

RESEARCH ARTICLE

WILEY

Root estimation in Galton–Watson trees

Anna M. Brandenberger | Luc Devroye | Marcel K. Goh

School of Computer Science, McGill University,
Montreal, Canada

Correspondence

Anna Brandenberger, School of Computer
Science, McGill University, Montreal, Canada.
anna.brandenberger@mail.mcgill.ca

Abstract

Given only the free-tree structure of a tree, the root estimation problem asks if one can guess which of the free tree's nodes is the root of the original tree. We determine the maximum-likelihood estimator for the root of a free tree when the underlying tree is a size-conditioned Galton–Watson tree and calculate its probability of being correct.

KEYWORDS

Galton–Watson trees, maximum-likelihood methods, probabilistic analysis, Root estimation

1 | INTRODUCTION

TREES ARE THE MOST ubiquitous nonlinear structures in computer science. There are two different, equally important, notions of a tree. The first is the *unrooted* or *free* tree, which is a connected unlabeled acyclic graph, and the second is the *rooted* tree, in which a single node is distinguished as the root and each edge has a direction from a child to its parent (so all edges point towards the root). Any free tree can be converted into a rooted tree by choosing a root node and setting all of the edge directions accordingly. Likewise, any rooted tree can be seen as a free tree by “forgetting” the directions of the edges. The root estimation problem asks for a method that will recover the root of the underlying rooted tree from the free-tree structure.

Given a free tree of size n , uniformly chosen from among all n -node free trees of a certain family, an easy strategy would be to pick a node uniformly at random; this estimator has a success probability of $1/n$. There are some trees for which this is the optimal estimator, but we will see that in most cases, we will be able to do much better. Of course, it is easy to cook up a family of trees whose structure ensures that the root can be guessed with certainty every time (an obvious example is the complete binary tree on $2^n - 1$ nodes). In many cases we will not be so fortunate, but often there is an estimator that guesses the root with probability asymptotically equal to c/n , where $c > 1$. We solve the root estimation problem on conditional Galton–Watson trees and exploit the connection between these trees and various families in the uniform tree model to give a general approach to root estimation.

1.1 | Background

Root-finding algorithms have been investigated in the literature, mostly for specific classes of trees. The problem was introduced by Haigh [13] in the context of uniform attachment trees, and this work obtains a maximum-likelihood estimate of the root along with the probability of correctness of this estimate as a function of the size of the tree. More recently, Bubeck et al. [4] show that on uniform attachment and preferential attachment trees, one can construct a confidence set of nodes containing the root, where this set has size independent of the number of nodes in the graph. The earlier work by Shah and Zaman [24] in network analysis estimates the source of a rumor in a social network under the susceptible-infected-recovered (SIR) model for viral epidemics, which can be viewed as uniform attachment on a background graph. Their estimation is based on the rumor centrality metric, a notion which is explored in further work reviewed in [26]. For instance, Shah et al. [25] extend their previous result to more generic classes of trees including d -regular trees and geometric trees, and further show that their rumor centrality estimator correctly detects the source in Galton–Watson trees with a strictly positive probability.

In a similar line of work to [4], including some follow up work, authors investigate uniform attachment and preferential attachment trees initialized with an original seed tree [5, 6, 8, 10, 18, 20]. The authors here seek to determine the original seed of a given graph, and study the influence of this seed and its properties on the structure of the graph as it grows. Recent work by Crane et al. [7] considers *shape-exchangable* trees, which encompass the aforementioned models such as uniform attachment, linear preferential attachment, and uniform attachment on a d -regular tree, and expand on the ideas of [4, 18] to provide algorithms for explicitly constructing a confidence set containing the root.

1.2 | The Galton–Watson model

A Galton–Watson tree [2] with offspring distribution ξ is a rooted ordered tree in which every node has i children with probability $p_i = \mathbf{P}\{\xi = i\}$. It is a well-known result that when $\mathbf{E}\{\xi\} \leq 1$, the tree is finite almost surely, except when $p_1 = 1$ and all other p_i are zero. The Galton–Watson branching process was first studied in 1845 by Bienaymé [3], who was interested in the disappearance of family names, and it derives its name from Galton and Watson [12], who studied the same phenomenon in England in 1874. In their model, nodes correspond to individuals in a population and p_i is the probability that an individual passes the family name down to i children. If the process results in a finite Galton–Watson tree, this means the family name goes extinct after some number of generations. We will consider finite, “critical” Galton–Watson trees. These are trees for which $\mathbf{E}\{\xi\} = 1$ and $\mathbf{V}\{\xi\} \in (0, \infty)$; ensuring a nonzero variance rules out the degenerate case $p_1 = 1$.

The Galton–Watson trees that we shall study are conditioned on $|T| = n$, where $|T|$ is the number of nodes in the tree. Conditional Galton–Watson trees were first studied by Kennedy [16] and a key correspondence was found between offspring distributions of conditional Galton–Watson trees and certain families of “simply generated trees” [21]:

1. When $\xi \sim \text{Binomial}(k, 1/k)$, the conditional Galton–Watson tree is a k -ary tree.
2. When $\xi \sim \text{Poisson}(1)$, we have a *Cayley tree*.
3. The distribution $p_0 = p_1 = p_2 = 1/3$ generates a random *Motzkin tree*, in which every node has ≤ 2 children whose order is significant.
4. A Geometric($1/2$) offspring distribution gives rise to a uniformly random *rooted ordered tree*, also known as a *planted plane tree*.

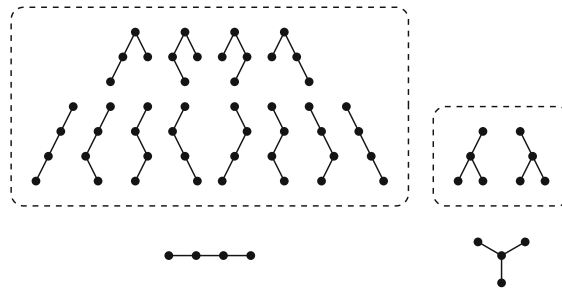


FIGURE 1 The free-tree structure of binary trees with four nodes

This gives us a way to pick uniformly at random from any such family of trees; we simply generate a conditional Galton–Watson tree, which can be done in linear expected time [9]. We will derive a root-estimation strategy for each of the aforementioned families of trees as special cases of our main result.

Our mission can be formalized as follows. Let a conditional Galton–Watson tree with n nodes be given and suppose the directions of the edges are erased, that is, we are shown only the free-tree structure F_n . The goal is to develop a strategy that determines the node with the highest likelihood to have been the root of the original Galton–Watson tree. We would also like to know the probability that we are correct.

1.3 | A concrete example

It is instructive to work through a small toy example using a naïve counting method. Suppose the offspring distribution is

$$p_0 = \frac{1}{4}, \quad p_1 = \frac{1}{2}, \quad p_2 = \frac{1}{4},$$

and all other $p_i = 0$. Conditioning on the number of nodes n generates a binary tree uniformly at random. Figure 1 illustrates the 14 possibilities when $n = 4$.

There are only two possible free trees with four nodes and one is much more likely to arise by this process than the other. If we are shown a path graph, we are best off choosing one of the endpoints, since an endpoint is the root in 8 of the 12 cases and we will guess the correct endpoint with probability $1/2$ (there are two identical endpoints). In this case, the probability of our guessing correctly is $1/3$. When the free tree is the star graph, we should also choose one of the endpoints, since the central node is never the root. Of course, we can still only be correct with probability $1/3$ because there are three identical endpoints.

1.4 | The probabilistic approach

This family of trees illustrated in Figure 1 was small enough to obtain a maximum-likelihood estimator (MLE) by simply counting, but for larger trees and more complex offspring distributions, this will not be feasible. The method we develop will be general and powerful enough to give an MLE for the root on conditional Galton–Watson trees with any offspring distribution p_i and any size n . We will find that the optimal strategy for picking a root is as follows:

1. If $p_i > 0$ and $p_{i-1} = 0$ for some $i \geq 1$ and there exists a node in the free tree with graph-degree i , then only one such node can exist and we select it as our guess. The probability that this node is the root, conditional on its existence in the free tree, is 1.

2. Otherwise, we choose a node uniformly from the nodes of graph-degree i that maximize ip_i/p_{i-1} (note that there could be multiple integers i for which this ratio is maximal).

Note that computing the MLE is computationally easy, and that the probability of correctness in case (ii) can also be explicitly given. We will also analyze the correctness of the MLE as the number of nodes in the tree tends to infinity. Indeed, we show in Theorem 5 that for Galton–Watson trees with offspring distribution satisfying $\sup_{i \geq 1} p_i/p_{i-1} < \infty$ and $0 < \sigma^2 < \infty$, the probability $\mathbf{P}\{C\}$ of the MLE being correct satisfies

$$\lim_{n \rightarrow \infty} n \cdot \mathbf{P}\{C\} = \sup_{i \geq 1} \frac{ip_i}{p_{i-1}}.$$

Thus, for a large class of tree families for which this supremum is finite, for example, k -ary, Cayley and Motzkin trees, the probability of correctness of the MLE decreases linearly with the size of the tree.

2 | AUTOMORPHISMS AND PROBABILITIES

We start off by establishing some terminology and notation. The setup is as follows. We will denote by F_n a free tree on n nodes. This is simply an acyclic graph on n vertices, and is *a priori* unlabeled, though we may choose labels for the nodes when convenient. If a node u is selected and the rest of the tree is allowed to hang from it as if by gravity, then we have the *u -rooted tree*, where the parent of a node is its immediate neighbor in the path towards u .

In the u -rooted tree, we define the *tree-degree* of a node v to be the number of children of v ; this is denoted $\deg_u(v)$. The *graph-degree* of v , written $\deg^*(v)$, is the original degree of v in the free tree F_n . For every node v different from u in the u -rooted tree, we have $\deg_u(v) = \deg^*(v) - 1$ and u is the only node for which the two degrees are equal. The number of nodes of a given tree-degree i in the u -rooted tree is denoted N_i ; the analogous value for the free tree is denoted N_i^* . The tree-degree and graph-degree are, in various places, referred to simply as “degree” (where the context explains which is meant).

An *automorphism* of a free tree F_n is a graph-isomorphism from F_n to itself, that is, a bijection from $V(F)$ to $V(F)$ that preserves the adjacency structure. The group of all such maps is denoted $\text{Aut}(F_n)$. We shall define the *multiplicity* $M(v)$ of a node $v \in F_n$ to be the size of its orbit under the action of $\text{Aut}(F_n)$.

The notion of free-tree automorphisms is used to define the multiplicity, but in fact the number of automorphisms of a *rooted tree* is more pertinent to our problem. Assuming some node u as the root, this is the number of ways that subtrees with the same parent can be permuted among themselves while leaving u firmly planted at the top of the tree. In group-theoretic parlance, this is the stabilizer subgroup $\text{Stab}(u)$ of the automorphism group of F_n .

Every Galton–Watson tree is a rooted ordered tree, and we note that if we reorder the children of any given node, we obtain another Galton–Watson tree with exactly the same tree-degree counts, and thus the same probability of occurrence. Repeat this at every node and let $\text{Perm}(T)$ be the number of possible such reorderings that one can perform on a given rooted ordered tree T ; it is clear that there are

$$\prod_{v \in T} \deg_u(v)! \quad (1)$$

such reorderings. But some permutations leave the tree unchanged (if two subtrees of a given node happened to be indistinguishable, then transposing them does not produce a new tree, in the unordered

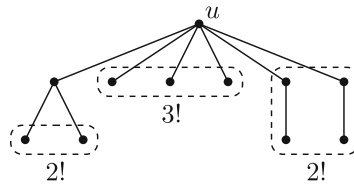


FIGURE 2 An example tree, in which $|\text{Stab}(u)| = 2! \cdot 3! \cdot 2! = 96$

sense). This happens when, at every node, the reordering only sends children to a slot previously occupied by a node in the same orbit of $\text{Stab}(u)$ (see for example Figure 2).

For a tree T with root node u , we let $\text{Perm}(u)$ be the number of distinct unlabeled rooted ordered trees that can be obtained from T by reordering children of nodes:

$$\text{Perm}(u) = \frac{1}{|\text{Stab}(u)|} \prod_v \deg_u(v). \quad (2)$$

Last but not least, we denote by $\text{Prob}(u)$ the Galton–Watson probability of the u -rooted tree. Since each node has a probability p_i of having i children, this is given by

$$\text{Prob}(u) = \prod_{i=0}^{\infty} p_i^{N_i}. \quad (3)$$

Now let F_n be a free tree obtained by removing the parent-child information from a conditional Galton–Watson tree. The probability of a node $u \in F_n$ being the root is the Galton–Watson probability of the u -rooted tree times the number of distinct rooted ordered trees one can obtain via permutations of children. But any node in u 's orbit under $\text{Aut}(F_n)$ could have been the root of an identical tree, so we must divide by $M(u)$. Hence the probability that u is the root is proportional to

$$\frac{\text{Prob}(u)\text{Perm}(u)}{M(u)} = \frac{\text{Prob}(u)}{M(u)|\text{Stab}(u)|} \prod_v \deg_u(v)! = \frac{\text{Prob}(u)}{|\text{Aut}(F_n)|} \prod_v \deg_u(v)!; \quad (4)$$

one must of course introduce a normalizing factor to ensure that this is indeed a valid probability distribution. Note that the last equality above is a consequence of the orbit-stabilizer theorem (see, e.g., [14]). Our maximum-likelihood estimator will thus need to choose a node u that maximizes this probability. Given a Galton–Watson offspring distribution, we will denote by C the event that the MLE is correct for any corresponding free tree of size n , and we seek to determine both $\mathbf{P}\{C\}$, the probability of success of the MLE, and $\mathbf{P}\{C|F_n\}$, the probability of success given a specific free tree F_n . Note that

$$\mathbf{P}\{C\} = \mathbf{E}_{F_n} \{\mathbf{P}\{C|F_n\}\}, \quad (5)$$

where the expected value is taken over all free trees of size n that could arise by the distribution.

3 | ESTIMATING THE ROOT

We are now ready to prove the first significant result. Since $|\text{Aut}(F_n)|$ does not depend on the choice of root, this boils down to maximizing the quantity $\text{Prob}(u) \prod_v \deg_u(v)!$. The following theorem shows

that this can be done knowing only the offspring distribution and the given free-tree structure F_n . To simplify notation, for $i \geq 1$ we define

$$R_i = \frac{ip_i}{p_{i-1}}.$$

Note that throughout the paper, we will assume that $0/0 = 0$, capturing the cases where both p_i and p_{i-1} are equal to zero.

Theorem 1. *Given a free tree F_n corresponding to some Galton–Watson tree with offspring distribution p_i , the strategy to maximize the probability of picking the original root is to select uniformly from the nodes of graph-degree i that maximize R_i , more specifically, defining*

$$\mathcal{M} = \max_{j \geq 1} \{R_j : p_j \neq 0 \text{ and there exists } u \in F_n \text{ such that } \deg^*(u) = j\},$$

the maximum-likelihood estimate for picking the root is to choose a node uniformly from the candidate set

$$\Omega = \{u \in F_n : \deg^*(u) = i, R_i = \mathcal{M}\}.$$

The probability of success of this maximum-likelihood estimator is

1. $P\{C|F_n\} = 1$, if $\mathcal{M} = \infty$;
2. when $\mathcal{M} < \infty$, we have $P\{C|F_n\} = \mathcal{M} / \sum_{v \in F_n} R_{\deg^*(v)}$.

Proof. The probability that any node $u \in F_n$ is the root is given by the formula (4). Thus the goal is to pick a node u that maximizes $\text{Prob}(u) \prod_v \deg_u(v)!$. Suppose we choose some u with $\deg^*(u) = i$, $i \geq 1$.

Note that all the nodes have graph degree one greater than their tree degree, except for the root u , where the two degrees are the same. So for all $j \notin \{i-1, i\}$, $N_j^u = N_{j+1}^*$ and $N_i = N_{i+1}^* + 1$, $N_{i-1}^u = N_i^* - 1$. We proceed, obtaining

$$\begin{aligned} \text{Prob}(u) \prod_v \deg_u(v)! &= \prod_{j=0}^{\infty} p_j^{N_j^u} \prod_v \deg_u(v)! \\ &= \prod_{j=0}^{\infty} p_j^{N_j^u} (j!)^{N_j^u} = \prod_j (j!p_j)^{N_j^u} \\ &= (i!p_i)^{N_i} ((i-1)!p_{i-1})^{N_{i-1}^u} \prod_{j \notin \{i, i-1\}} (j!p_j)^{N_j^u} \\ &= (i!p_i)^{N_{i+1}^*+1} ((i-1)!p_{i-1})^{N_i^*-1} \prod_{j \notin \{i, i-1\}} (j!p_j)^{N_{j+1}^*} \\ &= \frac{ip_i}{p_{i-1}} \prod_{j=0}^{\infty} (j!p_j)^{N_{j+1}^*}. \end{aligned} \tag{6}$$

The infinite product in the last line is the same for all u , so we need only maximize the ratio R_i . Considering the constraint that there must be a node of degree i in F_n , and the fact that there could be multiple degrees that maximize the required ratio (see the limit of k -ary trees as $k \rightarrow \infty$ in the following section), there are two cases for the probability of success of this MLE.

1. $\mathcal{M} = \infty$. This case is deceptively simple. If $\mathcal{M} = \infty$, then there exists $i \geq 1$ such that $p_{i-1} = 0$, $p_i \neq 0$, and there is some $u \in F_n$ with $\deg^*(u) = i$. Suppose, towards a contradiction, that this u were not the root. Then there must be some other node $v \neq u$ that is the root, and the v -tree degree of u would be $\deg_v(u) = \deg^*(u) - 1 = i - 1$. But this is impossible since $p_{i-1} = 0$. So u must be the root. It the only node in the candidate set Ω and our strategy determines the root correctly with probability $P\{C|F_n\} = 1$.
2. $\mathcal{M} < \infty$. In this case, since the probability of any node of degree i being the root is proportional to R_i , normalizing over all nodes in the free tree F_n , we obtain

$$\mathbf{P}\{C|F_n\} = \mathcal{M} / \sum_{v \in F_n} R_{\deg^*(v)}. \quad (7)$$

This is exactly the strategy specified in the theorem statement. ■

4 | APPLICATIONS TO k -ARY AND CAYLEY TREES

Theorem 1 can be applied to any family of trees that arises as a special case of conditional Galton–Watson trees. Without any further machinery, we are now able to give an MLE for conditional Galton–Watson trees of certain offspring distributions. Recall the computation that we performed on 4-node binary trees in the introduction. We were able to show that the best strategy to guess the root was to choose a random endpoint, which would be successful with probability $1/3$. It may come as a surprise that this MLE generalizes to k -ary trees of any size.

4.1 | Rooted k -ary trees

In a rooted k -ary tree, every node can have up to k children and the placement of the children is important; a node has k “slots” in which its children may be placed. As a result, a node can have i children in $\binom{k}{i}$ ways. When $k = 2$ these trees are often called *Catalan trees* because there are $\binom{2n}{n} / (n+1)$ such trees on n nodes.

We can generate an n -node k -ary tree uniformly at random by generating a conditional Galton–Watson tree with a $\text{Binomial}(k, 1/k)$ offspring distribution. Here we have

$$p_i = \binom{k}{i} \left(\frac{1}{k}\right)^i \left(\frac{k-1}{k}\right)^{k-i}$$

for every $i \in \{0, \dots, k\}$, whence

$$R_i = \frac{ip_i}{p_{i-1}} = i \binom{k}{i} \binom{k}{i-1}^{-1} \frac{1}{k} \cdot \frac{k}{k-1} = \frac{k-i+1}{k-1}. \quad (8)$$

So, for any free tree F_n , the probability of a given node u of degree $\deg^*(u) = i$ being the root is

$$R_i / \sum_v R_{\deg^*(v)} = \frac{k-i+1}{\sum_v (k - \deg^*(v) + 1)} = \frac{k-i+1}{nk - (2n-2) + n} = \frac{k-i+1}{(k-1)n+2}. \quad (9)$$

Following the MLE strategy, we pick uniformly at random out of the nodes in the free tree with degree $i = 1$ (of which at least one is guaranteed to exist). Note that this expression is independent of the shape of the free tree F_n , so the probability of success of the MLE is

$$\mathbf{P}\{C\} = \mathbf{P}\{C|F_n\} = \frac{k}{(k-1)n+2}. \quad (10)$$

4.2 | Cayley trees

From the formula (10), one can see that for random k -ary trees, our advantage decreases as k gets large. Indeed, taking the limit as $k \rightarrow \infty$, the Binomial($k, 1/k$) distributions approach a Poisson(1) distribution, with $p_i = (e \cdot i!)^{-1}$. This generates the family of *Cayley trees*, and in this case,

$$\frac{ip_i}{p_{i-1}} = \frac{i \cdot e \cdot (i-1)!}{e \cdot i!} = 1, \quad (11)$$

so every node is equally likely to be the root. Here there is no better strategy than picking uniformly from all nodes in the tree and the success probability is $1/n$.

In both of these cases, $\mathbf{P}\{C|F_n\}$ only depends on n , and we thus have $\mathbf{P}\{C\} = \mathbf{P}\{C|F_n\}$, lending to easy analysis of the MLE. This will not be true in all cases, so in the remainder of this paper, we will upgrade the probabilistic technology in our arsenal before reframing the maximum-likelihood estimator and its probability of correctness $\mathbf{P}\{C\}$ for more complex offspring distributions.

5 | TOOLS FOR COMPUTATION ON THE CONDITIONAL GALTON-WATSON TREE

We would like to be able to analyze the unconditional correctness of the MLE $\mathbf{P}\{C\}$ on a Galton–Watson tree with offspring probability p_i . In general, $\mathbf{P}\{C|F_n\}$ is a random variable that depends on the free-tree structure of the Galton–Watson tree, and we will need certain results from the theory of branching processes. This section contains a potpourri of lemmas and small results that will be useful in the upcoming sections and examples. The casual reader may choose to skim through them in anticipation of the main theorems of the next section, returning to enjoy the proofs after seeing the lemmas used in action.

5.1 | Events on the conditional tree

Let B be some event concerning an unconditional Galton–Watson tree T with offspring distribution ξ . We would like to establish useful tools for working with

$$\mathbf{P}\{B|T = n\},$$

using the random walk representation of conditional Galton–Watson trees. First, suppose that we number the nodes in T (in depth-first preorder, say). Each node i has degree ξ_i and if $\xi_1, \xi_2, \dots$ are independent and all distributed as ξ , then we have

$$\begin{aligned} |T| &= \min\{t > 0 : 1 + (\xi_1 - 1) + \dots + (\xi_t - 1) = 0\} \\ &= \min\left\{t > 0 : \sum_{i=1}^t \xi_i = t - 1\right\}. \end{aligned} \quad (12)$$

Defining two events

$$A^* = \left\{1 + \sum_{i=1}^t (\xi_i - 1) > 0 \text{ for all } t < n, \sum_{i=1}^n \xi_i = n - 1\right\}$$

and

$$A = \left\{ \sum_{i=1}^n \xi_i = n-1 \right\}, \quad (13)$$

we have, by Dwass's cycle lemma [11],

$$\mathbf{P}\{|T| = n\} = \mathbf{P}\{A^*\} = \frac{1}{n} \mathbf{P}\{A\}. \quad (14)$$

Now, B is an event on T , and is thus determined by $\xi_1, \dots, \xi_n$. If we assume rotation invariance (B remains true if applied to $\xi_i, \xi_{i+1}, \dots, \xi_n, \xi_{n+1}, \dots, \xi_{i-1}$ for all i), then we obtain, by another use of the cycle lemma,

$$\begin{aligned} \mathbf{P}\{B||T| = n\} &= \frac{\mathbf{P}\{B \cap |T| = n\}}{\mathbf{P}\{|T| = n\}} = \frac{\mathbf{P}\{B \cap A^*\}}{\mathbf{P}\{A^*\}} = \frac{\mathbf{P}\{B \cap A\}/n}{\mathbf{P}\{A\}/n} \\ &= \frac{\mathbf{P}\{B \cap A\}}{\mathbf{P}\{A\}} = \mathbf{P}\{B|A\}. \end{aligned} \quad (15)$$

This matters because one can study B by simply looking at sequences of i.i.d. random variables and without having to worry about trees.

5.2 | Sums of independent random variables

We will need two lemmas regarding the sums of random variables; these are well-known and will be given without proof. Let the *period* of a random variable ξ be the greatest common divisor of all the i 's for which $\mathbf{P}\{\xi = i\} > 0$. The first of these lemmas is due to Rogozin [23] and the statement as well as its proof can be found in [22].

Lemma A (Rogozin [23]). *If $X_1, \dots, X_n$ are i.i.d. random variables and*

$$p = \sup_x \mathbf{P}\{X_1 = x\},$$

then

$$\sup_x \mathbf{P}\{X_1 + \dots + X_n = x\} \leq \frac{\alpha}{\sqrt{n(1-p)}} \quad (16)$$

for a universal constant α .

The following lemma regards sums of i.i.d. random variables (e.g., as present in our event A) and is due to Kolchin [19].

Lemma B (Kolchin [19]). *Let $\xi_1, \dots, \xi_n$ be i.i.d. random variables on $[0, \infty)$ of mean 1 and variance $\sigma^2 > 0$. Let the period of ξ_1 be*

$$h = \gcd\{i \geq 1 : p_i > 0\}$$

and let X be the set of all integers x such that $(n+x) \bmod h = 0$. Then

$$\sup_{x \in X} \left| \sqrt{n} \mathbf{P}\{\xi_1 + \dots + \xi_n = n+x\} - \frac{h}{\sigma \sqrt{2\pi}} e^{-x^2/2n\sigma^2} \right| \rightarrow 0 \quad (17)$$

as $n \rightarrow \infty$. If $(n+x) \bmod h \neq 0$, then $\mathbf{P}\{\xi_1 + \dots + \xi_n = n+x\} = 0$.

5.3 | The number of nodes of degree i

Recall that we write N_i to indicate the number of nodes of tree-degree i in a Galton–Watson tree. We will show a result that as n gets large, the proportion of nodes in the tree of degree i approaches p_i . The following lemma is due to Aldous [1] and Janson [15].

Lemma C (Aldous [1] and Janson [15]). *Let T_n be a conditional Galton–Watson tree with offspring distribution ξ satisfying $0 < \sigma^2 < \infty$, and let*

$$N_i = \sum_{k=1}^n \mathbf{1}_{[\xi_k=i]}$$

be the number of nodes of degree i in T_n . For any i , $N_i/n \rightarrow p_i$ in probability as $n \rightarrow \infty$.

5.4 | The maximal degree

Another important random variable is the maximal degree M_n of T_n . Because this is rotation-invariant, one can study M_n just as one studies the maximum of independent random variables.

Lemma 2. *Let T be a conditional Galton–Watson tree of size n with offspring distribution ξ whose variance σ^2 satisfies $0 < \sigma^2 < \infty$ and let*

$$M_n = \max_{1 \leq i \leq n} \xi_i$$

be the maximal degree among all the nodes in T . Fix an integer x . Letting $o(1)$ stand for any quantity that tends to 0 as $n \rightarrow \infty$ independent of x , we have

$$\mathbf{P}\{M_n \geq x | |T| = n\} \leq (1 + o(1)) n \mathbf{P}\{\xi \geq x\} \quad (18)$$

and

$$\mathbf{P}\{M_n \leq x | |T| = n\} \leq (\beta + o(1)) \exp(-n \mathbf{P}\{\xi > x\}), \quad (19)$$

for a universal constant β .

Note that if we have a sequence of n i.i.d. random variables ξ_i , the same bounds can be derived, without the $(1 + o(1))$ and $(\beta + o(1))$ factors. This lemma shows that asymptotically, nothing is lost by conditioning on the size of a Galton–Watson tree.

Proof. We begin by expanding and applying the union bound, with A being the event as in (13), obtaining

$$\begin{aligned} \mathbf{P}\{M_n \geq x | |T| = n\} &= \mathbf{P}\{M_n \geq x, A\} / \mathbf{P}\{A\} \\ &\leq n \mathbf{P}\left\{\xi_i \geq x, \sum_{i=1}^n \xi_i = n-1\right\} / \mathbf{P}\{A\} \\ &= n \sum_{j=x}^{\infty} \left(\mathbf{P}\left\{\xi_1 = j, \sum_{i=2}^n \xi_i = n-1-j\right\} / \mathbf{P}\{A\} \right). \end{aligned} \quad (20)$$

Let h be the period of ξ_1 . By Lemma B, we can proceed as follows:

$$\begin{aligned}
 \mathbf{P}\{M_n \geq x | |T| = n\} &\leq n \sum_{j=x}^{\infty} \left(p_j \mathbf{P} \left\{ \sum_{i=2}^n \xi_i = n-1-j \right\} / \frac{h(1+o(1))}{\sigma \sqrt{2\pi n}} \right) \\
 &= n \sum_{j=x}^{\infty} \left(p_j \frac{h e^{-j^2/2\sigma^2(n-1)} + o(1)}{\sigma \sqrt{2\pi(n-1)}} / \frac{h(1+o(1))}{\sigma \sqrt{2\pi n}} \right) \\
 &\leq n \sqrt{\frac{n}{n-1}} \left(\sum_{j \geq x} p_j \right) (1+o(1)) \\
 &\leq (1+o(1)) n \mathbf{P}\{\xi \geq x\}.
 \end{aligned} \tag{21}$$

Next we tackle the lower bound, by an independence argument. First, we expand

$$\begin{aligned}
 \mathbf{P}\{M_n \leq x | |T| = n\} &= \frac{\mathbf{P}\{M_n \leq x, A\}}{\mathbf{P}\{A\}} \\
 &= \mathbf{P}\{M_n \leq x\} \frac{\mathbf{P}\{A | M_n \leq x\}}{\mathbf{P}\{A\}}.
 \end{aligned}$$

Well, $\mathbf{P}\{A | M_n \leq x\} = \mathbf{P}\{\sum_{i=1}^n \xi_i^* = n-1\}$, where $\xi_1^*, \dots, \xi_n^*$ are i.i.d. with

$$\mathbf{P}\{\xi_1^* = i\} = \begin{cases} \mathbf{P}\{\xi_1 = i\} / \mathbf{P}\{\xi_1 \leq x\}, & \text{if } i \leq x; \\ 0, & \text{if } i > x. \end{cases}$$

Let $\eta = \min\{i > 0 : p_i > 0\}$. Then, for $x \geq \eta$, we have $p := \max_{i \leq x} p_i / (p_0 + \dots + p_x) < 1$. Therefore, by Lemma A,

$$\mathbf{P} \left\{ \sum_{i=1}^n \xi_i^* = n-1 \right\} \leq \frac{\alpha}{\sqrt{n(1-p)}}, \tag{22}$$

for a general constant α . Putting

$$\beta = \frac{\alpha}{\sqrt{1-p}} \cdot \frac{\sigma \sqrt{2\pi}}{h},$$

we have, for $x \geq \eta$,

$$\begin{aligned}
 \mathbf{P}\{M_n \leq x | |T| = n\} &\leq \mathbf{P}\{M_n \leq x\} \cdot \beta (1+o(1)) \\
 &\sim \beta (\mathbf{P}\{\xi \leq x\})^n \\
 &\leq \beta \exp(-n \mathbf{P}\{\xi \geq x\}),
 \end{aligned} \tag{23}$$

On the other hand, if $x < \eta$, then $\mathbf{P}\{M_n \leq x | |T| = n\} = 0$ since $M_n < \eta$ implies $M_n = 0$ and thus $|T| = 1$, which is impossible for $n > 1$. The above bound therefore still holds. ■

5.5 | Weighted sums

In the derivation of $\mathbf{P}\{C\}$, one encounters the sum

$$\sum_v R_{\deg^*(v)}$$

for a given free tree in the denominator. When the nodes of a conditional Galton–Watson tree are numbered from 1 to n in preorder and each node i produces a number of offspring distributed as ξ_i , this sum is very close to the random variable

$$W_n = \sum_{i=1}^n \frac{1}{p_{\xi_i}} (\xi_i + 1) p_{\xi_i+1} \mathbf{1}_{[p_{\xi_i} \neq 0]}. \quad (24)$$

We give two lemmas that allow us to work with these weighted sums.

Lemma 3. *Consider the random variable W_n with ξ satisfying $0 < \sigma^2 < \infty$ and $\sup_{i \geq 1, i \notin S} p_i/p_{i-1} < \infty$, where $S = \{i \in \mathbb{N} : p_i > 0, p_{i-1} = 0\}$. Defining $\gamma = \sum_{j \notin S} j p_j \leq 1$, we have $W_n/(\gamma n) \rightarrow 1$ in probability as $n \rightarrow \infty$. Note that if $S = \emptyset$, then $W_n/n \rightarrow 1$.*

Proof. Note that

$$\mathbf{E} \left\{ \frac{(\xi + 1) p_{\xi+1}}{p_{\xi}} \mathbf{1}_{[p_{\xi} \neq 0]} \right\} = \sum_{j=0}^{\infty} \frac{p_j}{p_j} (j+1) p_{j+1} \mathbf{1}_{[p_j \neq 0]} = \sum_{j=1}^{\infty} j p_j \mathbf{1}_{[p_{j-1} \neq 0]} = \gamma, \quad (25)$$

so $\mathbf{E}\{W_n\} = \gamma n$. Also,

$$\begin{aligned} \mathbf{E} \left\{ \left(\frac{(\xi + 1) p_{\xi+1}}{p_{\xi}} \mathbf{1}_{[p_{\xi} \neq 0]} \right)^2 \right\} &= \sum_{j=0}^{\infty} \frac{p_j}{p_j^2} (j+1)^2 p_{j+1}^2 \mathbf{1}_{[p_j \neq 0]} \\ &\leq \sup_{j \geq 1, j \notin S} \frac{p_j}{p_{j-1}} \sum_{j=0}^{\infty} (j+1)^2 p_{j+1} \\ &= \sup_{j \notin S} \frac{p_j}{p_{j-1}} (\sigma^2 + \gamma^2). \end{aligned} \quad (26)$$

By Chebyshev's inequality, for any arbitrary $\epsilon > 0$,

$$\mathbf{P} \left\{ \left| \frac{W_n}{\gamma n} - 1 \right| > \epsilon \right\} \leq \frac{\mathbf{V}\{W_n\}}{n^2 \epsilon^2} \leq \frac{(\sigma^2 + \gamma^2) \sup_{j \notin S} p_j/p_{j-1}}{n \epsilon^2}.$$

Therefore, arguing as before and letting A be the event as in (13),

$$\mathbf{P} \left\{ \left| \frac{W_n}{\gamma n} - 1 \right| > \epsilon \mid |T| = n \right\} \leq \frac{\mathbf{P}\{|W_n/\gamma n - 1| > \epsilon\}}{\mathbf{P}\{A\}} = O\left(\frac{1}{\sqrt{n}}\right).$$

We would now like to show that $\mathbf{E}\{\gamma n/W_n \mid |T| = n\} \rightarrow 1$. This does not follow directly from Lemma 3, but we shall squeeze it out by means of some well-known inequalities and a little elbow grease.

Lemma 4. *Under the same assumptions as the previous lemma, we have*

$$\mathbf{E} \left\{ \frac{\gamma n}{W_n} \mid |T| = n \right\} \rightarrow 1 \quad \text{and} \quad \mathbf{E} \left\{ \frac{(\gamma n)^2}{W_n^2} \mid |T| = n \right\} \rightarrow 1 \quad (27)$$

as $n \rightarrow \infty$.

Proof. Let $\epsilon > 0$, and as before let A be the event as in (13). First, we observe that

$$\begin{aligned} \mathbf{E} \left\{ \frac{\gamma n}{W_n} \mid |T| = n \right\} &\geq \frac{\gamma n}{\gamma n(1 + \epsilon)} \cdot \frac{\mathbf{P}\{W_n < \gamma n(1 + \epsilon), A\}}{\mathbf{P}\{A\}} \\ &= \frac{1}{1 + \epsilon} \left(1 - \frac{\mathbf{P}\{W_n \geq \gamma n(1 + \epsilon), A\}}{\mathbf{P}\{A\}} \right) \\ &\geq \frac{1}{1 + \epsilon} - O\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \quad (28)$$

since $W_n/\gamma n \rightarrow 1$ in probability and $\mathbf{P}\{W_n \geq \gamma n(1 + \epsilon)\} = O(1/n)$, by the previous lemma. Similarly we have

$$\begin{aligned} \mathbf{E} \left\{ \frac{(\gamma n)^2}{W_n^2} \mid |T| = n \right\} &\geq \frac{(\gamma n)^2}{(\gamma n)^2(1 + \epsilon)^2} \frac{\mathbf{P}\{W_n \geq \gamma n(1 + \epsilon), A\}}{\mathbf{P}\{A\}} \\ &\geq \frac{1}{(1 + \epsilon)^2} - O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

It remains to show that $\mathbf{E} \left\{ \gamma n/W_n \mid |T| = n \right\} \leq 1 + o(1)$ and similarly for $(\gamma n)^2/W_n^2$. To that end, note that

$$W_n \geq \sum_{i=1}^n \mathbf{1}_{[\xi_i=0]} \cdot \frac{1}{p_0} p_1. \quad (29)$$

Letting $N_0 = \sum_{i=1}^n \mathbf{1}_{[\xi_i=0]}$, we remark that $N_0 \sim \text{Binomial}(n, p_0)$ and apply Hoeffding's bound to obtain, for $\delta < \min\{p_0, 1 - p_0\}$,

$$\mathbf{P}\{|N_0 - np_0| > \delta n\} \leq 2e^{-2n\delta^2}.$$

We choose $\delta = \epsilon/n^{1/\epsilon}$. Then, by rotation-invariance of W_n , we have

$$\mathbf{E} \left\{ \frac{\gamma n}{W_n} \mid |T| = n \right\} = \frac{\mathbf{E}\{(\gamma n/W_n)\mathbf{1}_A\}}{\mathbf{P}\{A\}} \quad (30)$$

and

$$\mathbf{E} \left\{ (\gamma n)^2/W_n^2 \mid |T| = n \right\} = \frac{\mathbf{E}\{((\gamma n)^2/W_n^2)\mathbf{1}_A\}}{\mathbf{P}\{A\}}, \quad (31)$$

where $\mathbf{P}\{A\} = \Theta(1/\sqrt{n})$. Recall that $a_n = \Theta(b_n)$ denotes the existence of constant real numbers $c, d > 0$ such that for large enough n , $a_n \leq cb_n$ and $a_n \geq db_n$. Also,

$$\begin{aligned} \mathbf{E} \left\{ \frac{\gamma n}{W_n} \mathbf{1}_A \right\} &\leq \mathbf{E} \left\{ \frac{\gamma n}{(1-\epsilon)\gamma n} \mathbf{1}_A \right\} + \mathbf{E} \left\{ \frac{\gamma n}{W_n} \mathbf{1}_{[W_n \leq (1-\epsilon)\gamma n]} \mathbf{1}_A \right\} \\ &\leq \frac{1}{1-\epsilon} \mathbf{P}\{A\} + \mathbf{E} \left\{ \frac{p_0}{p_1} \cdot \frac{\gamma n}{N_0} \cdot \mathbf{1}_{[N_0 \leq np_0/2]} \cdot \mathbf{1}_A \right\} \\ &\quad + \mathbf{E} \left\{ \frac{p_0}{p_1} \cdot \frac{2\gamma n}{np_0} \cdot \mathbf{1}_{[W_n \leq (1-\epsilon)\gamma n]} \right\}. \end{aligned} \quad (32)$$

Letting E_1 and E_2 denote the two expectation terms on the right-hand side, we note that since A implies that $N_0 \geq 1$,

$$E_1 \leq \frac{p_0}{p_1} \gamma n \mathbf{P}\{N_0 \leq np_0/2\} \leq \frac{p_0}{p_1} 2n \exp(-2\gamma n(p_0/2)^2).$$

Furthermore,

$$E_2 = \frac{2\gamma}{p_1} \mathbf{P}\{W_n \leq (1-\epsilon)\gamma n\} = O\left(\frac{1}{n}\right)$$

follows from Chebyshev's inequality, just as in the proof of Lemma 3. This implies that

$$\mathbf{E} \left\{ \frac{\gamma n}{W_n} \mid |T| = n \right\} \leq \frac{1}{1-\epsilon} + \frac{O(1/n)}{O(1/\sqrt{n})} = \frac{1+o(1)}{1-\epsilon}, \quad (33)$$

and we are done since ϵ was chosen arbitrarily. For the $(\gamma n)^2/W_n^2$ case, we proceed the same way to obtain

$$\begin{aligned} \mathbf{E} \left\{ \frac{(\gamma n)^2}{W_n^2} \mathbf{1}_A \right\} &\leq \frac{1}{(1-\epsilon)^2} \mathbf{P}\{A\} + \mathbf{E} \left\{ \frac{(\gamma n)^2}{W_n^2} \mathbf{1}_{[W_n \leq (1-\epsilon)(\gamma n)^2]} \mathbf{1}_A \right\} \\ &\leq \frac{1}{(1-\epsilon)^2} \mathbf{P}\{A\} + \mathbf{E} \left\{ \frac{p_0^2}{p_1^2} \cdot \frac{(\gamma n)^2}{N_0^2} \cdot \mathbf{1}_{[N_0 \leq np_0/2]} \cdot \mathbf{1}_A \right\} \\ &\quad + \mathbf{E} \left\{ \frac{p_0^2}{p_1^2} \cdot \frac{4(\gamma n)^2}{n^2 p_0^2} \cdot \mathbf{1}_{[W_n \leq (1-\epsilon)\gamma n]} \right\} \\ &\leq \frac{1}{(1-\epsilon)^2} \mathbf{P}\{A\} + \frac{p_0^2}{p_1^2} (\gamma n)^2 \mathbf{P}\{N_0 \leq np_0/2\} + \frac{4\gamma^2}{p_1^2} \mathbf{P}\{W_n \leq (1-\epsilon)\gamma n\}, \end{aligned}$$

completing the proof in the same manner. ■

6 | PROBABILITY OF CORRECTNESS OF THE MAXIMUM-LIKELIHOOD ESTIMATOR

We begin by setting up a few definitions to better deal with the two cases mentioned in Theorem 1 in the large n limit. Using this notation, we reformulate our maximum-likelihood estimator for the root, and compute its expected probability of correctness $\mathbf{P}\{C\}$.

Let an offspring distribution be fixed. If $p_i > 0$ and $p_{i-1} = 0$ for some positive integer i , we say that i is a *special integer* and we call a node in the free tree with graph degree i a *special node*. Remember

that finding a special node is akin to hitting the jackpot for the MLE. If i is a special integer and some node v in a free tree has graph-degree i , then v is the root with probability 1. We denote the set of all special integers by S . Note that $i = 1$ is never special, since $p_0 > 0$. We group all nonspecial integers i into equivalence classes $\{J_k\}_{k \geq 1}$ according to the equivalence

$$i \sim j \quad \text{if and only if} \quad \frac{ip_i}{p_{i-1}} = \frac{jp_j}{p_{j-1}}.$$

As before, we let $R_i = ip_i/p_{i-1}$ but for convenience, we will allow the notation R_{J_k} , which equals R_i for any $i \in J_k$. Lastly, we let N_{J_k} denote the number of nodes in the tree whose graph-degree belongs in the equivalence class J_k ; recalling that N_i^* is the number of nodes with graph-degree i , we have

$$N_{J_k} = \sum_{i \in J_k} N_i^*.$$

6.1 | The maximum-likelihood estimator

With these new definitions, we can formally redescribe the MLE and the probability of correctness. Given a free tree F_n of size n corresponding to a Galton–Watson tree with offspring distribution p_i , we guess the root as follows:

1. Let S_n denote the event that there exists a special node in a given free tree F_n . If S_n occurs, then select this special node. In this case,

$$\mathbf{P}\{C|F_n\}\mathbf{1}_{S_n} = \mathbf{1}_{S_n}.$$

2. Otherwise, let S_n^c denote the complement of S_n which occurs if there are either no special integers in the distribution or no nodes with the corresponding degrees in the free tree. On this event, select a node uniformly at random from the class J_λ , where

$$\lambda = \arg \max_{k \notin S} \{R_{J_k} : N_{J_k} > 0\},$$

where we note that this maximum can be taken since there are at most n nonempty equivalence classes. In this case,

$$\mathbf{P}\{C|F_n\}\mathbf{1}_{S_n^c} = \frac{R_\lambda}{\sum_k N_{J_k} R_{J_k}} \mathbf{1}_{S_n^c}. \quad (34)$$

6.2 | Distributions without special integers

We first consider the well-behaved (and more common) case in which there exist no special integers in the Galton–Watson distribution p_i . The following theorem will require the notion of Kesten's limit tree [17], which we will briefly describe. Recall that we are working with an offspring distribution ξ for which $\mathbf{E}\{\xi\} = \sum_{i \geq 1} ip_i = 1$. So if ζ is the random variable with $\mathbf{P}\{\zeta = i\} = ip_i$ for all $i \geq 1$, then ζ is a valid offspring distribution as well. Kesten's limit tree T_∞ is an infinite tree consisting of a central spine of nodes, one on each level, that each produce ζ children. Nodes that are not on the spine are the root of an unconditional Galton–Watson tree with distribution ξ (each of these is finite with probability 1). Let $\tau(T, h)$ denote the tree T , limited to levels $0, \dots, h$. Kesten's limit tree is important

to us because for all h and all infinite trees t , a Galton–Watson tree T_n conditioned to be of size n converges locally to it in the following sense:

$$\lim_{n \rightarrow \infty} \mathbf{P} \{ \tau(T_n, h) = \tau(t, h) \} = \mathbf{P} \{ \tau(T_\infty, h) = \tau(t, h) \}. \quad (35)$$

Theorem 5. *Given a random free tree of size n corresponding to a Galton–Watson tree with offspring distribution p_i with $0 < \sigma^2 < \infty$ and $\sup_{i \geq 1} p_i/p_{i-1} < \infty$. Then the probability of the MLE being correct satisfies*

$$\lim_{n \rightarrow \infty} n \cdot \mathbf{P}\{C\} = \sup_{i \geq 1} \frac{ip_i}{p_{i-1}}. \quad (36)$$

Note that this could be infinity.

Proof. Let λ indicate the equivalence class chosen by the MLE, as described above. First, we prove the upper bound:

$$\mathbf{P}\{C\} = \mathbf{E} \{ \mathbf{P}\{C|F_n\} \} = \mathbf{E} \left\{ \frac{R_\lambda}{\sum_k N_{J_k} R_{J_k}} \right\} \leq \sup_{i \geq 1} R_i \mathbf{E} \left\{ \frac{1}{\sum_k N_{J_k} R_{J_k}} \right\},$$

where we note that $\sum_k N_{J_k} R_{J_k} = \sum_v R_{\deg^*(v)}$ corresponds, up to a $O(1)$ error, to the random variable W_n from Lemma 4, which gave us that $\mathbf{E}\{n/W_n | |T| = n\} \rightarrow 1$ as $\gamma = 1$. We can thus conclude that

$$\limsup_{n \rightarrow \infty} n \mathbf{P}\{C\} \leq \sup_{i \geq 1} R_i. \quad (37)$$

Before moving to the lower bound, let us first show that for any degree $i \geq 1$ such that $p_i > 0$, as $n \rightarrow \infty$,

$$\mathbf{P}\{N_i^* = 0\} \rightarrow 0.$$

Note that by Lemma C, for any conditional Galton–Watson tree corresponding to the free tree of size n rooted at a node u , for all i , $N_i/n p_i \rightarrow 1$ in probability. Furthermore, since we assumed that our distribution has no special integers, for any degree i such that $p_i > 0$, we also have $p_{i-1} > 0$. This yields, for any $i \geq 1$,

$$\begin{aligned} \mathbf{P}\{N_i^* = 0\} &= \mathbf{P}\{N_i^* = 0, \deg^*(u) \notin \{i, i-1\}\} \\ &\quad + \mathbf{P}\{N_i^* = 0, \deg^*(u) = i-1\} + \mathbf{P}\{N_i^* = 0, \deg^*(u) = i\} \\ &= \mathbf{P}\{N_{i-1}^u = 0 | \deg^*(u) \notin \{i, i-1\}\} \mathbf{P}\{\deg^*(u) \notin \{i, i-1\}\} \\ &\quad + \mathbf{P}\{N_{i-1}^u - 1 = 0 | \deg^*(u) = i-1\} \mathbf{P}\{\deg^*(u) = i-1\}, \end{aligned}$$

which goes to 0. This follows from the fact that, as n gets large and the conditional Galton–Watson tree converges locally to Kesten’s limit tree, $\mathbf{P}\{\deg^*(u) = i\} = ip_i + o(1)$. Note that in the above argument, the random variables N_i^* , N_i^u and $\deg^*(u)$ all depend on n , but we avoid double-indexing for clarity of notation.

Now for the lower bound, we must consider two cases:

1. The supremum is *finite*: $\sup_{i \geq 1} R_i < \infty$.
2. The supremum is *infinite*: $\sup_{i \geq 1} R_i = \infty$.

In case (i), let $\epsilon > 0$. There exists some $j \geq 1$ with $p_j > 0$ such that $R_j \geq (1 - \epsilon) \sup_{i \geq 1} R_i$. We define $R = R_j$.

In case (ii), let $R \in \mathbb{R}$ be an arbitrarily large value. We have $\sup_{i \geq 1} R_i = \infty$, therefore for any choice of R , there must exist some j with $p_j > 0$ such that $R_j \geq R$.

Now, in both cases, define the set of equivalence class indices with a larger ratio:

$$\mathcal{J} = \{\ell : R_{J_\ell} \geq R\}.$$

The probability that the MLE chooses an equivalence class that is not a part of this set is the probability that $\mathcal{J}$ is empty,

$$\mathbf{P}\{\lambda \notin \mathcal{J}\} = \mathbf{P}\left\{\bigcap_{\ell \in \mathcal{J}} N_{J_\ell} = 0\right\} \leq \mathbf{P}\{N_j^* = 0\}, \quad (38)$$

which approaches 0 as $n \rightarrow \infty$.

We can thus bound the probability of success from below by

$$\begin{aligned} \mathbf{P}\{C\} &= \mathbf{E}\{\mathbf{P}\{C|F_n\}\} \\ &\geq \mathbf{E}\left\{\mathbf{1}_{[\lambda \in \mathcal{J}]} \frac{R_{J_\lambda}}{\sum_k N_{J_k} R_{J_k}}\right\} \\ &\geq \frac{R}{n(1+\epsilon)} \mathbf{E}\left\{\mathbf{1}_{[\lambda \in \mathcal{J}]} \mathbf{1}_{\left[\sum_k N_{J_k} R_{J_k} \leq n(1+\epsilon)\right]}\right\} \\ &\geq \frac{R}{n(1+\epsilon)} \left(1 - \mathbf{P}\{\lambda \notin \mathcal{J}\} - \mathbf{P}\left\{\sum_k N_{J_k} R_{J_k} > n(1+\epsilon)\right\}\right). \end{aligned} \quad (39)$$

As $n \rightarrow \infty$, we have that $\mathbf{P}\{\lambda \notin \mathcal{J}\} \rightarrow 0$ and, again noting that $\sum_k N_{J_k} R_{J_k}$ is within $O(1)$ of the random variable $W_n = \sum_v R_{\deg^*(v)}$ defined in Lemma 3, we also have $\mathbf{P}\left\{\sum_k N_{J_k} R_{J_k}/n > 1 + \epsilon\right\} \rightarrow 0$. Thus, in both cases (i) and (ii), the sum of terms in the parentheses approaches 1 as $n \rightarrow \infty$.

In case (i), we had $R \geq (1 - \epsilon) \sup_{i \geq 1} R_i$. Thus, since ϵ was arbitrary,

$$\liminf_{n \rightarrow \infty} n \mathbf{P}\{C\} \sup_{i \geq 1} R_i, \quad (40)$$

and we have equality in the limit.

In case (ii),

$$\liminf_{n \rightarrow \infty} n \mathbf{P}\{C\} \geq R \quad (41)$$

for any arbitrarily large choice of R . We thus have

$$\lim_{n \rightarrow \infty} n \mathbf{P}\{C\} = \infty = \sup_{i \geq 1} R_i, \quad (42)$$

completing case (ii). ■

This theorem applies to any distribution for which if there is a positive integer i without any probability mass, then all integers $j \geq i$ have $p_j = 0$ as well. Most of the important examples we consider satisfy this condition. We claimed earlier that in many cases, the probability of correctness is c/n in

the limit for some constant $c \geq 1$; indeed, Theorem 5 has shown that if there are no special nodes, then $c = \sup_{i \geq 1} R_i$ (when this is finite). In fact, since the only valid offspring distribution with mean 1 and $p_i/p_{i-1} = 1/i$ for all $i \geq 1$ is the Poisson(1) distribution, the only case where $c = 1$ is the family of Cayley trees, which we treated in Section 4. In most other cases, the MLE does better, asymptotically speaking, than choosing uniformly at random.

Although the limit of $n\mathbf{P}\{C\}$ may be infinite, the following lemma shows that it is always $o(n)$ if no special integer is observed. It will also apply to distributions containing special integers. We once again let S_n denote the event that there exists a special node in a given free tree F_n , and let S_n^c denote the complement of this event.

Lemma 6. *Let T be a random free tree of size n corresponding to a Galton–Watson tree with offspring distribution p_i . Let S be the set of special integers of this distribution. If $0 < \sigma^2 < \infty$, $\sup_{i \geq 1, i \notin S} p_i/p_{i-1} < \infty$, then the probability of correctness of the MLE satisfies*

$$\lim_{n \rightarrow \infty} \mathbf{P}\{C \cap S_n^c\} = 0.$$

Note that if there are no special integers in the distribution, this is exactly $\mathbf{P}\{C\}$.

Proof. For a conditional Galton–Watson tree of size n , recall the random variable $M_n = \max_{1 \leq i \leq n} \xi_i$ that we defined in Lemma 2 to describe the maximum degree. Next, we define $\kappa = \sup_{i \geq 1, i \notin S} p_i/p_{i-1} < \infty$. Letting $\lambda \notin S$ be the class chosen by the MLE, we note that the best ratio can be bounded by

$$R_\lambda \leq \kappa(M_n + 1) \leq 2\kappa M_n.$$

As for the sum of ratios over all nodes in the free tree, note that given the event S_n^c and letting the event W_n be as in (24), we have

$$\sum_k N_{J_k} R_{J_k} = W_n + \frac{D p_D}{p_{D-1}} - \frac{(D+1)p_{D+1}}{p_D}, \quad (43)$$

where D is the degree of the root of the tree. Then, since $p_{D+1}/p_D \leq \kappa$,

$$\sum_k N_{J_k} R_{J_k} \geq W_n - (D+1)\kappa.$$

Let E_n be the event that $W_n \geq 2\kappa\sqrt{n}$ and $D+1 \leq \sqrt{n}$. Observe that

$$\begin{aligned} \mathbf{P}\{E_n^c\} &\leq \mathbf{P}\{W_n < 2\kappa\sqrt{n}\} + \mathbf{P}\{D+1 \geq \sqrt{n}\} \\ &\leq 4\kappa^2 n \mathbf{E}\{(1/W_n)^2\} + \frac{\mathbf{E}\{D+1\}}{\sqrt{n}} \\ &= O\left(\frac{1}{n}\right) + \frac{\sigma^2 + 2 + o(1)}{\sqrt{n}}, \end{aligned}$$

where we used the fact that $\mathbf{E}\{1/(W_n)^2 | |T| = n\} = O(1/n^2)$ and that $\mathbf{E}\{D\} = \sigma^2 + 1$ for the Kesten tree, to which the conditional Galton–Watson tree locally converges. When E_n holds, we have $\sum_k N_k R_k \geq W_n - \kappa\sqrt{n} \geq W_n/2$.

By Lemmas 3 and 4, we have $W_n/\gamma n \rightarrow 1$ in probability given $|T| = n$, and $\mathbf{E}\{\gamma n/W_n ||T| = n\} \rightarrow 1$ as n tends to infinity. The probability of correctness of the MLE can thus be bounded by

$$\begin{aligned}
 \mathbf{P}\{C \cap S_n^c\} &\leq \mathbf{P}\{C \cap S_n^c \cap E_n\} + \mathbf{P}\{E_n^c\} = \mathbf{P}\{C \cap S_n^c \cap E_n\} + o(1) \\
 &= \mathbf{E}_{F_n} \left\{ \mathbf{P}\{C|F_n\} \mathbf{1}_{S_n^c \cap E_n} \right\} + o(1) \\
 &= \mathbf{E} \left\{ \frac{R_\lambda}{\sum_k N_{J_k} R_k} \mathbf{1}_{S_n^c \cap C_n} \right\} + o(1) \\
 &\leq \mathbf{E} \left\{ \frac{2\kappa M_n}{W_n/2} \mid |T| = n \right\} + o(1) \\
 &\leq 4\kappa \sqrt{\mathbf{E}\{M_n^2 ||T| = n\} \mathbf{E}\{1/W_n^2 ||T| = n\}} + o(1). \tag{44}
 \end{aligned}$$

To bound $\mathbf{E}\{M_n^2 ||T| = n\}$, let A once again denote the event defined in (13); we have

$$\begin{aligned}
 \mathbf{E}\{M_n^2 ||T| = n\} &= \frac{\mathbf{E}\{M_n^2 \mathbf{1}_A\}}{\mathbf{P}\{A\}} \leq \frac{n^2 \mathbf{P}\{M_n \geq n^{7/8}\} + n^{7/4} \mathbf{P}\{A\}}{\mathbf{P}\{A\}} \\
 &\leq \Theta(n^{5/2}) \mathbf{P}\{M_n \geq n^{7/8}\} + n^{7/4}.
 \end{aligned}$$

We proceed by applying the union bound to obtain

$$\begin{aligned}
 \mathbf{E}\{M_n^2 ||T| = n\} &\leq n\Theta(n^{5/2}) \sum_{i \geq n^{7/8}} p_i + n^{7/4} \\
 &\leq \Theta(n^{7/2}) \sum_{i \geq 1} \frac{i^2 p_i}{n^{7/4}} + n^{7/4} \\
 &= \Theta(n^{7/4}), \tag{45}
 \end{aligned}$$

where the last equality follows from the fact that $\sigma^2 < \infty$. Substituting everything into (44), we have

$$\mathbf{P}\{C \cap S_n^c\} = 2\kappa \sqrt{O(n^{7/4}) O(1/n^2)} = O\left(\frac{1}{n^{1/8}}\right).$$

6.3 | Distributions with special integers

We can now deal with the situation in which the distribution contains one or more special integers. It is clear that the MLE should do no worse here than in the nonspecial case, since there is now the possibility of stumbling upon a node that must be the root.

Theorem 7. Fix a random free tree of size n corresponding to a Galton–Watson tree with offspring distribution p_i . Let S denote the set of special integers and suppose that $S \neq \emptyset$, $0 < \sigma^2 < \infty$, and $\sup_{i \notin S} p_i/p_{i-1} < \infty$. The probability of the MLE being correct satisfies

$$\lim_{n \rightarrow \infty} \mathbf{P}\{C\} = \sum_{i \in S} i p_i + o(1). \tag{46}$$

Proof. The special integers $i \in S$ satisfy $p_i \neq 0$ and $p_{i-1} = 0$. Recall from case (i) of Theorem 5 that if there exists a node in the free tree with some special degree $i \in S$, then there can only be one

such node: $\sum_{i \in S} N_i^* \leq 1$. Thus we can split $\mathbf{P}\{C\}$ into two cases: Let S_n and S_n^c be defined as in the previous lemma. Then

$$\begin{aligned} \mathbf{P}\{C\} &= \mathbf{E}\{\mathbf{P}\{C|F_n\}\} \\ &= \mathbf{E}\{\mathbf{P}\{C|F_n\}\mathbf{1}_{S_n}\} + \mathbf{P}\{C\mathbf{1}_{S_n^c}\} \end{aligned} \quad (47)$$

The first term here is simply $\mathbf{P}\{S_n\}$, since the MLE satisfies $\mathbf{P}\{C|F_n\}\mathbf{1}_{S_n} = \mathbf{1}_{S_n}$. As stated in the proof of Theorem 5, a conditional Galton–Watson tree converges locally to Kesten’s limit tree as $n \rightarrow \infty$. Thus, the existence of a $u \in F_n$ with $\deg^*(u) \in S$ is the event that a random conditional Galton–Watson tree has root of degree $i \in S$, which occurs with probability $\sum_{i \in S} ip_i + o(1)$. Then, noting that $\mathbf{P}\{C \cap S_n^c\} = o(1)$ by Lemma 6, we have

$$\mathbf{P}\{C\} = \sum_{i \in S} ip_i + o(1).$$

Comparing this result with Theorem 5, we see that the MLE fares a lot better when there are special integers in the distribution. When there are no special integers, the product $n\mathbf{P}\{C\}$ approaches $\sup_{i \geq 1} R_i$ (and in many cases this supremum is a constant), but we have now shown that the presence of special integers causes $\mathbf{P}\{C\}$ itself to approach a nonzero constant.

7 | FURTHER EXAMPLES

We are now able to calculate the correctness of the MLE for Galton–Watson trees with much more general offspring distributions. We hope that the examples below will demonstrate the simplicity of our general approach to deriving and analyzing the MLE. A summary of these examples appears in Table 1.

7.1 | Full binary trees

This is an example of a distribution with a special integer. In a full binary tree, a node either has two children or none, so we have $p_0 = p_2 = 1/2$ and 2 is a special integer. If there is only one node, then it

TABLE 1 The probability of correctness of the MLE for some families of trees

Family	Distribution	R_i	$\frac{R_i}{\sum_v R_{\deg(v)}}$	MLE	$\mathbf{P}\{C\}$
k -ary	Binomial($k, 1/k$)	$\frac{1}{k-1}$	$\frac{k-i+1}{(k-1)n+2}$	Leaf	$\frac{k}{(k-1)n+2}$
Cayley	Poisson(1)	0	$1/n$	Choose uniformly	$1/n$
Full binary	Uniform{0, 2}	$\begin{cases} 0, & \text{if } i = 1, 3; \\ \infty, & \text{if } i = 2. \end{cases}$	$\begin{cases} 0, & \text{if } i = 1, 3; \\ \infty, & \text{if } i = 2. \end{cases}$	Degree 2	1
Planted plane	Geometric(1/2)	1/2	$\frac{i}{2(n-1)}$	Maximize degree	$\frac{\mathbf{E}\{M_n\}}{2(n-1)} \sim \frac{\log_2 n}{2n}$
Motzkin	Uniform{0, 1, 2}	$\begin{cases} i, & \text{if } i = 1, 2; \\ 0, & \text{if } i = 3. \end{cases}$	$\begin{cases} (i+o(1))/n, & \text{if } i = 1, 2; \\ 0, & \text{if } i = 3. \end{cases}$	Degree 2	$\frac{2+o(1)}{n}$

is certainly the root. Otherwise, the root has graph-degree 2. As asserted in the previous section, there can only be one node in the free tree with graph-degree 2. In other words, for $n \geq 2$, we are guaranteed to be in case (i) of the MLE and we can choose the root with probability 1.

7.2 | Motzkin trees

These are also known as unary-binary trees, because every node can have either one or two children. Unlike a Catalan tree, a node can have one child in only one way, so these trees arise by the probability distribution $p_0 = p_1 = p_2 = 1/3$. When the tree has $n \geq 2$ nodes, the root has either degree 1 or 2, and we have

$$R_i = \frac{ip_i}{p_{i-1}} = i \quad (48)$$

for $i = 1, 2$. The best strategy is to choose uniformly among all nodes with graph-degree 2, unless there are none, in which case we choose a leaf. By Theorem 5, we conclude that $n\mathbf{P}\{C\}$ approaches 2 as n gets large, so $\mathbf{P}\{C\} \sim 2/n$.

7.3 | Planted plane trees

Also called *rooted ordered trees*, this is the family of trees that can be embedded in the plane in a unique way; reordering the subtrees of a given node produces a different tree even if these subtrees are visually indistinguishable. Random planted plane trees correspond to conditional Galton–Watson trees with a Geometric(1/2) offspring distribution. Thus $p_i = 1/2^{i+1}$ for every i and we have

$$R_i / \sum_v R_{\deg^*(v)} = \frac{i/2}{\sum_v \deg^*(v)/2} = \frac{i}{2(n-1)}. \quad (49)$$

This is the probability that a node with degree i is the root. The optimal strategy here is therefore to pick uniformly at random among the nodes of highest degree.

The maximal degree M_n of T_n is a random variable, but we were able to give upper and lower bounds in Lemma 2. For an upper bound, we have

$$\mathbf{P}\{M_n \geq x\} \leq (1 + o(1))n\mathbf{P}\{\xi \geq x\} \sim n/2^x \quad (50)$$

and this tends to 0 if $x = \log_2 n + \omega(1)$. (The small-omega notation $\omega(1)$ denotes a term a_n such that $a_n \rightarrow \infty$ as $n \rightarrow \infty$.) Likewise, we can derive the lower bound

$$\mathbf{P}\{M_n \leq x\} \leq (\beta + o(1)) \exp(-n\mathbf{P}\{\xi \geq x\}) \sim \beta \exp(-n/2^{x+1}) \quad (51)$$

for the constant β given by Lemma 2 and this goes to 0 provided that $x = \log_2 n - \omega(1)$. In other words,

$$\lim_{n \rightarrow \infty} \mathbf{P}\{M_n \geq \log_2 n + \omega(1) | |T| = n\} = 0$$

and

$$\lim_{n \rightarrow \infty} \mathbf{P}\{M_n \leq \log_2 n - \omega(1) | |T| = n\} = 0,$$

that is, $M_n / \log_2 n \rightarrow 1$ in probability. This means that for a planted plane tree,

$$\mathbf{P}\{C\} = \frac{\mathbf{E}\{M_n\}}{2(n-1)} \sim \frac{\log_2 n}{2n}. \quad (52)$$

7.4 | *Large-tailed distributions

Assume that R_i is strictly increasing as a function of i and that $p_i/p_{i-1} \rightarrow 1$ as $i \rightarrow \infty$. For example, we may consider distributions with a polynomial tail

$$p_i = \frac{\theta}{(i+1)^\alpha},$$

for $i \geq 1$ and $\alpha > 3$. The bound on α ensures that $\sigma^2 < \infty$. Noting that $N_i^*/n \rightarrow p_{i-1}$, we obtain

$$\sum_{i=1}^{\infty} \frac{N_i^* R_i}{n} \rightarrow \sum_{i=1}^{\infty} p_{i-1} \frac{ip_i}{p_{i-1}} = 1$$

in probability, and thus

$$\left| \mathbf{P}\{C|F_n\} - \frac{M}{n} \right| \leq f(M, n) \quad (53)$$

where $f(M, n)/(M/n) \rightarrow 0$ in probability as $n \rightarrow \infty$. Thus we have, in general,

$$\mathbf{P}\{C\} \sim \frac{\mathbf{E}\{M\}}{n}. \quad (54)$$

For $p_i = \theta/(i+1)^\alpha$, we that $\mathbf{E}\{M\} = \Theta(n^{1/(\alpha-1)})$ and so our probability of correctness is $\Theta(n^{-(\alpha-2)/(\alpha-1)})$; varying α produces distributions with a whole range of correctness probabilities.

ACKNOWLEDGEMENTS

This paper was created entirely in social isolation during the 2020 coronavirus pandemic. We would like to thank the people who made Zoom, the people who invented the iPad, and our McGill comrades Konrad Anand, Jad Hamdan, Tyler Kastner, Gavin McCracken, Ndiamé Ndiaye, and Rosie Zhao for their support and valuable technical feedback. We are also grateful to the two anonymous referees for suggesting changes that substantially improved the clarity and readability of the paper.

REFERENCES

1. D. Aldous, *Asymptotic fringe distributions for general families of random trees*, Ann. Appl. Probab. **1** (1991), 228–266.
2. K. Athreya and P. Ney, *Branching processes*, Springer Verlag, Berlin, Germany, 1972.
3. I. J. Bienaymé, *De la loi de multiplication et de la durée des familles*, Soc. Philom. Paris Extraits **5** (1845), 37–39.
4. S. Bubeck, L. Devroye, and G. Lugosi, *Finding Adam in random growing trees*, Random Struct. Algorithms **50** (2017), no. 2, 158–172.
5. S. Bubeck, R. Eldan, E. Mossel, and M. Z. Rácz, *From trees to seeds: On the inference of the seed from large trees in the uniform attachment model*, Bernoulli **23** (2017), no. 4A, 2887–2916.
6. S. Bubeck, E. Mossel, and M. Z. Rácz, *On the influence of the seed graph in the preferential attachment model*, IEEE Trans. Netw. Sci. Eng. **2** (2015), no. 1, 30–39.
7. H. Crane and M. Xu, *Inference on the history of a randomly growing tree*, 2020. arXiv preprint arXiv:2005.08794.

8. N. Curien, T. Duquesne, I. Kortchemski, and I. Manolescu, *Scaling limits and influence of the seed graph in preferential attachment trees*, J. Éc. Polytech. Math. **2** (2015), 1–34.
9. L. Devroye, *Simulating size-constrained Galton–Watson trees*, SIAM J. Comput. **41** (2012), no. 1, 1–11.
10. L. Devroye and T. Reddad, *On the discovery of the seed in uniform attachment trees*, Internet Math. **1** (2019), no. 1, 1–29.
11. M. Dwass, *The total progeny in a branching process*, J. Appl. Probab. **6** (1969), 682–686.
12. F. Galton and H. Watson, *On the probability of extinction of families*, J. Anthropol. Inst. **4** (1874), 138–144.
13. J. Haigh, *The recovery of the root of a tree*, J. Appl. Probab. **7** (1970), no. 1, 79–88.
14. M. Hall Jr., *The theory of groups*, MacMillan Company, New York, NY, 1959.
15. S. Janson, *Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees*, Random Struct. Algorithms **48** (2016), 57–101.
16. D. Kennedy, *The Galton–Watson process conditioned on the total progeny*, J. Appl. Probab. **12** (1975), no. 4, 800–806.
17. H. Kesten, *Subdiffusive behavior of a random walk on a random cluster*, Ann. Inst. Henri Poincaré Probab. Stat. **22** (1986), 425–487.
18. J. Khim and P. L. Loh, *Confidence sets for the source of a diffusion in regular trees*, IEEE Trans. Netw. Sci. Eng. **4** (2016), no. 1, 27–40.
19. V. F. Kolchin, *Random mappings*, Optimisation Software Inc., New York, NY, 1986.
20. G. Lugosi and A. Pereira, *Finding the seed of uniform attachment trees*, Electron. J. Probab. **24** (2019), 1–15.
21. A. Meir and J. Moon, *On the altitude of nodes in random trees*, Can. J. Math. **30** (1978), no. 5, 997–1015.
22. V. Petrov, *Sums of independent random variables*, Springer, New York, NY, 1975.
23. B. A. Rogozin, *On an estimate of the concentration function*, Theory Probab. Appl. **6** (1961), 94–97.
24. D. Shah and T. Zaman, *Rumors in a network: Who's the culprit?* IEEE Trans. Inf. Theory **57** (2011), no. 8, 5163–5181.
25. D. Shah and T. Zaman, *Finding rumor sources on random trees*, Oper. Res. **64** (2016), no. 3, 736–755.
26. S. Shelke and V. Attar, *Source detection of rumor in social network – A review*, Online Soc. Netw. Media **9** (2019), 30–42.

How to cite this article: A. M. Brandenberger, L. Devroye, and M. K. Goh, *Root estimation in Galton–Watson trees*, Random Struct. Algorithms. **61** (2022), 520–542. <https://doi.org/10.1002/rsa.21072>