ESAIM: PS 20 (2016) 400–416
DOI: 10.1051/ps/2016019

ESAIM: Probability and Statistics
www.esaim-ps.org

CONDITIONED MULTI-TYPE GALTON−WATSON TREES

Eric Cator[1] and Henk Don[1]

Abstract. We consider multi-type Galton Watson trees, and find the distribution of these trees when conditioning on very general types of recursive events. It turns out that the conditioned tree is again a multi-type Galton Watson tree, possibly with more types and with offspring distributions depending on the type of the father node and on the height of the father node. These distributions are given explicitly. We give some interesting examples for the kind of conditioning we can handle, showing that our methods have a wide range of applications.

Mathematics Subject Classification. 60J80.

Received July 22, 2015. Revised February 19, 2016. Accepted July 13, 2016.

1. Introduction

The asymptotic shape of conditioned Galton−Watson trees has been widely studied. For example, one could condition on the number of nodes of the tree being n, and letting $n \to \infty$. In a lot of cases, the limiting tree is quite well understood, see for example the survey paper by Janson [4]. Some work on finite conditioned trees has been done by Geiger and Kersting [2], who study the shape of a tree conditioned on having height exactly equal to n. In this context, we also mention the spinal construction of a Galton−Watson tree conditioned to reach generation k, as derived in [3].

In this paper, we will investigate conditioning multi-type Galton−Watson trees on events of a recursive nature (as explained in Sect. 2), one example being conditioning on survival to a given level. The main idea is that we consider different classes of trees, where the class of a tree is determined by the types and classes of her children. The offspring distribution of a node depends on its type and on the level of the tree where this node is living. In fact, we show that the conditioned tree again is a multi-type Galton−Watson tree and how this can be used to directly construct such a conditioned tree. Our approach can be seen as a generalization of the well-known decomposition of a supercritical Galton−Watson tree into nodes whose offspring survives forever and nodes whose offspring eventually goes to extinction as discussed in [5].

Section 2.2 discusses a couple of examples that illustrate the applicability of our results. We give an example concerning mutants in a population, we discuss an alternative to Geiger's construction of a tree conditioned on having height exactly k and we show how to condition on the size of the kth generation.

Keywords and phrases. Multi-type Galton−Watson tree, conditioning, recursive events.

[1] Institute for Mathematics, Astrophysics, and Particle Physics, Faculty of Science, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands. e.cator@math.ru.nl, henkdon@gmail.com

Article published by EDP Sciences

1.1. Notation and preliminaries

We will consider rooted multi-type Galton–Watson trees with arbitrary offspring distribution, that can depend on the current generation. In such a tree each node has a type, which we indicate by a natural number $t \in \Theta := \{1, \dots, \theta\}$. If the root of a tree has type t, we use a bold-face t to denote this root. Define the set of trees of heigth 0 as $\mathcal{T}_0 = \{\mathbf{1}, \mathbf{2}, \dots, \boldsymbol{\theta}\}$. Then we define inductively for $k \geq 1$ the set of trees of height at most k by

$$\mathcal{T}_k = \mathcal{T}_0 \sqcup \bigsqcup_{n=1}^{\infty} \{(t, [T_1, \dots, T_n]) \mid T_i \in \mathcal{T}_{k-1}, t \in \Theta\},$$

and denote the set of all trees by $\mathcal{T} = \bigcup_{k=0}^{\infty} \mathcal{T}_k$. For a tree $T = (t, [T_1, \dots, T_n]) \in \Theta \times (\mathcal{T}_{k-1})^n$, the trees $T_1, \dots, T_n$ will be called the children (of the root) of T (notation: $T_i \prec T$). The type of T will be just the type of its root and will be denoted by $r(T)$. We now define a function $N : \mathcal{T} \to \mathbb{N}^\theta$ that counts how many children of each type a tree T has:

$$N(T) = (N_1(T), \dots, N_\theta(T)), \quad \text{where} \quad N_t(T) = \#\left\{\tilde{T} \prec T : r(\tilde{T}) = t\right\}.$$

The set of trees of heigth at most k having a root of type t is denoted by

$$\mathcal{T}_k^t = \{T \in \mathcal{T}_k : r(T) = t\} = \{\mathbf{t}\} \sqcup \bigsqcup_{n=1}^{\infty} \{(t, [T_1, \dots, T_n]) \mid T_i \in \mathcal{T}_{k-1}\}.$$

Let $\mathcal{T}^t = \bigcup_{k=0}^{\infty} \mathcal{T}_k^t$. Denote the offspring distribution of a type t node at height $l \geq 0$ by μ_l^t, for arbitrary probability measures μ_l^t on $\mathbb{N}^\theta = \{0, 1, \dots\}^\theta$. Define independent random variables $W_l^t \sim \mu_l^t$. For a vector $\mathbf{x} \in \mathbb{N}^\theta$, we write the corresponding multinomial coefficient as

$$D(\mathbf{x}) = \binom{|\mathbf{x}|_1}{x_1, \dots, x_\theta} = \frac{\left(\sum_{i=1}^{\theta} x_i\right)!}{\prod_{i=1}^{\theta} x_i!}.$$

We now introduce the Galton–Watson probability measures on $\mathcal{T}_k^t$. Firstly, let P_0^t be the trivial probability measure on $\mathcal{T}_0^t$, so $P_0^t(\mathbf{t}) = 1$. Now define inductively the probability measure P_{lk}^t for $0 \leq l \leq k$ and $k \geq 1$ as the following probability measure on $\mathcal{T}_{k-l}^t$: if $l = k$, then $P_{kk}^t = P_0^t$. Otherwise for all $T \in \mathcal{T}_{k-l}^t$

$$P_{lk}^t(T) = \frac{\mathbb{P}(W_l^t = N(T))}{D(N(T))} \prod_{\tilde{T} \prec T} P_{l+1,k}^{r(\tilde{T})}(\tilde{T}),$$

where empty products are taken to be 1. The intuition is that the second sub-index determines the height of the final tree we are considering, whereas the first sub-index determines at which level we are building up the tree (so P_{lk}^t generates trees of type t at level l of size $k - l$). We are interested in P_{0k}^t, which is the Galton–Watson probability measure on $\mathcal{T}_k^t$ (trees cut off at height k with a root of type t).

In the next section we will introduce a class of recursive-type events on which we would like to condition, and discuss several examples of such events. In Section 3 we will introduce the conditional measures corresponding to our events, and in Section 4 we show that these conditional measures indeed coincide with the original Galton–Watson measure, conditioned on our event.

2. Conditioning on recursive events

In this section we introduce a class of recursive-type events on which we would like to condition, such as the event that the tree survives until a specific level.

2.1. Partitioning the set of trees

We will now set up our general framework and show how some examples fit into it. We start by choosing $k_0 \in \mathbb{N}$ and partitioning $\mathcal{T}_{k_0}$ into m classes $A_{k_0}^{(1)}, \ldots, A_{k_0}^{(m)}$. Typically, all trees in such a class have some property that all trees in the other classes do not have. One of the simplest examples would be a partition into two classes, where trees that survive until some level k are in the first class and all other trees in the second. The partition of $\mathcal{T}_{k_0}$ will be the starting point to recursively define partitions of $\mathcal{T}_l$, $k_0 < l \leq k$ into sets $A_l^{(i)}, i = 1, \ldots, m$, where k is the (maximum) height of the trees that we are considering. Suppose that the partition of $\mathcal{T}_{l-1}$ is already defined. Then we are able to introduce a counting matrix for trees in $\mathcal{T}_l$. Define $C_l : \mathcal{T}_l \to \mathbb{N}^{m\times\theta}$ such that for $T \in \mathcal{T}_l$ the (i, j)th position is given by

$$C_l^{(i,j)}(T) = \#\left\{\tilde{T} \prec T : \tilde{T} \in A_{l-1}^{(i)} \cap \mathcal{T}^j\right\},$$

so this is the number of children of T having type j and being an element of the ith partition class. Now for $k_0 < l \leq k$ we partition $\mathbb{N}^{m\times\theta}$ into subsets $B_l^{(1)}, \ldots, B_l^{(m)}$. This partition is the key for the recursive definition of $A_l^{(i)}$. The set $A_l^{(i)}$ will contain exactly those trees for which the counting matrix $C_l(T)$ is in $B_l^{(i)}$:

$$A_l^{(i)} = \left\{T \in \mathcal{T}_l : C_l(T) \in B_l^{(i)}\right\}. \tag{2.1}$$

2.2. Examples

Before going into the details of the construction of conditioned trees, we will discuss some examples of recursive events that can be handled by our approach.

2.2.1. Genetic mutations

Suppose we have a population in which sometimes an individual (mutant) is born having a particular mutation in its genetic material. This mutation can be inherited by subsequent generations. Suppose we know the probability that the root is mutated. Such a population can be described as a two-type Galton−Watson process in which the offspring distribution is type- and possibly level-dependent. We take $\Theta = \{1, 2\}$ to be the set of types, where mutants have type 1.

Suppose we would like to condition on the event "there is at least one mutant in the kth generation". Choose $k_0 = 0$ and partition $\mathcal{T}_0 = \{\mathbf{1}, \mathbf{2}\}$ into the classes $A_0^{(1)} = \{\mathbf{1}\}$ and $A_0^{(2)} = \{\mathbf{2}\}$. For $0 < l \leq k$, we want to define $A_l^{(1)}$ and $A_l^{(2)}$ by

$$A_l^{(1)} = \{T \in \mathcal{T}_l : \text{there is a mutant at level } l\}, \qquad A_l^{(2)} = \mathcal{T}_l \backslash A_l^{(1)}.$$

These events satisfy a recursive relation: $A_l^{(1)}$ contains exactly those trees that have at least one child in $A_{l-1}^{(1)}$. For $T \in \mathcal{T}_l$, the first row of the 2×2-counting matrix $C_l(T)$ counts the children of T that are in $A_{l-1}^{(1)}$. Therefore, for all $0 < l \leq k$ we let

$$B_l^{(1)} = \left\{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{N}^{2\times 2} : a + b \geq 1\right\}, B_l^{(2)} = \mathbb{N}^{2\times 2} \backslash B_l^{(1)},$$

and now (2.1) gives the desired partition of $\mathcal{T}_l$. For the sake of illustration, we note that with a minor change, we can condition on "there is at least one mutant in the kth generation inheriting its mutation from the root". To achieve this, it suffices to merely redefine $B_l^{(1)}$ and $B_l^{(2)}$ for all $0 < l \leq k$ as follows

$$B_l^{(1)} = \left\{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{N}^{2\times 2} : a \geq 1\right\}, B_l^{(2)} = \mathbb{N}^{2\times 2} \backslash B_l^{(1)}.$$

In these two examples, we defined one partition class $A_l^{(1)} \subseteq \mathcal{T}_l$ by the event on which conditioning is required. The only other partition class was just the complement of the first one. Finding a suitable partition of the set of trees is not always that obvious, as is demonstrated in the next example. We will show how to condition on the slightly more complicated event "All mutants in the tree inherit their mutation from the root and at least one mutant is present in generation k". As before, define one partition class $A_l^{(1)}$ as the set of trees satisfying the condition. Here it is not sufficient to define only one other partition class. One obstacle is that some trees (namely those with a "spontaneous mutation") in the complement $\mathcal{T}_l \setminus A_l^{(1)}$ are forbidden as a child of trees in $A_{l+1}^{(1)}$ and others are not.

Nevertheless, with a slightly more elaborate partition, we can still handle this case. We distinguish four classes and partition $\mathcal{T}_0$ into

$$A_0^{(1)} = \{\mathbf{1}\}, \quad A_0^{(2)} = \{\mathbf{2}\} \quad \text{and} \quad A_0^{(3)} = A_0^{(4)} = \emptyset.$$

For $0 < l \le k$, we define the following subsets of $\mathbb{N}^{4\times 2}$:

$$B_l^{(1)} = \left\{ \begin{bmatrix} a & 0 \\ c & d \\ 0 & 0 \\ g & 0 \end{bmatrix} : a \ge 1 \right\}, \qquad B_l^{(2)} = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & d \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \right\}$$

$$B_l^{(3)} = \left\{ \begin{bmatrix} a & b \\ c & d \\ e & f \\ g & h \end{bmatrix} : b + e + f + h \ge 1 \right\}, B_l^{(4)} = \left\{ \begin{bmatrix} 0 & 0 \\ c & d \\ 0 & 0 \\ g & 0 \end{bmatrix} : c + g \ge 1 \right\}$$

As can be easily checked, these sets are disjoint and $\bigcup_i B_l^{(i)} = \mathbb{N}^{4\times 2}$, so this indeed is a partition. With this partition of $\mathbb{N}^{4\times 2}$, the classes $A_l^{(i)}$ are determined, see (2.1). We will give an explicit description of these classes in a moment. Recall that $a, \ldots, h$ count the children of a tree $T \in \mathcal{T}_l$, where mutants are counted in the first column of the matrix, non-mutants in the second and the row determines to which partition class $A_{l-1}^{(i)}$, $i = 1, \ldots, 4$ a child belongs. For instance, f is the number of non-mutated children in class $A_{l-1}^{(3)}$.

It follows by induction that the sets $A_l^{(i)}$ partition $\mathcal{T}_l$ in such a way that:

- $A_l^{(1)}, 0 < l \le k$ contains exactly the trees having
  - at least one mutated child of which the mutated progeny reaches level l, and
  - no "spontaneous" mutants in the progeny of their children.
- $A_l^{(2)}, 0 < l \le k$ contains the trees having only type 2 descendants.
- $A_l^{(3)}, 0 < l \le k$ contains the trees having a type 2 descendant with a type 1 child ("spontaneous mutation").
- $A_l^{(4)}, 0 < l \le k$ contains all other trees in $\mathcal{T}_l$.

Note that these classes are defined by properties of the children of a tree and not by the type of the tree itself. For example, a tree in $A_l^{(2)}$ can have a type 1 root, but all its descendants have type 2. The conditional measure we are interested in is now obtained by conditioning P_{0k}^1 on $A_k^{(1)}$.

2.2.2. Conditioning on the size of generation k

As a next example, we show how to condition a single-type Galton−Watson tree on having exactly G individuals in the kth generation. In this case, we partition $\mathcal{T}_0 = \{\mathbf{1}\}$ into $G + 2$ classes by defining

$$A_0^{(1)} = \{\mathbf{1}\}, \quad A_0^{(0)} = A_0^{(2)} = A_0^{(3)} = \ldots = A_0^{(G)} = A_0^{(G+1)} = \emptyset.$$

Define $x \in \mathbb{N}^{G+2}$ by $x := [0 \quad 1 \quad 2 \quad \ldots \quad G \quad G+1]^T$. For $0 < l \le k$, we define

$$B_l^{(i)} = \left\{ y \in \mathbb{N}^{G+2} : x^T y = i \right\}, \quad B_l^{(G+1)} = \left\{ y \in \mathbb{N}^{G+2} : x^T y \ge G + 1 \right\},$$

where $0 \le i \le G$. Partitioning $\mathcal{T}_l$ according to (2.1) gives the following: for $0 \le i \le G$, $A_l^{(i)}$ contains the trees of which the lth generation has exactly size i, while $A_l^{(G+1)}$ contains the trees of which the lth generation has at least size $G + 1$. Conditioning P_{0k} on $A_k^{(G)}$ gives the result we are looking for.

2.2.3. The tree has heigth exactly k

As a final illustration, we explain how to condition a Galton−Watson tree on having height exactly k, thus producing an alternative for the construction of Geiger and Kersting [2]. We consider trees in $\mathcal{T}_{k+1}$ that are conditioned to reach level k, but not level $k+1$. We start by choosing $k_0 = 2$, and partitioning $\mathcal{T}_2$ into three sets, namely correct trees, short trees and long trees:

$$A_2^{(1)} = \{T \in \mathcal{T}_2 \mid T \text{ reaches level 1, but not level 2}\},$$

$$A_2^{(2)} = \{T \in \mathcal{T}_2 \mid T \text{ does not reach level 1}\} = \{\mathbf{0}\},$$

$$A_2^{(3)} = \{T \in \mathcal{T}_2 \mid T \text{ reaches level 2}\}.$$

Define for each $2 < l \leq k+1$

$$B_l^{(1)} = \{n \in \mathbb{N}^3 \mid n_1 \geq 1, n_3 = 0\},$$

$$B_l^{(2)} = \{n \in \mathbb{N}^3 \mid n_1 = 0, n_3 = 0\},$$

$$B_l^{(3)} = \{n \in \mathbb{N}^3 \mid n_3 \geq 1\},$$

and let $\mathcal{T}_l$ be partitioned as in (2.1). This construction guarantees that if a tree $T \in \mathcal{T}_{k+1}$ is an element of $A_{k+1}^{(1)}$, then it has at least one child that reaches level k, and no children that reach level $k+1$. If $T \in A_{k+1}^{(2)}$, all its children do not reach level k, and if $T \in A_{k+1}^{(3)}$, then at least one child reaches level $k+1$. Conditioning on being in $A_{k+1}^{(1)}$ therefore gives the desired result.

2.3. Remarks following the examples

As it turns out from the examples in the previous section, the setup allows to condition on quite a variety of events. A fundamental requirement on these events is that they are determined only by the *number* of children of a tree having particular properties. So we can (for instance) not distinguish between trees having the same children in a different order.

An additional example is discussed in detail in [1]. As an application of the theory developed in the present paper, the cost of searching a tree to a given level is determined. The proposed model takes into account costs for having a lot of children, but also for walking into dead ends. So both a high expected offspring and a low expected offspring would give high search costs. This gives rise to an optimization problem: which offspring distribution gives minimal costs? For this model the conditional probability measures are explicitly constructed, leading to recursions that enable us to calculate the costs and solve the optimization problem for Poisson offspring.

Conditioning on recursive events as in the examples allows us to compute (conditional) probabilities that are defined in terms of such events. As an illustration: in the example on genetic mutations we can easily compute the probability that the root is mutated, given that there is at least one mutant in generation k. What makes the results even more useful is that they show how to directly construct a tree conditioned on some event. This means that trees conditioned on (rare) events can be studied by just simulating them.

3. Conditional measures

In this section we construct an alternative measure $\tilde{P}_{lk}^t$ on $\mathcal{T}_{k-l}^t$, that depends on the event we want to condition on. As soon as we have this measure, conditioning on the desired event is a triviality. In the next section, we will show that in fact the two measures P_{lk}^t and $\tilde{P}_{lk}^t$ are the same.

3.1. Definition of the conditional measure

Recall that k_0 is the smallest index for which the partition classes $A_{k_0}^{(i)}$ are defined. Starting from this index, we can define the probability on a tree of type t in each of the classes as follows. Define ${}^t p_{lk}^{(i)}$ for $0 \leq l \leq k-k_0$ by

$$ {}^t p_{lk}^{(i)} = P_{lk}^t \left(A_{k-l}^{(i)} \cap \mathcal{T}^t \right). $$

We can calculate this probability in a recursive way. Denote, for $q \in [0,1]^m$ with $\sum q_i = 1$, by $\mathrm{Multi}(n,q) \in \mathbb{N}^m$ the multinomial distribution where we distribute n elements over m classes, according to the probabilities q_i. We also choose independent random vectors $W_l^t \sim \mu_l^t$ according to the offspring distribution of a type t node at level l and denote the jth coordinate by $W_{l,j}^t$. Then, for $l < k - k_0$

$$ {}^t p_{lk}^{(i)} = \mathbb{P}\left(\bigotimes_{j=1}^{\theta} \mathrm{Multi}\left(W_{l,j}^t, ({}^j p_{l+1,k}^{(1)}, \ldots, {}^j p_{l+1,k}^{(m)}) \right) \in B_{k-l}^{(i)} \right), \tag{3.1} $$

where, for $a_0, \ldots, a_\theta \in \mathbb{N}^m$, we defined $\Lambda := \bigotimes_{j=1}^{\theta} a_j \in \mathbb{N}^{m \times \theta}$ to be the matrix for which $\Lambda_{ij} = a_j(i)$.

We proceed by defining the conditional measure ${}^t\tilde{Q}_{lk}^{(i)}$ on $A_{k-l}^{(i)} \cap \mathcal{T}^t$. To do this, define for each $t \in \{1, \ldots, \theta\}$ and $0 \leq l \leq k - k_0 - 1$ on the same probability space as W_l^t, the random matrices

$$ {}^t X_{lk} = \begin{pmatrix} {}^t X_{lk}^{(1,1)} & \ldots & {}^t X_{lk}^{(1,\theta)} \\ \vdots & \ddots & \vdots \\ {}^t X_{lk}^{(m,1)} & \ldots & {}^t X_{lk}^{(m,\theta)} \end{pmatrix}, $$

such that conditional on W_l^t, all columns are independent and the distribution of the jth column satisfies

$$ \left({}^t X_{lk}^{(1,j)}, \ldots, {}^t X_{lk}^{(m,j)} \right) \mid W_l^t \sim \mathrm{Multi}\left(W_{l,j}^t, ({}^j p_{l+1,k}^{(1)}, \ldots, {}^j p_{l+1,k}^{(m)}) \right). $$

Note that with these definitions

$$ {}^t p_{lk}^{(i)} = \mathbb{P}\left({}^t X_{lk} \in B_{k-l}^{(i)} \right). $$

The full joint distribution of $(W_l^t, {}^t X_{lk})$ is now determined. For a type t node at level l, the distribution of its children over the θ types is given by the random vector W_l^t. Furthermore, the jth column of ${}^t X_{lk}$ represents how the type j children of this type t node are distributed over the m classes. For $l = k - k_0$, we define for each $T \in A_{k_0}^{(i)} \cap \mathcal{T}^t$

$$ {}^t\tilde{Q}_{k-k_0,k}^{(i)}(T) = \frac{P_{k-k_0,k}^t(T)}{P_{k-k_0,k}^t(A_{k_0}^{(i)} \cap \mathcal{T}^t)}, $$

as a probability measure on $A_{k_0}^{(i)} \cap \mathcal{T}^t$. Next, we inductively define the probability measures ${}^t\tilde{Q}_{lk}^{(i)}$ on $A_{k-l}^{(i)} \cap \mathcal{T}^t$ for each $0 \leq l \leq k - k_0 - 1$ such that for each $T \in A_{k-l}^{(i)} \cap \mathcal{T}^t$

$$ {}^t\tilde{Q}_{lk}^{(i)}(T) = \frac{\mathbb{P}({}^t X_{lk} = C_{k-l}(T) \mid {}^t X_{lk} \in B_{k-l}^{(i)})}{D(C_{k-l}(T))} \prod_{j=1}^{m} \prod_{\tilde{T} \prec T : \tilde{T} \in A_{k-l-1}^{(j)}} {}^{r(\tilde{T})}\tilde{Q}_{l+1,k}^{(j)}(\tilde{T}), $$

where we extended the definition of D (see Sect. 1.1) to integer-valued matrices, *i.e.* for a matrix A

$$ D(A) = \frac{\left(\sum_{i,j} A_{i,j} \right)!}{\prod_{i,j} A_{i,j}!}, \tag{3.2} $$

where once again empty products are taken to be 1. Note that this definition is valid for all $T \in \mathcal{T}^t_{k-l}$: we simply get ${}^t\tilde{Q}^{(i)}_{lk}(T) = 0$ whenever $T \notin A^{(i)}_{k-l}$. We can now define the alternative measure $\tilde{P}^t_{lk}$ on $\mathcal{T}^t_{k-l}$:

$$\tilde{P}^t_{lk}(T) = \sum_{i=1}^{m} {}^t p^{(i)}_{lk} \, {}^t\tilde{Q}^{(i)}_{lk}(T).$$

3.2. Constructing a tree according to the conditional measure

We will now show that constructing a tree according to the measure $\tilde{P}^t_{lk}$ boils down to generating an unconditional multi-type non-homogeneous Galton Watson tree. This tree has more types, but the advantage of the alternative measure $\tilde{P}^t_{lk}$ is that conditioning on $A^{(i)}_{k-l}$ is trivial. In the next section, we will show that the alternative measure $\tilde{P}^t_{lk}$ is equal to the original measure P^t_{lk}.

We can describe the random tree $\mathbb{T} \sim \tilde{P}^t_{lk}$ as follows. The root of the tree has type t. To construct the tree, we first toss an m-sided coin to determine in which of the m classes $\mathbb{T}$ is, giving probability ${}^t p^{(i)}_{lk}$ to the ith class $A^{(i)}_{k-l}$. If $\mathbb{T} \in A^{(i)}_{k-l}$, then we choose it according to ${}^t\tilde{Q}^{(i)}_{lk}$. This means that we choose $(\tilde{W}^t_{lk}, {}^t\tilde{X}_{lk})$, where $\tilde{W}^t_{lk}$ counts the numbers of children of $\mathbb{T}$ of each type and ${}^t\tilde{X}_{lk}$ counts for each type the numbers of children that will lie in each of the m classes, according to

$$(\tilde{W}^t_{lk}, {}^t\tilde{X}_{lk}) \sim (W^t_l, {}^t X_{lk}) \mid {}^t X_{lk} \in B^{(i)}_{k-l}.$$

The $\sum_{i,j} {}^t\tilde{X}^{(i,j)}_{lk}$ children are distributed over the $\sum_j \tilde{W}^t_{lk,j}$ positions uniformly at random. Then for each child of type j in $A^{(i)}_{k-l-1}$ we draw a tree according to ${}^j\tilde{Q}^{(i)}_{l+1,k}$.

In this way we have described the random tree as a Galton−Watson tree with $m \times \theta$ 'types' of children and type- and level-dependent offspring distribution. Note that conditioning $\tilde{P}^t_{lk}$ on $A^{(i)}_{k-l}$ is trivial: we simply have to draw $\mathbb{T}$ according to ${}^t\tilde{Q}^{(i)}_{lk}$.

4. The two trees have the same distribution

The following theorem shows that the construction procedure of Section 3 in fact generates trees with the same probabilities as under the original Galton−Watson measure. Fix k and k_0 and define all measures as before.

Theorem 4.1. *For all $0 \le l \le k - k_0$, $t \in \Theta$ and $T \in \mathcal{T}^t_{k-l}$,*

$$P^t_{lk}(T) = \tilde{P}^t_{lk}(T).$$

Proof. The theorem is true by construction for $l = k - k_0$. Now suppose that we have already shown that $P^t_{l+1,k} = \tilde{P}^t_{l+1,k}$ for all t. Choose $T \in \mathcal{T}_{k-l}$. Then there exists a unique i such that $T \in A^{(i)}_{k-l}$. Before we show that $\tilde{P}^t_{lk}(T) = P^t_{lk}(T)$, we collect some useful observations. First of all, note that the number of ways to distribute the individuals over the positions in C_{k-l} can be written as a product by first assigning a type to each individual and then distributing all individuals of a given type over the classes (writing C_{k-l} for $C_{k-l}(T)$ and N for $N(T)$):

$$D(C_{k-l}) = D(N) \prod_{j=1}^{\theta} D\left(\left(C^{(i,j)}_{k-l}\right)_{i=1}^{m}\right).$$

Secondly, note that by definition $\mathbb{P}({}^tX_{lk} \in B^{(i)}_{k-l})$ is equal to ${}^tp^{(i)}_{lk}$. Next, since $C_{k-l}(T)$ determines $N(T)$, we have:

$$\begin{aligned}
\mathbb{P}({}^tX_{lk} = C_{k-l}) &= \mathbb{P}\left({}^tX_{lk} = C_{k-l}, W^t_l = N\right)\\
&= \mathbb{P}\left(W^t_l = N\right)\mathbb{P}\left({}^tX_{lk} = C_{k-l} \mid W^t_l = N\right)\\
&= \mathbb{P}\left(W^t_l = N\right)\prod_{j=1}^{\theta}\mathbb{P}\left(\left({}^tX^{(i,j)}_{lk}\right)_{i=1}^{m} = \left(C^{(i,j)}_{k-l}\right)_{i=1}^{m} \mid W^t_{l,j} = N_j\right)\\
&= \mathbb{P}\left(W^t_l = N\right)\prod_{j=1}^{\theta} D\left(\left(C^{(i,j)}_{k-l}\right)_{i=1}^{m}\right)\prod_{i=1}^{m}\left({}^jp^{(i)}_{l+1,k}\right)^{C^{(i,j)}_{k-l}}\\
&= \frac{\mathbb{P}\left(W^t_l = N\right)D(C_{k-l})}{D(N)}\prod_{j=1}^{\theta}\prod_{i=1}^{m}\left({}^jp^{(i)}_{l+1,k}\right)^{C^{(i,j)}_{k-l}}.
\end{aligned}$$

Combining these observations gives, since $T \in A^{(i)}_{k-l}$,

$$\begin{aligned}
\tilde{P}^t_{lk}(T) &= {}^tp^{(i)}_{lk}\,{}^t\tilde{Q}^{(i)}_{lk}(T)\\
&= {}^tp^{(i)}_{lk}\frac{\mathbb{P}({}^tX_{lk} = C_{k-l}(T) \mid {}^tX_{lk} \in B^{(i)}_{k-l})}{D(C_{k-l}(T))}\prod_{j=1}^{m}\prod_{\tilde{T}\prec T:\tilde{T}\in A^{(j)}_{k-l-1}} {}^{r(\tilde{T})}\tilde{Q}^{(j)}_{l+1,k}(\tilde{T})\\
&= {}^tp^{(i)}_{lk}\frac{\mathbb{P}({}^tX_{lk} = C_{k-l}(T))}{D(C_{k-l}(T))\mathbb{P}({}^tX_{lk} \in B^{(i)}_{k-l})}\prod_{j=1}^{m}\prod_{\tilde{t}\in\Theta}\prod_{\tilde{T}\prec T:\tilde{T}\in A^{(j)}_{k-l-1}\cap\mathcal{T}^{\tilde{t}}} {}^{\tilde{t}}\tilde{Q}^{(j)}_{l+1,k}(\tilde{T})\\
&= \left(\frac{\mathbb{P}\left(W^t_l = N\right)}{D(N)}\prod_{\tilde{t}=1}^{\theta}\prod_{j=1}^{m}\left({}^{\tilde{t}}p^{(j)}_{l+1,k}\right)^{C^{(j,\tilde{t})}_{k-l}}\right)\left(\prod_{j=1}^{m}\prod_{\tilde{t}\in\Theta}\prod_{\tilde{T}\prec T:\tilde{T}\in A^{(j)}_{k-l-1}\cap\mathcal{T}^{\tilde{t}}} {}^{\tilde{t}}\tilde{Q}^{(j)}_{l+1,k}(\tilde{T})\right)\\
&= \frac{\mathbb{P}\left(W^t_l = N\right)}{D(N)}\prod_{j=1}^{m}\prod_{\tilde{t}\in\Theta}\prod_{\tilde{T}\prec T:\tilde{T}\in A^{(j)}_{k-l-1}\cap\mathcal{T}^{\tilde{t}}} {}^{\tilde{t}}p^{(j)}_{l+1,k}\,{}^{\tilde{t}}\tilde{Q}^{(j)}_{l+1,k}(\tilde{T})\\
&= \frac{\mathbb{P}\left(W^t_l = N\right)}{D(N)}\prod_{j=1}^{m}\prod_{\tilde{t}\in\Theta}\prod_{\tilde{T}\prec T:\tilde{T}\in A^{(j)}_{k-l-1}\cap\mathcal{T}^{\tilde{t}}} \tilde{P}^{\tilde{t}}_{l+1,k}(\tilde{T})\\
&= \frac{\mathbb{P}\left(W^t_l = N\right)}{D(N)}\prod_{\tilde{T}\prec T} P^{r(\tilde{T})}_{l+1,k}(\tilde{T})\\
&= P^t_{lk}(T).
\end{aligned}$$

□

5. Example: Genetic mutations revisited

In this section we use our results to work out one of the examples of Section 2.2, to demonstrate how the method works in general. For these calculations it will turn out to be very useful that our conditioned tree is again a multi-type Galton−Watson tree. We consider a population with mutants and let the set of types be $\Theta = \{1, 2\}$, where type 1 denotes a mutant. The theory developed in the preceding sections allows for any kind of offspring distribution. In the current example the number of children Σ of a type t node will have a geometric distribution (denoted $\mathrm{Geo}(\pi_t)$):

$$\mathbb{P}(\Sigma = n) = (1 - \pi_t)^n\pi_t, \qquad n = 0, 1, 2, \ldots, \tag{5.1}$$

which gives expectation

$$\mathbb{E}[\Sigma] = \frac{1-\pi_t}{\pi_t}\cdot \tag{5.2}$$

Furthermore, we assume that each child of a type t node has probability p_t to be a mutant itself, independent of all other children. Our method also works without this independence. We could take any probability distribution on pairs $(n_m, n-n_m)$, where n_m is the number of mutants. In Section 5.2 there is a remark explaining how to handle this. The offspring distributions could also be made level-dependent, without essential changes in the calculations.

5.1. Recursive definition of events

We will condition on the event that there is at least one mutant in the kth generation. The corresponding partition (see also Sect. 2.2.1) of $\mathcal{T}_l$ is given by the classes

$$\begin{aligned} A_l^{(1)} &= \{T \in \mathcal{T}_l : \text{there is a mutant at level } l\} = \left\{T \in \mathcal{T}_l : C_l(T) \in B_l^{(1)}\right\}, \\ A_l^{(2)} &= \mathcal{T}_l \backslash A_l^{(1)}, \end{aligned}$$

where $B_l^{(1)}$ is the set of matrices

$$B_l^{(1)} = \left\{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{N}^{2\times 2} : a+b \geq 1\right\}.$$

Remember that this means that a tree in $\mathcal{T}_l$ is an element of $A_l^{(1)}$ if and only if it has a type 1 or a type 2 child in $A_{l-1}^{(1)}$. Also, to start the recursion, we take $A_0^{(1)} = \{\mathbf{1}\}$ and $A_0^{(2)} = \{\mathbf{2}\}$, so $k_0 = 0$.

5.2. Unconditional probability distributions

We will now derive the recursions for the probabilities ${}^t p_{lk}^{(i)}$ (the probability that a type t node rooted at level l in a tree of height k belongs to class $A_{k-l}^{(i)}$). A type t subtree that starts on level $l < k$ has two types of children, and each child is in one of the two classes. Type and class of a child are, by definition, independent of all other children's properties. The total number of children Σ has the geometric distribution (5.1). So we can introduce four new 'types', occurring according to the following joint distributions for $t = 1, 2$:

$$\begin{cases} \Sigma \sim \mathrm{Geo}(\pi_t) \\ \left({}^tX_{lk}^{(1,1)}, {}^tX_{lk}^{(1,2)}, {}^tX_{lk}^{(2,1)}, {}^tX_{lk}^{(2,2)}\right) \mid \Sigma \sim \mathrm{Multi}(\Sigma, \mathrm{q}), \end{cases} \tag{5.3}$$

where the probability vector of the multinomial is given by

$$q = \left({}^1p_{l+1,k}^{(1)} \cdot p_t, \quad {}^2p_{l+1,k}^{(1)} \cdot (1-p_t), \quad {}^1p_{l+1,k}^{(2)} \cdot p_t, \quad {}^2p_{l+1,k}^{(2)} \cdot (1-p_t)\right) \tag{5.4}$$

In this notation ${}^tX_{lk}^{(a,b)}$ stands for the number of type b-children in class a of a type t node at level l. Note that the probabilities in q indeed add up to 1. We also want to mention that the quantities ${}^tX_{lk}^{(a,b)}$ are dependent. Nevertheless, as this example shows, we will still be able to completely describe the distribution of the conditioned tree.

Remark 5.1. Suppose we replace the geometric distribution by a general offspring distribution on pairs $(n_m, n-n_m)$, thus introducing dependence between types of children. Then the joint distributions (5.3) change in the sense that, conditioned on the pair of numbers, each type would have its own multinomial.

From the probability distributions above, it follows that the probability that a mutant (type 1) on level l does not generate a mutant on level k (*i.e.* belongs to class 2) satisfies

$$\begin{aligned} {}^1p^{(2)}_{l,k} &= \sum_{n=0}^{\infty} \mathbb{P}(\Sigma = n) \cdot \mathbb{P}(\text{all children are in class 2} \mid \Sigma = \text{n}) \\ &= \sum_{n=0}^{\infty} (1-\pi_1)^n \pi_1 \left({}^1p^{(2)}_{l+1,k} \cdot p_1 + {}^2p^{(2)}_{l+1,k} \cdot (1-p_1) \right)^n \\ &= \frac{\pi_1}{1-(1-\pi_1)\left({}^1p^{(2)}_{l+1,k} \cdot p_1 + {}^2p^{(2)}_{l+1,k} \cdot (1-p_1) \right)} \end{aligned}$$

Similarly it follows that

$${}^2p^{(2)}_{l,k} = \frac{\pi_2}{1-(1-\pi_2)\left({}^1p^{(2)}_{l+1,k} \cdot p_2 + {}^2p^{(2)}_{l+1,k} \cdot (1-p_2) \right)}$$

The other probabilities ${}^1p^{(1)}_{l,k}$ and ${}^2p^{(1)}_{l,k}$ follow from the fact that

$${}^1p^{(1)}_{l,k} + {}^1p^{(2)}_{l,k} = 1, \quad {}^2p^{(1)}_{l,k} + {}^2p^{(2)}_{l,k} = 1.$$

The corresponding initial conditions are

$${}^1p^{(1)}_{kk} = {}^2p^{(2)}_{kk} = 1, \qquad {}^1p^{(2)}_{kk} = {}^2p^{(1)}_{kk} = 0.$$

Figure 1 shows the behavior of these probabilities for the following choice of parameters:

$$\pi_1 = \frac{1}{2}, \qquad \pi_2 = \frac{2}{5}, \qquad p_1 = 1, \qquad p_2 = 10^{-9}.$$

This choice means that the expected number of children is 1 for mutants and $\frac{3}{2}$ for non-mutants. Furthermore, mutants can only generate mutants and a child of a non-mutant has a very small probability to be a mutant. The dashed line shows the probability that a tree with mutated root has a mutant on the kth level as a function of k. This is a critical tree with only mutants that eventually goes extinct. The solid line shows the probability that a tree with non-mutated root has a mutant on the kth level. This tree is supercritical, with reproduction rate (very close to) $\frac{3}{2}$. In a tree with Geo(π) offspring, the extinction probability s is the smallest solution of

$$s = \sum_{n=0}^{\infty} (1-\pi)^n \cdot \pi \cdot s^n = \frac{\pi}{1-(1-\pi)s},$$

which gives

$$s = \begin{cases} \dfrac{\pi}{1-\pi} & \text{if } \pi < \dfrac{1}{2}, \\ 1 & \text{if } \pi \geq \dfrac{1}{2}. \end{cases}$$

For $\pi = \frac{2}{5}$, this gives survival probability $1 - s = \frac{1}{3}$, which is given as a dotted line in Figure 1. If the tree does not die out, then eventually there will be mutants almost surely, since the population grows exponentially. This can be clearly seen in the picture: the survival probability is asymptotically approached by the probability to have a mutant on level k. The population is of order 10^9 around generation $\frac{9\log(10)}{\log(3/2)} \approx 51$, which explains the location of the increase of the solid line.

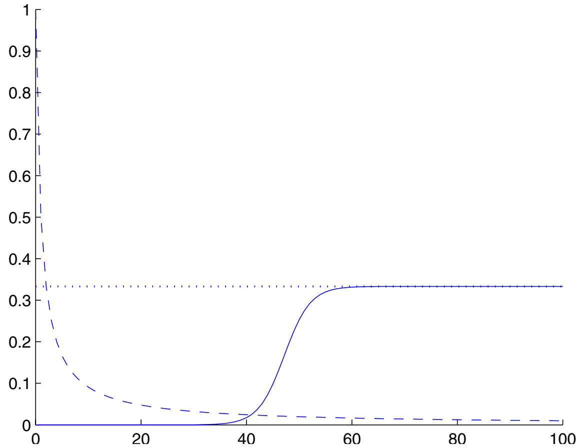

FIGURE 1. *Dashed*: ${}^1p_{0k}^{(1)}$. *Solid*: ${}^2p_{0k}^{(1)}$. *Dotted*: Survival probability of a tree with Geo(2/5) offspring.

5.3. Conditional probability distributions

In this section we demonstrate how to find the full conditional probability distributions, when conditioning on $A_k^{(i)}$. This completely describes a tree conditioned on having (or not having) a mutant on level k. As our general results already showed, we will see that the conditioned two-type Galton−Watson tree in fact is an unconditioned Galton−Watson tree with four types:

$$\begin{aligned} &\text{type } \alpha: \text{ mutant, class 1}\\ &\text{type } \beta: \text{ non-mutant, class 1}\\ &\text{type } \gamma: \text{ mutant, class 2}\\ &\text{type } \delta: \text{ non-mutant, class 2}\end{aligned}$$

Remember that an individual belongs to class 1 if it generates a mutant on level k and to class 2 otherwise. The offspring distribution of a node of one of these types is equal to a conditioned version of the distributions in (5.3). Write (W, X, Y, Z) for $\left({}^tX_{lk}^{(1,1)}, {}^tX_{lk}^{(1,2)}, {}^tX_{lk}^{(2,1)}, {}^tX_{lk}^{(2,2)}\right)$ and (q_W, q_X, q_Y, q_Z) for q as defined in (5.4). Conditioning on $W + X = 0$ gives the offspring distributions of class 2 nodes (*i.e.* type γ and δ):

$$\mathbb{P}\left((w,x,y,z)|W+X=0\right) = \begin{cases} \dfrac{\mathbb{P}\left((0,0,y,z)\right)}{\mathbb{P}(W+X=0)} & \text{if } w=x=0,\\ 0 & \text{otherwise,}\end{cases} \tag{5.5}$$

where we abbreviate $\mathbb{P}\left((W,X,Y,Z)=(w,x,y,z)\right)$ to $\mathbb{P}\left((w,x,y,z)\right)$. The two probabilities in the fraction on the right are given by

$$\begin{aligned}\mathbb{P}(W+X=0) &= \sum_{n=0}^{\infty} \mathbb{P}(\Sigma=n)\cdot\mathbb{P}(W=X=0 \mid \Sigma=n)\\ &= \sum_{n=0}^{\infty}(1-\pi_t)^n\pi_t(q_Y+q_Z)^n\\ &= \frac{\pi_t}{1-(1-\pi_t)(q_Y+q_Z)},\end{aligned} \tag{5.6}$$

and

$$\begin{aligned}\mathbb{P}\left((0,0,y,z)\right) &= \mathbb{P}(\Sigma = y+z)\cdot\mathbb{P}((0,0,y,z) \mid \Sigma = y+z) \\ &= (1-\pi_t)^{y+z}\pi_t\binom{y+z}{y}q_Y^y q_Z^z.\end{aligned}$$

Type γ and δ nodes only can get type γ and δ children. The calculations above give the probability for these types to get $0, 0, y$ and z children of the four types respectively:

$$\left(1-(1-\pi_t)\left(q_Y+q_Z\right)\right)(1-\pi_t)^{y+z}\binom{y+z}{y}q_Y^y q_Z^z, \tag{5.7}$$

where $t = 1$ for a type γ father and $t = 2$ for a type δ father. To find the offspring distributions for type α and β nodes, we have to condition on $W + X \geq 1$. For $w + x \geq 1$, the probability to get w, x, y and z children equals

$$\begin{aligned}\mathbb{P}((w,y,x,z) \mid W+X \geq 1) = \frac{\mathbb{P}\left((w,x,y,z)\right)}{\mathbb{P}(W+X\geq 1)} &= \frac{(1-\pi_t)^{w+x+y+z}\pi_t\binom{w+x+y+z}{w,x,y,z}q_W^w q_X^x q_Y^y q_Z^z}{1-\frac{\pi_t}{1-(1-\pi_t)(q_Y+q_Z)}} \\ &= \frac{(1-(1-\pi_t)(q_Y+q_Z))(1-\pi_t)^{w+x+y+z}\pi_t\binom{w+x+y+z}{w,x,y,z}q_W^w q_X^x q_Y^y q_Z^z}{(1-\pi_t)(q_W+q_X)},\end{aligned} \tag{5.8}$$

where $t = 1$ for an $\alpha-$father and $t = 2$ for a $\beta-$father. The equations (5.7) and (5.8) describe the full offspring distributions for the four types and show that the conditioned tree is again a Galton−Watson tree. The calculations are straightforward (allthough notationally somewhat complicated) and demonstrate that it is actually quite easy to find the distribution of a conditioned tree as soon as the right description in terms of recursive events has been found. Using these distributions, one can easily study or simulate the conditioned tree.

5.4. Expectations in the conditioned tree

The fact that the conditioned tree again is a Galton−Watson tree guarantees some nice properties, for example the expected total progeny can be found by multiplying offspring matrices. As an illustration we will consider a tree with a non-mutated root, conditioned on having a mutant on level k. We will show how to compute the expected number of mutants on each level in such a conditioned tree.

The expected offspring of a node at level $l < k$ is described by a matrix $M_{lk}(i,j)$, where $i, j \in \{\alpha, \beta, \gamma, \delta\}$. In this matrix, the rows give the expected offspring of a type α, β, γ or δ node respectively at level l in the tree. The offspring distributions are given in (5.7) and (5.8) and in Appendix 6 we show how to compute the corresponding expected offspring vectors for all types. Just to give an impression, for $p_2 \downarrow 0$, below are the offspring matrices M_{01}, M_{02} and $\lim_{k\to\infty} M_{0k}$ respectively.

$$M_{01} = \begin{pmatrix} 2&0&0&0\\ 1&0&0&3\\ 0&0&0&0\\ 0&0&0&\frac{3}{2}\end{pmatrix}, \quad M_{02} = \begin{pmatrix} \frac{3}{2}&0&\frac{5}{6}&0\\ \frac{1}{4}&\frac{3}{4}&0&3\\ 0&0&\frac{1}{3}&0\\ 0&0&0&\frac{3}{2}\end{pmatrix}, \quad \lim_{k\to\infty} M_{0k} = \begin{pmatrix} 1&0&2&0\\ 0&\frac{3}{2}&0&\frac{5}{3}\\ 0&0&1&0\\ 0&0&0&\frac{2}{3}\end{pmatrix}.$$

We gave the results for $p_2 \downarrow 0$ to get nice numbers. In our current example, where p_2 is still strictly positive, the numbers are practically the same. Now the conditioned tree we are interested in is just a tree with a root of type β. The types α and γ correspond to mutated individuals. Therefore, the expected number of mutants on level l in the conditioned tree is given by

$$\begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix} M_{0,k}\ldots M_{l-2,k}M_{l-1,k}\begin{pmatrix}1\\0\\1\\0\end{pmatrix}.$$

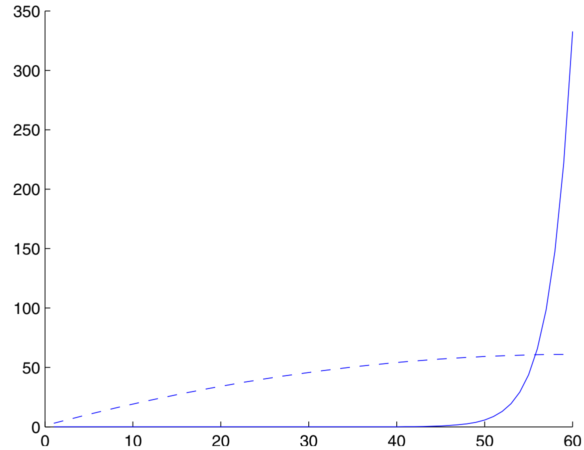

FIGURE 2. Expected number of mutants in each generation of a tree conditioned to have at least one mutant in generation 60. *Solid*: The root is not a mutant. *Dashed*: The root is a mutant.

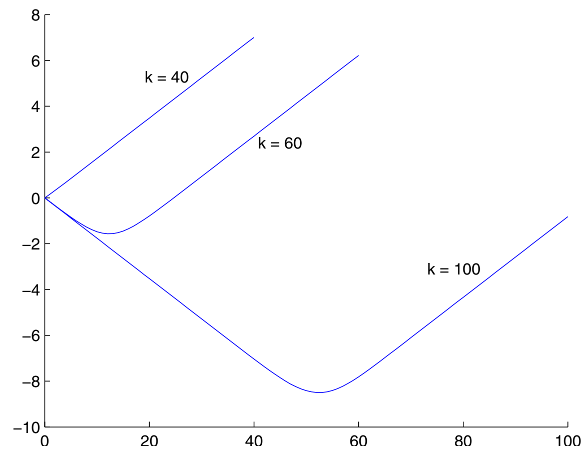

FIGURE 3. Base 10 log of expected size of each generation in a tree conditioned to have no mutant in generation k, for three values of k. In all cases, the root is not a mutant. Note the huge difference in expected population size.

See Figure 2, for a plot of these expected numbers as a function of the generation.

As a second example, we computed the expected total number of individuals in a tree with a non-mutant root and conditioned to have *no* mutant in generation k. If k is small, occurrence of mutants is unlikely anyway, so then the tree just grows exponentially. See Figure 3. If the population grows beyond order 10^9, then the condition has a serious influence on the expected size of the tree. For example, taking $k = 60$, the population size in the unconditioned tree would be of order 10^{10}, but in the conditioned tree it is only of order 10^6. For even larger k the condition to have no mutant in generation k is very restrictive. Apparently, the condition more or less forces the tree to die out exponentially fast. In the unlikely case that it survives for more than 50 generations, it starts to grow exponentially. If an individual is born in the last 40 generations, its progeny is hardly influenced by the condition that no mutant is present in generation 90.

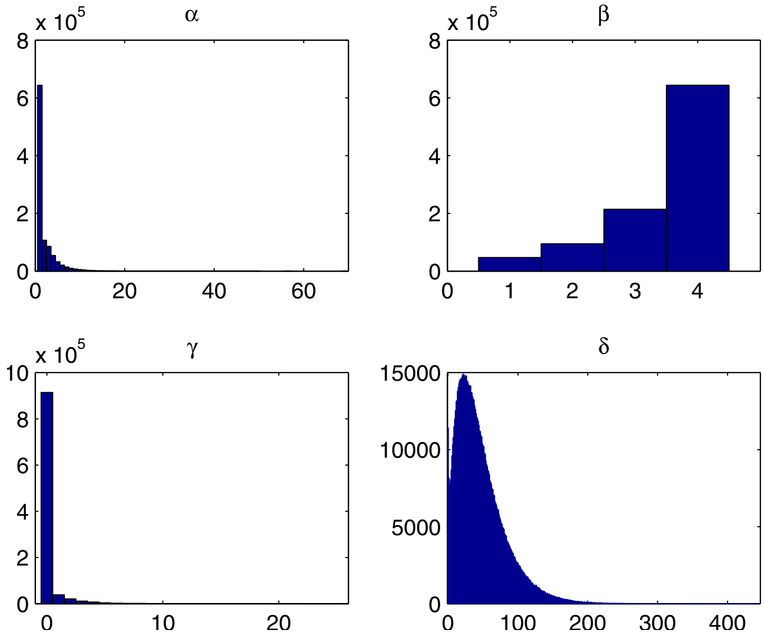

FIGURE 4. Simulated numbers of individuals in a tree of height five, conditioned to have a mutant in the last generation. Mutants: $\alpha + \gamma$.

5.5. Simulation

In this subsection we give an illustration of the simulation of a tree which is conditioned on a rare event. We study six generations of a family, in which at least one individual in the last generation is mutated, while the root is not. Given the root, this event has probability ${}^2p_{05}^{(1)} \approx 1.2 \times 10^{-8}$. Relevant questions for such trees are for example if there are more mutants in the family and in which generation the mutation first appeared.

We generated 10^6 replicates of such a conditioned tree and produced histograms of the total number of each type in the tree, see Figure 4. The majority of the population consists of δ-indivuduals, so non-mutants without a mutant descendant in the last generation. The number of mutants (type α and γ) is most likely to be equal to one, which is also the lowest possible number since there is at least one individual of type α. In our sample the maximum number of mutants in the tree was as high as 84 (69 times α and 25 times γ). From the simulation, we observe that the distribution of the number of mutants looks like a geometric distribution but has heavier tails. The average number of mutants is about 2.41. So mutants are rare, but given there is at least one, there could well be some more.

Also the number of β's is interesting: in the sample it only takes the values one, two, three and four. The minimum is obvious, since the root is of type β. By definition of the types, the root should give birth to at least one α or β child. But more than one is extremely unlikely, since then there would be more than one mutant with a non-mutant father. So from the first α that appears in the tree, there is a line of descent back to the root consisting of the only β's in the tree, which explains that there are at most 4 β's. It also explains the fact that for all trees in our sample, the number of β's was exactly equal to the generation in which the first mutant appeared (average: about 3.4).

It is clear that a simulation like this is very valuable in revealing characteristics of conditioned trees. Simulating this kind of rare events without a direct construction would be very time-consuming. Obtaining sample size 10^6 by just generating unconditioned trees and waiting for trees meeting the condition, would in this particular example require in the order of 10^{14} replications.

6. Scope of our results

We have demonstrated how to condition multi-type Galton−Watson trees on events having some recursive nature. Such trees can be represented as unconditioned multi-type Galton−Watson trees, with possibly more types.

We looked at partitions of the set of trees in which each partition set is defined by some tree property. Suppose this partition is defined for trees of a given height k. Then the children of the root of a tree T of height $k+1$ can be counted to find the number in each partition class and of each type. If the number of types is θ and the number of partition classes is m, then there are $\theta \times m$ possible combinations. Properties of T that are defined in terms of the frequencies of these combinations can be used to define a partition of the set of trees of height $k+1$.

Our procedure applies to all events that can be defined recursively as described above. Examples include presence or absence of a given type at level k, population size of a given type at level k, height of the tree, the tree being a binary tree and the event that every individual of a given type has exactly g grandchildren. Also unions and intersections of such events fit into the framework.

There are no restrictions on the offspring distribution of the conditioned tree. For each of the types, it can be any discrete probability distribution on $\mathbb{N}^\theta$. The offspring distributions in the representation as unconditioned Galton−Watson tree will just be type- and level-dependent conditioned versions of the original offspring distributions, as is also demonstrated in our example in Section 5. In particular cases it might be hard to find exact expressions for the new offspring distributions, but in any case our description allows for efficient simulation. This is especially useful to study trees conditioned on rare events, because our explicit construction procedure allows to directly generate such a tree.

Appendix A. Calculation of conditional expectations

In this appendix we consider the joint distributions

$$\begin{cases} \varSigma \sim \mathrm{Geo}(\pi) \\ (W,X,Y,Z) \mid \varSigma \sim \mathrm{Multi}(\Sigma, (\mathrm{q_W}, \mathrm{q_X}, \mathrm{q_Y}, \mathrm{q_Z})). \end{cases} \tag{A.1}$$

We will calculate all conditional expectations when conditioning on $W+X=0$ or its complement $W+X \geq 1$. First we list (without proof) two useful identities:

$$\sum_{k=0}^{\infty} k \cdot p^k = \frac{p}{(1-p)^2} \qquad \text{for } |p|<1, \tag{A.2}$$

$$\sum_{k=0}^{\infty} (1-p)^n p^k \binom{n+k}{n} = \frac{1}{1-p} \qquad \text{for } |p|<1, n \in \mathbb{N}. \tag{A.3}$$

The unconditional expectations are straightforward:

$$\mathbb{E}[\varSigma] = \sum_{k=0}^{\infty} k(1-\pi)^k \pi = \frac{1-\pi}{\pi}, \tag{A.4}$$

$$\mathbb{E}[A] = \mathbb{E}[\mathbb{E}[A \mid \varSigma]] = \mathbb{E}[q_A \varSigma] = \frac{q_A(1-\pi)}{\pi}, \tag{A.5}$$

where $A \in \{W, X, Y, Z\}$. Next we condition on $W + X = 0$. Clearly, under this condition, W and X have expectation 0. So we proceed by computing the conditional expectation of Y. First we find $\mathbb{P}(Y = y \mid W = X = 0)$ using (5.7) and (A.3):

$$\begin{aligned}
\mathbb{P}(Y = y \mid W, X = 0) &= \sum_{z=0}^{\infty} \mathbb{P}(Y = y, Z = z \mid W, X = 0) \\
&= \sum_{z=0}^{\infty} \left(1 - (1-\pi)\left(q_Y + q_Z\right)\right)(1-\pi)^{y+z}\binom{y+z}{y} q_Y^y q_Z^z \\
&= \left(1 - (1-\pi)\left(q_Y + q_Z\right)\right)\left((1-\pi)q_Y\right)^y \sum_{z=0}^{\infty} \left((1-\pi)q_Z\right)^z \binom{y+z}{y} \\
&= \frac{1 - (1-\pi)\left(q_Y + q_Z\right)}{1 - (1-\pi)q_Z} \cdot \left(\frac{(1-\pi)q_Y}{1-(1-\pi)q_Z}\right)^y.
\end{aligned}$$

Now we apply (A.2) to obtain the conditional expectation:

$$\begin{aligned}
\mathbb{E}[Y \mid W + X = 0] &= \sum_{y=0}^{\infty} y \cdot \mathbb{P}(Y = y \mid W = X = 0) \\
&= \frac{1 - (1-\pi)(q_Y + q_Z)}{1 - (1-\pi)q_Z} \sum_{y=0}^{\infty} y \cdot \left(\frac{(1-\pi)q_Y}{1-(1-\pi)q_Z}\right)^y \\
&= \frac{1 - (1-\pi)(q_Y + q_Z)}{1 - (1-\pi)q_Z} \cdot \frac{(1-\pi)q_Y}{1-(1-\pi)q_Z} \cdot \left(1 - \frac{(1-\pi)q_Y}{1-(1-\pi)q_Z}\right)^{-2} \\
&= \frac{(1-\pi)q_Y}{1 - (1-\pi)(q_Y + q_Z)}. \qquad \text{(A.6)}
\end{aligned}$$

Analogously, for the expectation of Z conditioned on $W + X = 0$, we find

$$\mathbb{E}[Z \mid W + X = 0] = \frac{(1-\pi)q_Z}{1 - (1-\pi)(q_Y + q_Z)}. \qquad \text{(A.7)}$$

So far, we collected all expectations when conditioning on $W + X = 0$. To find the expectations under the condition $W + X \geq 1$, we will use that

$$\mathbb{E}[A] = \mathbb{E}[A \mid W + X \geq 1] \cdot \mathbb{P}(W + X \geq 1) + \mathbb{E}[A \mid W + X = 0] \cdot \mathbb{P}(W + X = 0), \qquad \text{(A.8)}$$

for $A \in \{W, X, Y, Z\}$. Using (5.6), this leads to

$$\begin{aligned}
\mathbb{E}[W \mid W + X \geq 1] &= \frac{\mathbb{E}[W] - \mathbb{E}[W \mid W + X = 0] \cdot \mathbb{P}(W + X = 0)}{\mathbb{P}(W + X \geq 1)} \\
&= \frac{\mathbb{E}[W]}{1 - \mathbb{P}(W + X = 0)} \\
&= \frac{q_W(1-\pi)}{\pi}\left(1 - \frac{\pi}{1 - (1-\pi)(q_Y + q_Z)}\right)^{-1} \\
&= \frac{q_W}{q_W + q_X} \cdot \frac{1 - (1-\pi)(q_Y + q_Z)}{\pi}. \qquad \text{(A.9)}
\end{aligned}$$

And in a similar way

$$\mathbb{E}[X \mid W + X \geq 1] = \frac{q_X}{q_W + q_X} \cdot \frac{1 - (1-\pi)(q_Y + q_Z)}{\pi}. \qquad \text{(A.10)}$$

Finally, we derive the expectations of Y and Z, conditioned on $W + X \geq 1$:

$$\begin{aligned}\mathbb{E}[Y \mid W + X \geq 1] &= \frac{\mathbb{E}[Y] - \mathbb{E}[Y \mid W + X = 0] \cdot \mathbb{P}(W + X = 0)}{\mathbb{P}(W + X \geq 1)} \\ &= \left(\frac{q_Y(1-\pi)}{\pi} - \frac{\pi(1-\pi)q_Y}{(1-(1-\pi)(q_Y+q_Z))^2} \right) \cdot \left(\frac{(1-\pi)(q_W+q_X)}{1-(1-\pi)(q_Y+q_Z)} \right)^{-1} \\ &= \frac{q_Y}{q_W+q_X} \cdot \frac{(1-(1-\pi)(q_Y+q_Z))^2 - \pi^2}{(1-(1-\pi)(q_Y+q_Z)) \cdot \pi}. \end{aligned} \tag{A.11}$$

and analogously it follows that

$$\mathbb{E}[Z \mid W + X \geq 1] = \frac{q_Z}{q_W+q_X} \cdot \frac{(1-(1-\pi)(q_Y+q_Z))^2 - \pi^2}{(1-(1-\pi)(q_Y+q_Z)) \cdot \pi}. \tag{A.12}$$

References

[1] E.A. Cator and H. Don, Constructing and searching conditioned Galton−Watson trees. Preprint arXiv:1412.5890 (2014).

[2] J. Geiger and G. Kersting, The Galton−Watson tree conditioned on its height. In *Probability Theory Mathematical Statistics: Proceedings of the 7th Vilnius Conference 1998*. TEV, Vilnius (1999) 277–286.

[3] J. Geiger, Elementary new proofs of classical limit theorems for Galton−Watson processes. *J. Appl. Probab.* **36** (1999) 301–309.

[4] S. Janson, Simply generated trees, conditioned Galton−Watson trees, random allocations and condensation. *Probab. Surveys* **9** (2012) 103–252.

[5] R. Lyons and Y. Peres, *Probability on Trees and Networks*. Cambridge University Press. Available at http://pages.iu.edu/~rdlyons/ (2016). To appear in *Cambridge Series in Stat. Probab. Math.* (2017).