

$$\mathrm{dist}(H\mathcal{G}^*_\mu \,|\, \#\mathcal{G}^*_\mu = n) = \mathrm{dist}(H\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n) = \mathrm{dist}(H\mathcal{U}_n)$$

and $\#\mathcal{G}^*_\mu \stackrel{d}{=} N^*_\mu$ it follows that for μ close to 1 the distribution of $(1-\mu)(H\mathcal{G}^*_\mu)$ must be close to that of $(1-\mu)\sqrt{N^*_\mu}R$ where R is independent of N^*_μ, and hence also close to that of $|Z|R$ where R is independent of Z. On the other hand, we know that $H\mathcal{G}^*_\mu$ has geometric$(1-\mu)$ distribution, so it is elementary that the asymptotic distribution of $(1-\mu)(H\mathcal{G}^*_\mu)$ is that of a standard exponential variable ϵ. Thus we deduce the non-trivial identity in distribution

$$|Z|R \stackrel{d}{=} \epsilon. \tag{111}$$

In terms of probability densities, this amounts to the well known formula

$$\sqrt{\frac{2}{\pi}}\int_0^\infty e^{-(y^2+t^2/y^2)/2}dy = e^{-t}$$

which can be verified by showing that the functions of t on both sides are equal at 0 and satisfy the same ordinary differential equation. Let Γ_t denote a random variable with the gamma(t) distribution defined by the density $\Gamma(t)^{-1}x^{t-1}e^{-x}$ for $x > 0$. Since

$$|Z| \stackrel{d}{=} \sqrt{2\Gamma_{1/2}}; \quad R \stackrel{d}{=} \sqrt{2\Gamma_1}; \quad \epsilon \stackrel{d}{=} \Gamma_1$$

the identity (111) is the special case $t = 1/2$ of the identity in distribution

$$4\Gamma_t\Gamma_{t+1/2} \stackrel{d}{=} \Gamma_{2t}^2 \quad \forall\, t > 0 \tag{112}$$

where Γ_t and $\Gamma_{t+1/2}$ are independent, which is due to Wilks [51]. Evaluation of moments shows that both (111) and (112) are equivalent to the duplication formula for the gamma function

$$\Gamma(2z) = 2^{2z-1}\Gamma(z)\Gamma(z+1/2)/\Gamma(1/2).$$

See Gordon [18] for further probabilistic interpretations of gamma function identities.

4.8 Size-Modified PGW-trees

Several results of the previous sections have natural generalizations to the following class of distributions on the set $\mathbf{T}$ of finite family trees. Call a random tree $\mathcal{G}^\circ$ a *size-modified Poisson-Galton-Watson tree (SMPGW tree)*, or use the same acronym for its distribution, if $\mathcal{G}^\circ$ has distribution of the form

$$P(\mathcal{G}^\circ = \mathbf{t}) = f(\#\mathbf{t})P(\mathcal{G}_1 = \mathbf{t}) \quad \forall \mathbf{t} \in \mathbf{T} \tag{113}$$

for some f with $E(f(\#\mathcal{G}_1)) = 1$, where $\mathcal{G}_1$ is a PGW(1) tree. The distribution of such a tree $\mathcal{G}^\circ$ is determined by its *size distribution* $Q(\cdot)$, that is the distribution of $\#\mathcal{G}^\circ$ on the positive integers which is given by $Q(n) = f(n)P_1(n)$ where $P_1(\cdot)$ is the Borel(1) distribution of $\mathcal{G}_1$. Let $(\mathcal{U}_n)$ be a sequence of random trees such that $\mathcal{U}_n$ has uniform distribution on the set $\mathbf{R}_{[n]}$ of n^{n-1} rooted trees labeled by $[n]$, and let $\mathcal{T}_n := \mathrm{fam}(\mathcal{U}_n)$. By the $\mathcal{U}_n$-representation (15) of $\mathrm{dist}(\mathcal{G}_\mu \,|\, \#\mathcal{G}_\mu = n)$, formula (113) is equivalent to

$$\mathrm{dist}(\mathcal{G}^\circ \,|\, \#\mathcal{G}^\circ = n) = \mathrm{dist}(\mathcal{T}_n) \;\; \forall\, n : P(\#\mathcal{G}^\circ = n) > 0. \tag{114}$$

That is to say

$$P(\mathcal{G}^\circ = \mathbf{t}) = \sum_{n=1}^{\infty} Q(n) P(\mathcal{T}_n = \mathbf{t}) \tag{115}$$

where $P(\mathcal{T}_n = \mathbf{t})$ given by formula (15). So the most general SMPGW distribution is obtained as the distribution of a random tree $\mathcal{G}^\circ$ constructed as follows. Independent of the sequence of random trees $(\mathcal{T}_n)$ let S have distribution $Q(\cdot)$. Then $\mathcal{G}^\circ := \mathcal{T}_S$ has the distribution displayed in (115). The set of all SMPGW distributions on $\mathbf{T}$ is therefore a simplex whose extreme points are the distributions of $\mathcal{T}_n$ for $n = 1, 2, \ldots$.

Clearly, PGW(μ) for $\mu \in [0,1]$ and PGW$^*(\mu)$ for $\mu \in [0,1)$ are SMPGW distributions. Typically a result for SMPGW can be established first for the extreme distributions of $\mathcal{T}_n$, either by a combinatorial argument or by conditioning a result already obtained for the PGW or PGW* family, and then extended to the SMPGW family by mixing over the extreme distributions. Following are several illustrations of this theme.

Proposition 34 *Suppose $(\mathcal{G}^\circ_u, 0 \le u \le 1)$ is a uniform pruning of $\mathcal{G}^\circ$ with a SMPGW distribution. Then*

(i) The process $(\#\mathcal{G}^\circ_u, 0 \le u \le 1)$ is Markov with the following co-transition probabilities: for $0 < q < 1$

$$P(\mathcal{G}^\circ_{qu} = \ell \,|\, \mathcal{G}^\circ_u = m) = P(\#\mathcal{U}_{m,q} = \ell) \quad \forall\ 1 \le \ell \le m \tag{116}$$

where $\mathcal{U}_{m,q}$ is the component subtree containing root$(\mathcal{U}_m)$ *after each edge of $\mathcal{U}_m$ is deleted independently with probability $1-q$.*

(ii) (Moon [33]) *For all $0 < q < 1$ and $1 \le \ell \le m$*

$$P(\#\mathcal{U}_{m,q} = \ell) = \binom{m}{\ell} \frac{(1-q)}{q} \left(\frac{q}{m}\right)^{m-1} \ell^{\ell} \left(\frac{m}{q} - \ell\right)^{m-\ell-1} \tag{117}$$

(iii) For each $0 \le u \le 1$ the tree $\mathcal{G}^\circ_u$ is an SMPGW tree whose size distribution is given by

$$P(\#\mathcal{G}^\circ_u = \ell) = \sum_{m=1}^{\infty} P(\#\mathcal{U}_{m,u} = \ell) P(\#\mathcal{G}^\circ_1 = m) \tag{118}$$

Proof. In the particular case when $\mathcal{G}^\circ_1 = \mathcal{G}_\mu$ has a PGW(μ) distribution for some $0 \le \mu \le 1$, the Markov property of $(\#\mathcal{G}^\circ_u, 0 \le u \le 1)$ was established in Corollary 21. By conditioning on $\#\mathcal{G}_1^\circ = n$ in this special case, it follows that $(\#\mathcal{G}_u^\circ, 0 \le u \le 1)$ must also be Markov with the same co-transition probabilities in the extreme case when $\mathcal{G}_1^\circ = \mathcal{T}_n$, hence also by mixing for any SMPGW tree $\mathcal{G}_1^\circ$. The formula (116) for the co-transition probabilities follows easily from (114). Moon found formula (117) by a combinatorial argument. By application of Bayes rule, this formula for the co-transition probabilities of $(\#\mathcal{G}_u^\circ, 0 \le u \le 1)$ can be obtained from the forwards transition probabilities (75) of the same process, or vice versa. For part (iii), it is enough to consider the case $\mathcal{G}_1^\circ = \mathrm{fam}(\mathcal{U}_n)$, in which case $(\mathcal{G}_u^\circ)$ can be constructed as $\mathcal{G}_u^\circ = \mathrm{fam}(\mathcal{U}_{n,u})$ after using independent uniform variables to define $(\mathcal{U}_{n,u}, 0 \le u \le 1)$ as an increasing process of subtrees of $\mathcal{U}_n$. The problem is to show that

$$\mathrm{dist}(\mathrm{fam}(\mathcal{U}_{n,u}) \,|\, \#\mathrm{fam}(\mathcal{U}_{n,u}) = m) = \mathrm{dist}(\mathrm{fam}(\mathcal{U}_m)) \tag{119}$$

By an easy combinatorial argument, for each subset V of $[n]$ with $\#V = m$, given $\mathrm{verts}(\mathcal{U}_{n,u}) = V$ the tree $\mathcal{U}_{n,u}$ has uniform distribution on the set of all m^{m-1} rooted trees labeled by V. It follows that for each $V \subseteq [n]$ with $\#V = m$

$$\mathrm{dist}(\mathrm{fam}(\mathcal{U}_{n,u}) \,|\, \mathrm{verts}(\mathcal{U}_{n,u}) = V) = \mathrm{dist}(\mathrm{fam}(\mathcal{U}_m))$$

which of course implies (119). □

Consider now the closure $\overline{\mathrm{SMPGW}}$ of SMPGW, that is the set of all probability distributions on $\mathbf{T}^{(\infty)}$ obtainable as weak limits of some sequence of SMPGW distributions. By an easy variation of the autonomous proof of Corollary 24, every distribution in $\overline{\mathrm{SMPGW}}$ is a mixture of the $\mathrm{PGW}^\infty(1)$ distribution of $\mathcal{G}_1^\infty$ and some SMPGW distribution. That is to say, $\mathcal{G}^\circ$ has a $\overline{\mathrm{SMPGW}}$ distribution iff

$$P(\mathcal{G}^\circ \in \cdot) = \sum_{n \in \{1,2,\ldots,\infty\}} P(\#\mathcal{G}^\circ = n) P(\mathcal{T}_n \in \cdot) \tag{120}$$

where $\mathcal{T}_\infty \stackrel{d}{=} \mathcal{G}_1^\infty$ has $\mathrm{PGW}^\infty(1)$ distribution. The following characterization of the PGW^* process now follows from Lemma 7 and Corollary 25:

Proposition 35 *Suppose that $(\mathcal{G}_u^\circ, 0 \le u < 1)$ is a uniform pruning process such that $\mathcal{G}_u^\circ$ has a SMPGW distribution for each $0 \le u < 1$. Then $(\mathcal{G}_u^\circ, 0 \le u \le 1)$ is a uniform pruning of $\mathcal{G}_1^\circ := \lim_{u \uparrow 1} \mathcal{G}_1^\circ$, which has a $\overline{\mathrm{SMPGW}}$ distribution, and the following conditions are equivalent:*

(i) $\lim_{u\uparrow 1} P(\#\mathcal{G}_u^\circ \leq n) = 0$ *for every* $n = 1, 2, \ldots$
(ii) $\mathrm{dist}(\mathcal{G}_u^\circ) = PGW^*(u)$ *for every* $u \in (0, 1]$
(iii) $\mathrm{dist}(\mathcal{G}_1^\circ) = PGW^\infty(1)$.

The spinal decomposition for a SMPGW tree. It is instructive to consider the analog for a SMPGW tree of the spinal decomposition of PGW$^*(\mu)$ stated in Corollary 30. Let V^* be a uniform random vertex of $\mathcal{G}^\circ$, let $H\mathcal{G}^\circ$ be the height of V^*, and construct family trees $\mathcal{G}^\circ(i)$ as before by cutting all edges along the path from 0 to V^*. Then instead of (i),(ii) and (iii) in Corollary 30 it is clear that by application of that Corollary and (107) we have

(i)

$$P(H\mathcal{G}^\circ = h, \#\mathcal{G}^\circ = n) = \frac{(h+1)(n-1)!}{n^{h+1}(n-h-1)!} P(\#\mathcal{G}^\circ = n) \ \ \forall\ 0 \leq h \leq n-1$$

(ii) given $H\mathcal{G}^\circ = h$ and $\#\mathcal{G}^\circ = n$ the $\mathcal{G}^\circ(i)$ for $0 \leq i \leq h$ are distributed like $h+1$ independent PGW(1) trees conditionally given that the sum of their sizes is n.

(iii) The conditional distribution of the path from 0 to V^* given $H\mathcal{G}^\circ = h \geq 1$ and these family trees $\mathcal{G}^\circ(i)$ for $0 \leq i \leq h$ is just as described in (iii) of Corollary 30 for $\mathcal{G}^\circ = \mathcal{G}_\mu^*$.

It follows easily that the trees $\mathcal{G}^\circ(i)$ are i.i.d. iff $\mathcal{G}^\circ$ has PGW$^*(\mu)$ distribution for some $\mu \in (0, 1]$. To more precise about the converse:

Corollary 36 *If a SMPGW tree $\mathcal{G}^\circ$ is such that given $H\mathcal{G}^\circ = 1$ the sizes $\#\mathcal{G}^\circ(0)$ and $\#\mathcal{G}^\circ(1)$ are independent, then $\mathcal{G}^\circ$ has $PGW^*(\mu)$ distribution for some $\mu \in (0, 1]$.*

In particular, for $\mathcal{G}^\circ$ with PGW(μ) distribution the $\#\mathcal{G}^\circ(i)$ for $0 \leq i \leq H\mathcal{G}^\circ$ are not conditionally i.i.d. given $H\mathcal{G}^\circ$. But they are exchangeable:

Corollary 37 *For a SMPGW tree $\mathcal{G}^\circ$, the family trees $\mathcal{G}^\circ(i)$ for $0 \leq i \leq H\mathcal{G}^\circ$ are conditionally exchangeable given $H\mathcal{G}^\circ$.*

This consequence of the spinal decomposition of $\mathcal{G}^\circ$ can also be proved by first checking it combinatorially for $\mathcal{G}^\circ = \mathcal{T}_n$. Indeed, this case is implicit in the combinatorial results [23, 28].

References

[1] D.J. Aldous. A random tree model associated with random graphs. *Random Structures Algorithms*, 1:383–402, 1990.

[2] D.J. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, 3:450–465, 1990.

[3] D.J. Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1:228–266, 1991.

[4] D.J. Aldous. The continuum random tree I. *Ann. Probab.*, 19:1–28, 1991.

[5] D.J. Aldous. The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.

[6] D.J. Aldous. Deterministic and stochastic models for coalescence: a review of the mean-field theory for probabilists. Unpublished. Available via homepage http://www.stat.berkeley.edu/users/aldous, 1996.

[7] D.J. Aldous and J. Pitman. Tree- and forest-valued Markov chains. In Preparation, 1997.

[8] N. Alon and J.H. Spencer. *The Probabilistic Method.* Wiley, New York, 1992.

[9] K.B. Athreya and P. Ney. *Branching Processes.* Springer, 1972.

[10] S. Berg and J. Jaworski. Probability distributions related to the local structure of a random mapping. In A. Frieze and T. Łuczak, editors, *Random Graphs*, volume 2, pages 1–21. Wiley, 1992.

[11] S. Berg and L. Mutafchiev. Random mappings with an attracting center: Lagrangian distributions and a regression function. *J. Appl. Probab.*, 27:622 – 636, 1990.

[12] E. Borel. Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'un infinité de coefficients. Application au probleme de l'attente á un guichet. *C.R. Acad. Sci. Paris*, 214:452–456, 1942.

[13] P.C. Consul. *Generalized Poisson Distributions.* Dekker, 1989.

[14] N. Dershowitz and S. Zaks. Enumerations of ordered trees. *Discrete Mathematics*, 31:9–28, 1980.

[15] M. Dwass. The total progeny in a branching process. *J. Appl. Probab.*, 6:682–686, 1969.

[16] W. Feller. *An Introduction to Probability Theory and its Applications, Vol 1,3rd ed.* Wiley, New York, 1968.

[17] P. Fitzsimmons, J. Pitman, and M. Yor. Markovian bridges: construction, Palm interpretation, and splicing. In E. Çinlar, K.L. Chung, and M.J. Sharpe, editors, *Seminar on Stochastic Processes, 1992*, pages 101–134. Birkhäuser, Boston, 1993.

[18] L. Gordon. A stochastic approach to the gamma function. *Amer. Math. Monthly*, 101:858–865, 1994.

[19] G. R. Grimmett. Random labelled trees and their branching networks. *J. Austral. Math. Soc. (Ser. A)*, 30:229–237, 1980.

[20] H. Haase. On the incipient cluster of the binary tree. *Arch. Math. (Basel)*, 63:465–471, 1994.

[21] F.A. Haight and M.A. Breuer. The Borel-Tanner distribution. *Biometrika*, 47:143–150, 1960.

[22] T.E. Harris. *The Theory of Branching Processes.* Springer-Verlag, New York, 1963.

[23] A. Joyal. Une théorie combinatoire des séries formelles. *Adv. in Math.*, 42:1–82, 1981.

[24] D.P. Kennedy. The Galton-Watson process conditioned on the total progeny. *J. Appl. Probab.*, 12:800–806, 1975.

[25] H. Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.*, 22:425–487, 1987.

[26] V.F. Kolchin. Branching processes, random trees, and a generalized scheme of arrangements of particles. *Mathematical Notes of the Acad. Sci. USSR*, 21:386–394, 1977.

[27] V.F. Kolchin. *Random Mappings*. Optimization Software, New York, 1986. (Translation of Russian original).

[28] G. Labelle. Une novelle démonstration combinatoire des formules d'inversion de lagrange. *Adv. in Math.*, 42:217–247, 1981.

[29] R. Lyons. Random walks, capacity, and percolation on trees. *Ann. Probab.*, 20:2043–2088, 1992.

[30] R. Lyons. Probability on trees. Book in preparation, 1996.

[31] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proof of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23:1125–1138, 1995.

[32] A. Meir and J.W. Moon. The distance between points in random trees. *J. Comb. Theory*, 8:99–103, 1970.

[33] J. W. Moon. A problem on random trees. *J. Comb. Theory B*, 10:201–205, 1970.

[34] J. Neveu. Arbres et processus de Galton-Watson. *Ann. Inst. H. Poincaré Probab. Statist.*, 22:199 – 207, 1986.

[35] R. Otter. The multiplicative process. *Ann. Math. Statist.*, 20:206–224, 1949.

[36] A. G. Pakes and T. P. Speed. Lagrange distributions and their limit theorems. *SIAM Journal on Applied Mathematics*, 32:745 – 754, 1977.

[37] R. Pemantle. Uniform random spanning trees. In J. Laurie Snell, editor, *Topics in Contemporary Probability*, pages 1–54, Boca Raton, FL, 1995. CRC Press.

[38] J. Pitman. Coalescent random forests. Technical Report 457, Dept. Statistics, U.C. Berkeley, 1996.

[39] J. Pitman. Enumerations of trees and forests related to branching processes and random walks. Technical Report 482, Dept. Statistics, U.C. Berkeley, 1997.

[40] C.R. Rao and H. Rubin. On a characterization of the Poisson distribution. *Sankhyā, Ser. A*, 26:294–298, 1964.

[41] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales*, volume 1,Foundations. Wiley, 1994. 2nd. edition.

[42] R.K. Sheth. Merging and hierarchical clustering from an initially Poisson distribution. *Mon. Not. R. Astron. Soc.*, 276:796–824, 1995.

[43] R.K. Sheth. Galton-Watson branching processes and the growth of gravitational clustering. *Mon. Not. R. Astron. Soc.*, 281:1277–1289, 1996.

[44] R.K. Sheth and J. Pitman. Coagulation and branching process models of gravitational clustering. To appear in *Mon. Not. R. Astron. Soc.*, 1997.

[45] M. Sibuya, N. Miyawaki, and U. Sumita. Aspects of Lagrangian probability distributions. *Studies in Applied Probability. Essays in Honour of Lajos Takács (J. Appl. Probab.)*, 31A:185–197, 1994.

[46] R. Stanley. Enumerative combinatorics, vol. 2. Book in preparation, to be published by Cambridge University Press, 1996.

[47] L. Takács. Queues, random graphs and branching processes. *J. Applied Mathematics and Simulation*, 1:223–243, 1988.

[48] L. Takács. Limit distributions for queues and random rooted trees. *J. Applied Mathematics and Stochastic Analysis*, 6:189–216, 1993.

[49] J.C. Tanner. A problem of interference between two queues. *Biometrika*, 40:58–69, 1953.

[50] J.C. Tanner. A derivation of the Borel distribution. *Biometrika*, 1961:222–224, 1961.

[51] S. S. Wilks. Certain generalizations in the analysis of variance. *Biometrika*, 24:471–494, 1932.