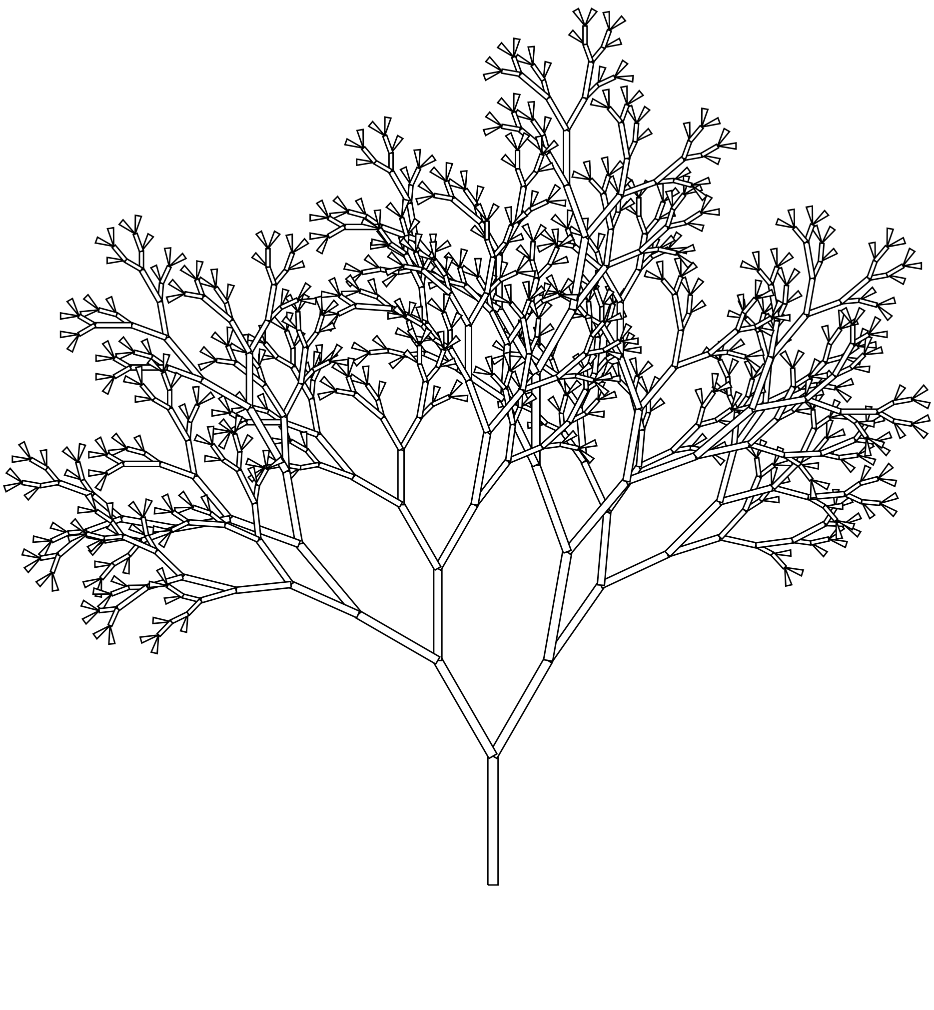
Structural properties of conditional Galton–Watson trees

Marcel Kieren Goh

Department of Mathematics and Statistics
McGill University, Montréal

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Master of Science, April 2022



ABSTRACT

THIS THESIS presents three results concerning conditional Galton–Watson trees. Each of these results involves, in some way, the structure or shape of the underlying tree.

First we tackle the *root estimation problem* in Galton–Watson trees, whose setup is as follows. A Galton–Watson tree is generated with a known offspring distribution, conditioned on its having n nodes. Next, the parent-child directions of the edges are erased, so that only the simple-graph structure of the tree remains. The root estimation problem asks for a maximum-likelihood estimator for the root of the original tree. We give such an estimator and determine its probability of being correct.

Next we study the sizes of orbits of nodes in a conditional Galton–Watson tree under graph automorphism. In the root estimation problem above, it was convenient to define the *multiplicity* of a node to be the size of the node's orbit under (a certain subgroup of) the free tree's automorphism group. The problem of computing the maximum of the multiplicities of the nodes turns out to be rather difficult, so we introduce a finer partition of the nodes and give an asymptotic calculation of the size of the largest equivalence class under this stricter definition of multiplicity.

Finally, we study several parameters of a conditional Galton–Watson tree that are related to independent sets. We give a formula for the independence number of a conditional Galton–Watson tree in terms of the offspring generating function of the distribution. We also analyse the running time of a commonly-used algorithm that computes the independent set. This corresponds to a structural parameter of the tree that we call the *peel number* and we also consider a similar parameter which we call the *leaf height*.

ABRÉGÉ

CETTE THÈSE présente trois résultats concernant les arbres de Galton–Watson. Chacun de ces résultats implique, d'une certaine façon, la structure ou la forme de l'arbre sous-jacent.

Premièrement, on attaque le problème d'estimer la racine d'un arbre de Galton–Watson. Un arbre de Galton–Watson est généré avec une certaine loi de réproduction, conditionné à avoir n nœuds. Ensuite, les directions des arêtes sont effacées pour créer un graphe simple, et on est demandé de deviner le nœud qui a la plus grande chance d'avoir été la racine de l'arbre original. On donne un estimateur du maximum de vraisemblance et calcule la probabilité qu'il est exact.

Ensuite, on étudie les cardinalités des orbites des nœuds dans un arbre de Galton–Watson conditionné sous une action par automorphismes. Dans le problème d'estimation ci-dessus, il était commode de définer la *multiplicité* d'un nœud d'être la cardinalité de son orbite sous (un sous-groupe du) groupe d'automorphismes de l'arbre. Le problème de calculer le maximum des multiplicités des nœuds est difficile, alors on introduit une partition plus fine des nœuds et l'on donne une calculation asymptotique de la cardinalité de la classe d'équivalence plus grande sous cette notion plus stricte de la multiplicité.

Enfin, on étudie quelques paramètres d'un arbre de Galton–Watson qui sont reliés aux stables. En utilisant la fonction génératrice de l'arbre, on donne une formule pour la taille du stable le plus grand dans un arbre de Galton–Watson conditionné. On analyse aussi le nombre d'étapes utilisées par un algorithme bien connu pour trouver un stable maximal dans un arbre de Galton–Watson. Ceci correspond à un paramètre structurel de l'arbre qu'on appelle le *nombre de pelage* et on considère aussi un paramètre similaire qu'on appelle la *hauteur foliaire*.

ACKNOWLEDGEMENTS

I am indebted first and foremost to my supervisor, Luc, for taking a chance on me when I was a third-year undergraduate, despite my unimpressive mark in probability, and introducing me to the joys of research. Luc's boundless curiosity taught me how important it is to keep asking questions and witnessing him use his formidable mathematical toolset has inspired me to dedicate my life to learning how to answer some of them. The sizable stipend Luc allotted to me helped me stay heartily fed throughout my degree, but Luc's generosity extends far beyond financial support and mathematical assistance. When I faced difficulties in my personal life early in my M.Sc., Luc scheduled a special meeting to lend an ear and cheer me up. I will cherish memories of the random conversations we had about anything and everything, as well as the summer afternoons we spent working in the sun at a precariously-propped chalkboard.

A heartfelt thanks goes also to Louigi Addario-Berry, who gave a preliminary draft of this thesis a very thorough reading. He found a number of errors that have since been corrected and offered many suggestions that have greatly improved the readability of the thesis.

This thesis could not exist without the help of my good friends Anna and Rosie. I am grateful to them for always being there to help me with a tricky proof or a confusing definition, even when my frantic wall of messages arrives at two in the morning. Their top-notch mathematical prowess has gotten me out of a tight spot on countless occasions, and their endless enthusiasm and positivity made even the most tedious parts of the research process a good time. A special thanks goes also to Jad, who did not work on these problems but who brought a great deal of good (and often crass) humour to our group and with whom I worked on a different problem that unfortunately did not make it into this thesis. Other friends who made the past year in Montréal a blast include Jaya, Adam, Jason, Mikey, Rachel, Matej, and last but certainly not least, my skrrt skrrt homies Jonah, Camille, and Amanda, whom I met on the road and who enabled me to escape the city for a dozen weekends thereafter.

The compilation of this manuscript was done on an ironing board while quarantining for ten days in the spare room of my grandmother's house, as I contracted the coronavirus while on winter break in Singapore. I would like to thank my mother Joanne and sister Cheyenne for bringing me food three times a day from the outside world, as well as my brother Mikael for also being sick and therefore being a source of human interaction during this time. They have been supportive at various other times in my life as well, as has my father Kien Hwee.

CONTENTS

CHAPTER ONE

PRELIMINARIES

1.1. Introduction

THE MODERN CONCEPT of a probabilistic branching process dates back to 1845, when the French mathematician I.-J. Bienaymé used such a model in his work to study the disappearance of family names [6]. Roughly three decades later, F. Galton and H. W. Watson independently studied the same problem in England [26], and arrived at a model which is essentially identical to Bienaymé's. For this reason, Galton–Watson trees are sometimes called Bienaymé trees or Bienaymé–Galton–Watson trees. See [3] for more on the history of branching processes.

In Galton and Watson's model of family names, nodes correspond to male individuals in a patrilineal population, each of whom passes on the family name to a random number of sons. All nodes are assumed to reproduce independently and according to the same distribution. One can illustrate this process by means of a tree, and if this random tree is finite, then one concludes that the family name goes extinct after some number of generations. Formally, for any nonnegative-integer-valued random variable ξ, a *Galton–Watson* tree is a random tree in which every node has i children independently with probability $p_i = \mathbf{P}\{\xi = i\}$. This random variable ξ is called the *offspring distribution* of the tree. Standard references on the topic include [4], [29], and [40].

When the expected number of children that each node has is at most 1, the family name dies out with probability 1; in other words, the Galton–Watson tree is finite. By conditioning on the event that this finite tree has n nodes, one obtains a distribution on the space of all rooted trees with n nodes, and these *conditional Galton–Watson trees*, which were first studied in 1975 by D. P. Kennedy [35], are the principal object of study in this thesis. Given an offspring distribution ξ, we will usually denote by T the unconditional Galton–Watson tree corresponding to ξ, and we let T_n be the tree T, conditioned on having n nodes.

An important correspondence between conditional Galton–Watson trees and certain families of trees (called *simply generated trees*) was found in 1978 by A. Meir and J. W. Moon [44], This link has given Galton–Watson trees a new lease on life in applications to computer science, since for many important families of simply generated trees, there exists ξ with $\mathbf{E}\{\xi\} = 1$ such that sampling T_n is the same as choosing uniformly from all trees of size n in the family. As can be seen in the following examples, this covers many of the trees that arise in the design and analysis of algorithms.

i) When $\xi \sim$ Binomial$(d, 1/d)$, the conditional Galton–Watson tree is a *d-ary tree* or a *binomial tree*. These are trees in which every node may have up to d children and the placement of the children is important; every node has d "slots" in which children can be placed. As a result, a node can have i children in $\binom{d}{i}$ ways. When $d = 2$, these trees are often called *Catalan trees*, because in this case the number of trees on n nodes is the nth Catalan number $\binom{2n}{n}/(n+1)$.

ii) When $\xi \sim$ Poisson(1), we have a random rooted *Cayley tree*. These can otherwise be sampled by choosing uniformly from all n^{n-2} free trees (connected acyclic graphs) on n labelled nodes, and then picking a random node to be the root. We then hang the tree from the root, ordering the children of each vertex from smallest to largest label. Finally, we remove the labels to obtain a rooted ordered tree.

iii) The distribution $p_0 = p_1 = p_2 = 1/3$ generates a random *Motzkin tree*. These are also sometimes called *unary-binary trees*, since each node can either have one or two children. Motzkin trees are similar to Catalan trees, except that a node can have exactly one child in only one way, whereas in a Catalan tree, every node has two ways of having one child.

iv) A Geometric(1/2) offspring distribution gives rise to a uniformly random rooted ordered tree. Since every Galton–Watson tree is technically a rooted ordered tree, whenever we wish to be clear that the distribution over the space of all rooted ordered trees of size n is uniform, we shall call trees in this family *planted plane trees*.

v) For $t \geq 2$, the distribution with $p_0 = 1 - 1/t$ and $p_t = 1/t$ generates a tree in which every child has either 0 nodes or t nodes. We call these trees *Flajolet t-ary trees*, after the French computer scientist P. Flajolet. When $t = 2$, these are sometimes also called Catalan trees, since there is a bijection between Flajolet 2-ary trees on $2n + 1$ nodes and the Catalan trees we described above on n nodes. To avoid any confusion, we will always call Flajolet 2-ary trees *full binary trees* instead of Catalan trees.

These correspondences give 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 [18].

As n gets large, the conditional Galton–Watson tree approaches a certain infinite tree which we call *Kesten's limit tree*, as it was first described by H. Kesten [37]. It will come up a number of times in our work, so we briefly describe it here. We work with an offspring distribution ξ for which $\mathbf{E}\{\xi\} = \sum_{i \geq 1} i p_i = 1$. So if ζ is the random variable with $\mathbf{P}\{\zeta = i\} = i p_i$ 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 roots of unconditional Galton–Watson trees with distribution ξ (each of these is finite with probability 1). Let $\tau(T_\infty, h)$ denote the tree T_∞, limited to levels $0, \ldots, h$. 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 sense that

$$\lim_{n\to\infty} \mathbf{P}\big\{\tau(T_n,h)=\tau(t,h)\big\} = \mathbf{P}\big\{\tau(T_\infty,h)=\tau(t,h)\big\}. \tag{1}$$

Letting TV denote total variation distance, we can say something stronger. It is known (see [36] and [50]) that if $\mathbf{V}\{\xi\}<\infty$ and $k=o(\sqrt{n})$, then

$$\lim_{n\to\infty} \mathrm{TV}\big(\tau(T_\infty,k),\tau(T_n,k)\big)=0, \tag{2}$$

and we will make use of this in Chapters 3 and 4.

The remainder of this chapter will be dedicated to laying out the preliminary definitions and technical results that will be used in subsequent chapters. We give many of these well-known results without proof, but there are some minor lemmas which, to our knowledge, do not have proofs in the literature (though they too are for the most part essentially well-known in the field). For these lemmas we shall supply a proof. The barrage of technical propositions may be somewhat unmotivated upon first reading, so the reader should consider skipping from the end of this section to the beginning of Chapter 2, returning to this chapter to read proofs of the technical results only after seeing them in action.

The statements and proofs that will comprise the rest of this chapter appear in a dedicated section of a paper jointly written with A. M. Brandenberger and L. Devroye [9]. The second chapter of this thesis discusses the main result of this paper, namely, a solution to the *root estimation problem* in conditional Galton–Watson trees. In particular, we show that, barring any oddities in the distribution that make it possible to find the root with probability 1, the best strategy is to choose uniformly from all the nodes of graph-degree i that maximise the ratio ip_i/p_{i-1}. Of course, there may be multiple possible degrees i to choose from. We also show that in this case we have

$$\lim_{n\to\infty} n\,\mathbf{P}\{\mathcal{C}\} = \sup_{i\ge 1} \frac{ip_i}{p_{i-1}}, \tag{3}$$

where $\mathbf{P}\{\mathcal{C}\}$ denotes the probability that the estimator is correct.

The third chapter is based on joint work with A. M. Brandenberger, L. Devroye, and R. Y. Zhao [10]. It concerns the notion of the *multiplicity* $\mu_{\mathrm{F}}(v)$ of a node v, which was important to the derivation of the maximum likelihood estimator developed in Chapter 2. Let M_n be the maximum value of $\mu_{\mathrm{F}}(v)$ over all nodes v in a conditional Galton–Watson tree T_n. We were not able to find matching upper and lower bounds on M_n, but we define a stricter notion of multiplicity, with corresponding random variable S_n, and study S_n instead. We are able to prove that $S_n=\Omega(\log n)$ asymptotically in probability, and under the further assumption that $\mathbf{E}\{2^\xi\}$ is finite, we have $S_n=O(\log n)$ asymptotically in probability as well. Explicit formulas, depending on the distribution, for the constants in the asymptotic bound are given in both cases.

The fourth and final chapter is unrelated to the second and third, focusing on maximum-size independent sets and the running-time of algorithms used to find them. We give a formula for the independence number (the size of the largest independent set) of a conditional Galton–Watson tree. Letting f denote the reproduction generating function of the offspring distribution ξ, we find that the independence number is in probability asymptotic to qn, where q is the unique solution in $[0,1]$ to the equation $q = f(1-q)$. One of the many algorithms used to find independent sets in trees loops through the tree, peeling away layers of nodes until no nodes are left. We say that a tree has *peel number* ρ if the algorithm takes ρ loops to compute a maximum-size independent set in the tree. We show that the peel number of a conditional Galton–Watson tree is in probability asymptotic to $\log n / \log\bigl(1/f'(1-q)\bigr)$. The notion of a peel number is rather similar in nature to another parameter of a rooted tree, namely, the length of the shortest path from the root to a leaf. This is called the *protection number* in the literature, but we have decided to call it the *leaf height* in our work, as we believe this name to be more illustrative. One can also define the leaf height of any node in the tree by simply considering it as the root of a subtree. When $p_1 > 0$, we show that the maximum leaf height over all nodes in T_n is in probability asymptotic to $\log n / \log(1/p_1)$. If $p_1 = 0$ and κ is the first integer $i > 1$ with $p_i > 0$, then the leaf height is in probability asymptotic to $\log_\kappa \log n$. This last chapter is based on joint work with L. Devroye and R. Y. Zhao [19].

1.2. Basic probabilistic notions

The probability theory we use in this thesis is discrete, down-to-earth, and unpretentious; no notions from measure theory shall explicitly be used. We write $\mathbf{P}\{E\}$ to denote the *probability* of an event E, and write $\mathbf{1}_E$ for the random variable that equals 1 when E is true and 0 when E is false. The expectation of a random variable X is denoted by $\mathbf{E}\{X\}$ and the variance $\mathbf{E}\bigl\{(X - \mathbf{E}\{X\})^2\bigr\}$ is denoted by $\mathbf{V}\{X\}$. When $\mathbf{E}\{X\}$ exists, the fact that it is an "average" of some sort is supported by the *law of large numbers*, which, in its weak form, asserts that for a sequence of independent random variables $X_1, X_2, \ldots$, all distributed as X, we have

$$\lim_{n\to\infty} \mathbf{P}\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbf{E}\{X\} \right| < \epsilon \right\} = 1. \tag{4}$$

for arbitrary $\epsilon > 0$. In general, if for some sequence X_n of random variables there exists a random variable X such that for all $\epsilon > 0$, $\mathbf{P}\bigl\{|X_n - X| > \epsilon\bigr\} \to 0$, then we say that X_n *converges in probability* to X. In most of the cases we encounter, the random variable X is actually a constant, as it is in the statement of the weak law of large numbers. If a nonnegative sequence X_n is such that there exists a function $f(n)$ with

$$\lim_{n\to\infty} \mathbf{P}\bigl\{(1-\epsilon)f(n) < X_n < (1+\epsilon)f(n)\bigr\} = 0 \tag{5}$$

for arbitrary $\epsilon > 0$, then we say that X_n is *in probability asymptotic to* $f(n)$. If for any $\epsilon > 0$ we are only able to show that

$$\lim_{n\to\infty} \mathbf{P}\big\{X_n < (1+\epsilon)\cdot Cf(n)\big\} = 0, \tag{6}$$

for some constant $C > 0$, then we say that $X_n = O\big(f(n)\big)$ *asymptotically in probability*; likewise, if for some constant $c > 0$ we have

$$\lim_{n\to\infty} \mathbf{P}\big\{(1-\epsilon)\cdot cf(n) < X_n\big\} = 0, \tag{7}$$

then $X_n = \Omega\big(f(n)\big)$ asymptotically in probability.

As can be seen from the definitions of convergence of random variables, establishing probabilistic inequalities will be the name of the game in the chapters to come. The first tool for doing so is the indispensable *union bound*

$$\mathbf{P}\Big\{\bigcup_{i=1}^{\infty} E_i\Big\} \le \sum_{i=1}^{\infty} \mathbf{P}\{E_i\}, \tag{8}$$

valid for any countable collection $E_1, E_2, \ldots$ of events. We will also make use of some simple *tail inequalities* for a real-valued random variable X. The first is *Markov's inequality*, which states that for any $\epsilon > 0$,

$$\mathbf{P}\{X \ge \epsilon\} \le \frac{\mathbf{E}\{X\}}{\epsilon}. \tag{9}$$

Replacing X by $X - \mathbf{E}\{X\}$ and ϵ by ϵ^2 yields *Chebyshev's inequality*

$$\mathbf{P}\big\{|X - \mathbf{E}\{X\}| \ge \epsilon\big\} \le \frac{\mathbf{V}\{X\}}{\epsilon^2}. \tag{10}$$

This inequality was stated without proof by I.-J Bienaymé in 1853 [7] and proved by P. Chebyshev in an 1867 paper [14]. Markov's inequality is named after Chebyshev's student A. Markov, but it was also used extensively in the work of Chebyshev himself, and many references (e.g., [49]) attribute both propositions to him.

1.3. The cycle lemma

When it is finite, a Galton–Watson tree, like all rooted ordered trees, is determined entirely by its sequence of the degrees of its nodes, enumerated in depth-first preorder. Since each of these nodes has a degree equal to an independent copy of the random variable ξ, we shall list the degrees $\xi_1, \xi_2, \ldots, \xi_n$. On the other hand, if a sequence of degrees $\xi_1, \xi_2, \ldots$ is such that there exists some $t > 0$ with

$$\sum_{i=1}^{t} \xi_i = t - 1 \qquad \text{and} \qquad \sum_{i=1}^{k} \xi_i \ge k$$

for all $1 \le k < t$, then the tree that the first t degrees define in depth-first preorder is finite, with exactly t nodes (and $\xi_{t+1}, \xi_{t+2}, \ldots$ are meaningless from the point of view of the tree, since the process has already gone extinct).

We now state a result of Dwass [21] that underpins much of our work on the conditional Galton–Watson tree.

Theorem 1.1 (*Dwass' cycle lemma*, 1969). *Let ξ be an offspring distribution with mean 1 and let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables distributed as ξ. Let T be the tree that this sequence defines when considered to be degrees of nodes listed in preorder. The probability that $|T| = n$ equals*

$$\frac{1}{n}\,\mathbf{P}\Big\{\sum_{i=1}^{n} \xi_i = n-1\Big\}. \quad \blacksquare \tag{11}$$

The event that $\sum_{i=1}^{n} \xi_i = n-1$ is thus very important to us and we shall often denote it by A. Given any sequence $\xi_1, \ldots, \xi_n$ that sums to $n-1$, exactly one of the n cyclic permutations of the variables defines a tree (i.e., the kth partial sums are $\geq k$ for $1 \leq k < n$). Thus for any event B on an unconditional Galton–Watson tree T that only concerns the first n degrees $\xi_1, \ldots, \xi_n$ and which holds for any cyclic permutation of the ξ_i, we have

$$\mathbf{P}\{B \mid |T| = n\} = \frac{\mathbf{P}\{B \cap |T| = n\}}{\mathbf{P}\{|T| = n\}} = \frac{\mathbf{P}\{B \cap A\}/n}{\mathbf{P}\{A\}/n} = \frac{\mathbf{P}\{B \cap A\}}{\mathbf{P}\{A\}} = \mathbf{P}\{B \mid A\}. \tag{12}$$

The cycle lemma allows us to study events on a conditional Galton–Watson tree by considering sums of independent random variables. In view of this, we collect here three theorems concerning such sums. When the random variables involved are bounded from above and below, we have the following inequality, due to W. Hoeffding [31].

Theorem 1.2 (*Hoeffding*, 1963). *Let $[a, b] \subseteq \mathbf{R}$ be any closed interval and let $X_1, X_2, \ldots, X_n$ be real-valued random variables with $X_i \in [a, b]$ for all $1 \leq i \leq n$. Then for all $\epsilon > 0$,*

$$\mathbf{P}\bigg\{\Big|\sum_{i=1}^{n}\big(X_i - \mathbf{E}\{X_i\}\big)\Big| > \epsilon\bigg\} \leq 2\exp\bigg(\frac{-2\epsilon^2}{n(b-a)^2}\bigg). \quad \blacksquare \tag{13}$$

When the X_i are all Bernoulli(p) random variables (which take values in the closed interval $[0, 1]$), their sum X is distributed as Binomial(n, p), and Hoeffding's inequality tells us that

$$\mathbf{P}\big\{|X - np| > \epsilon\big\} \leq 2e^{-2\epsilon^2/n} \tag{14}$$

for all $\epsilon > 0$. This result was found earlier by H. Chernoff [16].

Next, we have a theorem due to B. A. Rogozin [47]. The statement and proof of this inequality can be found in [45].

Theorem 1.3 (*Rogozin*, 1961). *Let $X_1, \ldots, X_n$ be i.i.d. random variables and let $s = \sup_x \mathbf{P}\{X_1 = x\}$. There is some universal constant α such that*

$$\sup_x \mathbf{P}\{X_1 + \cdots + X_n = x\} \leq \frac{\alpha}{\sqrt{n(1-s)}}. \quad \blacksquare \tag{15}$$

Let the *period* of a nonnegative-integer-valued random variable X be the greatest common divisor of all the integers i for which $\mathbf{P}\{X = i\} > 0$. The final theorem regarding sums of independent random variables that we list here is due to V. F. Kolchin [38].

Theorem 1.4 (*Kolchin*, 1986). *Let $X_1, \ldots, X_n$ be i.i.d. random variables on $[0, \infty)$ of mean 1 and variance $\sigma^2 > 0$. Let h denote the period of X_1 and let S be the set of all integers x such that $n + x \equiv 0 \pmod h$. Then*

$$\sup_{x \in S} \left| \sqrt{n}\, \mathbf{P}\{X_1 + \cdots + X_n = n + x\} - \frac{h}{\sigma\sqrt{2\pi}} e^{-x^2/(2n\sigma^2)} \right| \to 0 \tag{16}$$

as $n \to \infty$. If $n + x \not\equiv 0 \pmod h$, then $\mathbf{P}\{X_1 + \cdots + X_n = n + x\} = 0$. ▮

In asymptotic terms, Kolchin's estimate tells us that if $\xi_1, \ldots, \xi_n$ are independent copies of a random variable ξ with mean 1 and variance σ^2, then

$$\mathbf{P}\Big\{\sum_{i=1}^{n} \xi_i = n - 1\Big\} = \frac{h\big(1 + o(1)\big)}{\sigma\sqrt{2\pi n}}, \tag{17}$$

where h is the period of ξ. Combining this with Dwass's theorem above, if ξ is the offspring distribution of an unconditional Galton–Watson tree T, then

$$\mathbf{P}\{|T| = n\} = \frac{h\big(1 + o(1)\big)}{\sigma\sqrt{2\pi} n^{3/2}} = \Theta(n^{3/2}). \tag{18}$$

1.4. Degree statistics

We now divert our attention to the distribution of node degrees in a conditional Galton–Watson tree, namely, the number of nodes of a given degree i as well as the maximum degree. The key observation is that since the random variables $\xi_1, \ldots, \xi_n$ are themselves the degrees, both of these statistics stay the same when the ξ_i are permuted. The following lemma is due to D. Aldous [2] and S. Janson [33], but we give an alternative proof using the cycle lemma and Kolchin's estimate.

Lemma 1.5 (*Aldous*, 1991; *Janson*, 2016). *Let T_n be a conditional Galton–Watson tree with offspring distribution ξ satisfying $\mathbf{E}\{\xi\}$ and $0 < \mathbf{V}\{\xi\} < \infty$, and let*

$$N_{i,n} = \sum_{k=1}^{n} \mathbf{1}_{[\xi_k = i]} \tag{19}$$

be the number of nodes of degree i in T_n. For any i, $N_{i,n}/n \to p_i$ in probability as $n \to \infty$. ▮

Proof. Let $\epsilon > 0$ be given. Let A be the event that $\sum_{i=1}^{n} \xi_i = n - 1$ and let B be the event that $|N_{i,n}/n - p_i| > \epsilon$. Note that B is rotation-invariant. So by the cycle lemma, we have

$$\mathbf{P}\{B \mid |T| = n\} = \mathbf{P}\{B \mid A\} = \frac{\mathbf{P}\{B \cap A\}}{\mathbf{P}\{A\}} \le \frac{\mathbf{P}\{B\}}{\mathbf{P}\{A\}}. \tag{20}$$

Next, Kolchin's estimate states that

$$\mathbf{P}\{A\} = \frac{h\bigl(1+o(1)\bigr)}{\sigma\sqrt{2\pi n}}, \tag{21}$$

where h is the period of ξ_1. Also, since $\mathbf{E}\{N_{i,n}\} = np_i$ and

$$\mathbf{V}\{N_{i,n}/n\} = \frac{\mathbf{V}\{\mathbf{1}_{[\xi_1=i]}\}}{n} = \frac{p_i(1-p_i)}{n}, \tag{22}$$

we have, by Chebyshev's inequality,

$$\mathbf{P}\{B\} \le \frac{\mathbf{V}\{N_{i,n}/n\}}{\epsilon^2} \le \frac{p_i(1-p_i)}{n\epsilon^2}, \tag{23}$$

and substituting (21) and (23) into (20) yields

$$\mathbf{P}\{B \mid |T| = n\} \le \frac{p_i(1-p_i)\sigma}{h\bigl(1+o(1)\bigr)\epsilon^2}\sqrt{\frac{2\pi}{n}}, \tag{24}$$

which goes to 0 as $n \to \infty$. ▮

Of course, the result of this lemma is entirely predictable, since if $\xi_1, \ldots, \xi_n$ are independent copies of ξ, then the fraction of the ξ_i that are equal to i converges to p_i by the weak law of large numbers. The next result about the maximum degree is just as predictable, though a bit harder to prove.

Lemma 1.6. *Let T_n be a conditional Galton–Watson tree with offspring distribution ξ with mean 1 and variance σ^2 satisfying $0 < \sigma^2 < \infty$, and let M_n be the maximal degree among all the nodes in T_n. Fix an integer x. Letting $o(1)$ stand for any quantity that tends to 0 as $n \to \infty$ independent of x, we have*

$$\mathbf{P}\{M_n \ge x\} \le \bigl(1+o(1)\bigr)n\,\mathbf{P}\{\xi \ge x\} \tag{25}$$

and

$$\mathbf{P}\{M_n \le x\} \le \bigl(\beta+o(1)\bigr)\exp\bigl(-n\,\mathbf{P}\{\xi > x\}\bigr), \tag{26}$$

for a constant β depending on ξ but not n.

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

Proof. Let $\xi_1, \ldots, \xi_n$ be independent random variables distributed as ξ and let T be the tree built using the ξ_i. Let A be the event that $\sum_{i=1}^n \xi_i = n-1$ and note that

$$\mathbf{P}\{M_n \ge x\} = \mathbf{P}\Bigl\{\max_{1\le i\le n} \xi_i \ge x \mid |T| = n\Bigr\}. \tag{27}$$

Of course, the event that $\max_{1\le i\le n}\xi_i \ge x$ is invariant under rotation, and so is the analogous statement for $\mathbf{P}\{M_n \le x\}$. For the upper bound, we expand using the cycle lemma and apply the union bound, obtaining

$$\begin{aligned}\mathbf{P}\Big\{\max_{1\le i\le n}\xi_i \ge x \mid |T| = n\Big\} &= \frac{\mathbf{P}\big\{\max_{1\le i\le n}\xi_i \ge x,\ A\big\}}{\mathbf{P}\{A\}}\\ &\le \frac{n}{\mathbf{P}\{A\}}\,\mathbf{P}\Big\{\xi_1 \ge x,\ \sum_{i=1}^{n}\xi_i = n-1\Big\}\\ &= \frac{n}{\mathbf{P}\{A\}}\sum_{j=x}^{\infty}\mathbf{P}\{\xi_1 = j\}\,\mathbf{P}\Big\{\sum_{i=2}^{n}\xi_i = n-1-j\Big\}.\end{aligned} \tag{28}$$

Let h be the period of ξ_1. Using Kolchin's estimate twice, we obtain the upper bound

$$\begin{aligned}\mathbf{P}\{M_n \ge x\} &\le \frac{n^{3/2}\sigma\sqrt{2\pi}}{h\big(1+o(1)\big)}\sum_{j=x}^{\infty}p_j\,\frac{he^{-j^2/(2\sigma^2(n-1))}+o(1)}{\sigma\sqrt{2\pi(n-1)}}\\ &\le n\sqrt{\frac{n}{n-1}}\Big(\sum_{j\ge x}p_j\Big)\big(1+o(1)\big)\\ &\le \big(1+o(1)\big)n\,\mathbf{P}\{\xi \ge x\}.\end{aligned} \tag{29}$$

Next we tackle the lower bound, by an independence argument. We have

$$\begin{aligned}\mathbf{P}\{M_n \le x\} &= \frac{\mathbf{P}\big\{\max_{1\le i\le n}\xi_i \le x, A\big\}}{\mathbf{P}\{A\}}\\ &= \mathbf{P}\Big\{\max_{1\le i\le n}\xi_i \le x\Big\}\frac{\mathbf{P}\big\{A \mid \max_{1\le i\le n}\xi_i \le x\big\}}{\mathbf{P}\{A\}}.\end{aligned} \tag{30}$$

Well, $\mathbf{P}\big\{A \mid \max_{1\le i\le n}\xi_i \le x\big\} = \mathbf{P}\{\sum_{i=1}^{n}\xi_i^* = n-1\}$, where $\xi_1^*,\ldots,\xi_n^*$ are i.i.d. with

$$\mathbf{P}\{\xi_1^* = i\} = \begin{cases}\mathbf{P}\{\xi_1 = i\}/\,\mathbf{P}\{\xi_1 \le x\}, & \text{if } i \le x;\\ 0, & \text{if } i > x.\end{cases} \tag{31}$$

Let $\eta = \min\{i > 0 : p_i > 0\}$. Then, for $x \ge \eta$, we let $p = \max_{i\le x} p_i/(p_0+\cdots+p_x)$ and note that $p < 1$. Therefore, by Rogozin's inequality, there is a universal constant α such that

$$\mathbf{P}\Big\{\sum_{i=1}^{n}\xi_i^* = n-1\Big\} \le \frac{\alpha}{\sqrt{n(1-p)}}. \tag{32}$$

Putting

$$\beta = \frac{\alpha}{\sqrt{1-p}}\cdot\frac{\sigma\sqrt{2\pi}}{h}, \tag{33}$$

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

$$\begin{aligned}\mathbf{P}\{M_n \le x\} &\le \mathbf{P}\Big\{\max_{1\le i\le n} \xi_i \le x\Big\} \cdot \beta\big(1+o(1)\big)\\ &\sim \beta\,\mathbf{P}\{\xi \le x\}^n \\ &\le \beta \exp\big(-n\,\mathbf{P}\{\xi > x\}\big),\end{aligned} \tag{34}$$

On the other hand, if $x < \eta$ and $n > 1$, then $\mathbf{P}\{M_n \le x\} = 0$ since $\max_{1\le i\le n} \xi_i < \eta$ implies $\max_{1\le i\le n} \xi_i = 0$ and thus $|T_n| = 1$. The above bound therefore still holds. ▌

1.5. Weighted sums

We close this chapter with an asymptotic analysis of a certain weighted sum. This random variable appears at a crucial point in Chapter 2. Though it can hardly be considered an important proposition in its own right and is used nowhere else in the thesis, we have decided to place its proof here, as, being rather long and technical, it would detract from the flow of the second chapter.

Lemma 1.7. *Let ξ be an offspring distribution with mean 1 and nonzero finite variance. Write $p_i = \mathbf{P}\{\xi = i\}$. Setting $\mathcal{S} = \{i \in \mathbf{N} : p_i > 0, p_{i-1} = 0\}$, assume furthermore that the probabilities p_i that define ξ satisfy $\sup_{i\ge 1, i\notin \mathcal{S}} p_i/p_{i-1} < \infty$. Writing*

$$\gamma = \sum_{j\notin\mathcal{S}} j p_j, \tag{35}$$

the random variable

$$W_n = \sum_{i=1}^{n} \frac{1}{p_{\xi_i}} (\xi_i + 1) p_{\xi_i+1} \mathbf{1}_{[p_{\xi_i} \ne 0]} \tag{36}$$

satisfies

i) *$W_n/(\gamma n) \to 1$ in probability as $n \to \infty$;*

ii) *$\mathbf{E}\{\gamma n/W_n \mid \xi_1 + \cdots + \xi_n = n-1\} \to 1$; and*

iii) *$\mathbf{E}\big\{(\gamma n/W_n)^2 \;\big|\; \xi_1 + \cdots + \xi_n = n-1\big\} \to 1$.*

Note that $\gamma \le 1$ and if $\mathcal{S} = \emptyset$, then $\gamma = \mathbf{E}\{\xi\} = 1$ and in this case part (i) of the lemma says that $W_n/n \to 1$.

Proof. Note that

$$\mathbf{E}\Big\{\frac{(\xi+1)p_{\xi+1}}{p_\xi}\mathbf{1}_{[p_j\ne 0]}\Big\} = \sum_{j=0}^{\infty} \frac{p_j}{p_j}(j+1)p_{j+1}\mathbf{1}_{[p_j\ne 0]} = \sum_{j=1}^{\infty} j p_j \mathbf{1}_{[p_{j-1}\ne 0]} = \gamma, \tag{37}$$

so that $\mathbf{E}\{W_n\} = \gamma n$, and also

$$\mathbf{E}\Big\{\Big(\frac{(\xi+1)p_{\xi+1}}{p_\xi}\mathbf{1}_{[p_j\neq 0]}\Big)^2\Big\} = \sum_{j=0}^{\infty}\frac{p_j}{{p_j}^2}(j+1)^2 p_{j+1}^2\mathbf{1}_{[p_j\neq 0]}$$
$$\leq \Big(\sup_{j\geq 1, j\notin\mathcal{S}}\frac{p_j}{p_{j-1}}\Big)\sum_{j=0}^{\infty}(j+1)^2 p_{j+1} \qquad (38)$$
$$= \sup_{j\notin\mathcal{S}}\frac{p_j}{p_{j-1}}(\sigma^2+1).$$

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

$$\mathbf{P}\Big\{\Big|\frac{W_n}{\gamma n}-1\Big| > \epsilon\Big\} \leq \frac{\mathbf{V}\{W_n\}}{n^2\epsilon^2} \leq \frac{(\sigma^2+1)\sup_{j\notin\mathcal{S}} p_j/p_{j-1}}{n\epsilon^2}. \qquad (39)$$

Therefore, arguing as before and letting A be the event that $\sum_{i=1}^n \xi_i = n-1$, we have

$$\mathbf{P}\Big\{\Big|\frac{W_n}{\gamma n}-1\Big| > \epsilon \;\Big|\; |T| = n\Big\} \leq \frac{\mathbf{P}\big\{|W_n/\gamma n - 1| > \epsilon\big\}}{\mathbf{P}\{A\}} = O\Big(\frac{1}{\sqrt{n}}\Big). \qquad (40)$$

We have thus proved part (i).

Parts (ii) and (iii) take a bit of elbow grease. Strictly speaking, (iii) implies (ii), but for simplicity of presentation, we show how to prove (ii) and then describe where the proof changes for (iii). Let $\epsilon > 0$ once again be arbitrary. First, we observe that

$$\mathbf{E}\Big\{\frac{\gamma n}{W_n}\;\Big|\; |T| = n\Big\} \geq \frac{\gamma n}{\gamma n(1+\epsilon)}\cdot\frac{\mathbf{P}\{W_n < \gamma n(1+\epsilon),\ A\}}{\mathbf{P}\{A\}}$$
$$= \frac{1}{1+\epsilon}\Big(1 - \frac{\mathbf{P}\{W_n \geq \gamma n(1+\epsilon),\ A\}}{\mathbf{P}\{A\}}\Big) \qquad (41)$$
$$\geq \frac{1}{1+\epsilon} - O\Big(\frac{1}{\sqrt{n}}\Big),$$

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

$$\mathbf{E}\left\{\frac{(\gamma n)^2}{{W_n}^2}\;\Big|\; |T| = n\right\} \geq \frac{(\gamma n)^2}{(\gamma n)^2(1+\epsilon)^2}\frac{\mathbf{P}\{W_n \geq \gamma n(1+\epsilon), A\}}{\mathbf{P}\{A\}}$$
$$\geq \frac{1}{(1+\epsilon)^2} - O\Big(\frac{1}{\sqrt{n}}\Big). \qquad (42)$$

It remains to show that $\mathbf{E}\big\{\gamma n/W_n \;\big|\; |T| = n\big\} \leq 1 + o(1)$ as well as the analagous statement for $(\gamma n)^2/{W_n}^2$. To that end, note that

$$W_n \geq \sum_{i=1}^n \mathbf{1}_{[\xi_i = 0]}\cdot\frac{1}{p_0}p_1. \qquad (43)$$

Letting $N_0 = \sum_{i=1}^n \mathbf{1}_{[\xi_i=0]}$ so that $W_n \geq N_0 p_1/p_0$, we observe that $N_0 \sim \text{Binomial}(n, p_0)$ and apply the binomial case of Hoeffding's bound to obtain, for $\delta < \min\{p_0, 1-p_0\}$,

$$\mathbf{P}\big\{|N_0 - np_0| > \delta n\big\} \le 2e^{-2n\delta^2}. \tag{44}$$

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

$$\mathbf{E}\left\{ \frac{\gamma n}{W_n} \;\middle|\; |T| = n \right\} = \frac{\mathbf{E}\big\{(\gamma n/W_n)\,\mathbf{1}_A\big\}}{\mathbf{P}\{A\}} \tag{45}$$

by the cycle lemma and

$$\mathbf{E}\big\{(\gamma n/W_n)^2 \;\big|\; |T| = n\big\} = \frac{\mathbf{E}\big\{(\gamma n/W_n)^2\,\mathbf{1}_A\big\}}{\mathbf{P}\{A\}}, \tag{46}$$

and we recall that $\mathbf{P}\{A\} = \Theta(1/\sqrt{n})$. Also,

$$\begin{aligned}
\mathbf{E}\left\{ \frac{\gamma n}{W_n}\,\mathbf{1}_A \right\} &\le \mathbf{E}\left\{ \frac{\gamma n}{(1-\epsilon)\gamma n}\,\mathbf{1}_A \right\} + \mathbf{E}\left\{ \frac{\gamma n}{W_n}\,\mathbf{1}_{[W_n \le (1-\epsilon)\gamma n]}\,\mathbf{1}_A \right\} \\
&\le \frac{1}{1-\epsilon}\,\mathbf{P}\{A\} + \mathbf{E}\left\{ \frac{p_0}{p_1}\cdot\frac{\gamma n}{N_0}\cdot\mathbf{1}_{[N_0 \le np_0/2]}\cdot\mathbf{1}_A \right\} \\
&\qquad + \mathbf{E}\left\{ \frac{p_0}{p_1}\cdot\frac{2\gamma n}{np_0}\cdot\mathbf{1}_{[W_n \le (1-\epsilon)\gamma n]} \right\}.
\end{aligned} \tag{47}$$

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

$$E_1 = (\gamma n/W_n)^2 \gamma n\,\mathbf{P}\{N_0 \le np_0/2\} \le \frac{p_0}{p_1} 2n \exp\big(-2\gamma n (p_0/2)^2\big). \tag{48}$$

Furthermore,

$$E_2 = \frac{2\gamma}{p_1}\,\mathbf{P}\big\{W_n \le (1-\epsilon)\gamma n\big\} = O\left(\frac{1}{n}\right) \tag{49}$$

follows from Chebyshev's inequality, just as in the proof of (i) above. This implies that

$$\mathbf{E}\left\{ \frac{\gamma n}{W_n} \;\middle|\; |T| = n \right\} \le \frac{1}{1-\epsilon} + \frac{O(1/n)}{O(1/\sqrt{n})} = \frac{1+o(1)}{1-\epsilon}, \tag{50}$$

and we have settled part (ii), since ϵ was chosen arbitrarily. For part (iii), we proceed the same way to obtain

$$\begin{aligned}
\mathbf{E}\left\{ \frac{(\gamma n)^2}{{W_n}^2}\,\mathbf{1}_A \right\} &\le \frac{1}{(1-\epsilon)^2}\,\mathbf{P}\{A\} + \mathbf{E}\left\{ \frac{(\gamma n)^2}{{W_n}^2}\,\mathbf{1}_{[W_n \le (1-\epsilon)(\gamma n)^2]}\,\mathbf{1}_A \right\} \\
&\le \frac{1}{(1-\epsilon)^2}\,\mathbf{P}\{A\} + \mathbf{E}\left\{ \frac{{p_0}^2}{{p_1}^2}\cdot\frac{(\gamma n)^2}{{N_0}^2}\cdot\mathbf{1}_{[N_0 \le np_0/2]}\cdot\mathbf{1}_A \right\} \\
&\qquad + \mathbf{E}\left\{ \frac{{p_0}^2}{{p_1}^2}\cdot\frac{4(\gamma n)^2}{n^2 p_0^2}\cdot\mathbf{1}_{[W_n \le (1-\epsilon)\gamma n]} \right\} \\
&\le \frac{1}{(1-\epsilon)^2}\,\mathbf{P}\{A\} + \frac{{p_0}^2}{{p_1}^2}(\gamma n)^2\,\mathbf{P}\{N_0 \le np_0/2\} \\
&\qquad + \frac{4\gamma^2}{{p_1}^2}\,\mathbf{P}\{W_n \le (1-\epsilon)\gamma n\},
\end{aligned} \tag{51}$$

from which we complete the proof in the same manner. ▮

CHAPTER TWO

ROOT ESTIMATION

2.1. Introduction

There are two different, equally important, notions of a tree. The first is the *unrooted* or *free* tree, which is a connected unlabelled 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 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. The problem was first considered by J. Haigh for uniform attachment trees [28]. More recently, S. Bubeck, L. Devroye, and G. Lugosi showed that on uniform attachement and preferential attachment trees, one can construct a confidence set of nodes that contains the root, where the size of this confidence set does not depend on the number of nodes in the graph [13].

Our mission can be formalised as follows. Let a conditional Galton–Watson tree with n nodes be given and suppose that 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.

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

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

and $p_i = 0$ for all other i. Conditioning on the number of nodes n generates a Catalan tree uniformly at random. Fig. 2.1 illustrates the fourteen possibilities when $n = 4$.

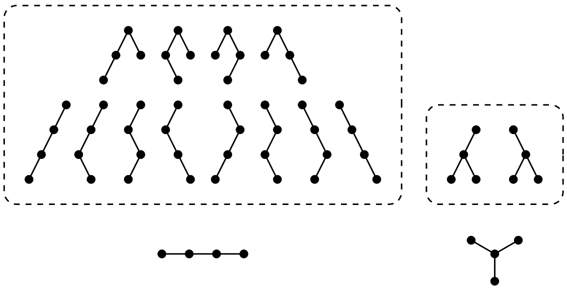


Fig. 2.1. The free-tree structure of Catalan trees with four nodes.

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.

The probabilistic approach. This family of trees illustrated in Fig. 2.1 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 critical offspring distribution ξ and any size n. We will find that the optimal strategy for picking a root is as follows.

i) 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.

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

Note that determining the MLE for a given tree is computationally easy, and we can give an explicit formula for the probability of correctness in the second case. We also analyse the correctness of the MLE as the number of nodes in the tree tends to infinity, showing in Theorem 2.2 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 $\mathbf{P}\{\mathcal{C}\}$ of the MLE being correct satisfies

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

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

2.2. 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* unlabelled, 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 neighbour 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^u; 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 context explains which is meant).

An *automorphism* of a free tree F_n is a graph isomorphism from F_n to itself, i.e., a bijection from the set of vertices of F to itself that preserves the adjacency structure. The group of all such maps is denoted $\mathrm{Aut}(F_n)$. We shall define the *multiplicity* $\mu_{\mathrm{F}}(v)$ of a node $v \in F_n$ to be the size of its orbit under the action of $\mathrm{Aut}(F_n)$. In Chapter 3, we shall define other notions of multiplicity, and this version μ_{F} of multiplicity will be referred to as the *free multiplicity*.

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 amongst themselves while leaving u firmly planted at the top of the tree. In group-theoretic parlance, this is the stabiliser subgroup $\mathrm{Stab}(u)$ 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 $\mathrm{Perm}(T)$ be the number of possible such reorderings that one can perform on a given rooted ordered tree

T; it is clear that there are

$$\prod_{v\in T} \deg_u(v)! \tag{2}$$

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 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 $\mathrm{Stab}(u)$.

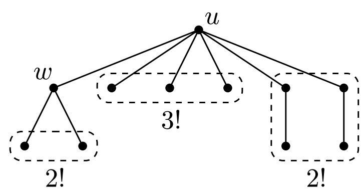


Fig. 2.2. An example tree, in which $\mathrm{Stab}(u) = 2! \cdot 3! \cdot 2! = 96$.

For a tree T with root node u, we let $\mathrm{Perm}(u)$ be the number of distinct unlabelled rooted ordered trees that can be obtained from T by reordering children of nodes. It is given by the formula

$$\mathrm{Perm}(u) = \frac{1}{\left|\mathrm{Stab}(u)\right|} \prod_v \deg_u(v) \tag{3}$$

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

$$\mathrm{Prob}(u) = \prod_{i=0}^{\infty} p_i^{N_i}. \tag{4}$$

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 $\mathrm{Aut}(F_n)$ could have been the root of an identical tree, so we must divide by $\mu_{\mathrm{F}}(u)$. Hence the probability that u is the root is proportional to

$$\frac{\mathrm{Prob}(u)\,\mathrm{Perm}(u)}{\mu_{\mathrm{F}}(u)} = \frac{\mathrm{Prob}(u)}{\mu_{\mathrm{F}}(u)\left|\mathrm{Stab}(u)\right|} \prod_v \deg_u(v)! = \frac{\mathrm{Prob}(u)}{\left|\mathrm{Aut}(F_n)\right|} \prod_v \deg_u(v)!; \tag{5}$$

one must of course introduce a normalising factor to ensure that this is indeed a valid probability distribution. Note that the last equality above is a consequence of the orbit-stabiliser theorem (see, e.g., [34]). Our maximum likelihood estimator will thus need to choose a node u that maximises this probability. Given a Galton–Watson offspring distribution, we will denote by $\mathcal{C}$ the event that the MLE is correct for any corresponding free tree of size n, and we seek to determine both $\mathbf{P}\{\mathcal{C}\}$, the probability of success of the MLE, and $\mathbf{P}\{\mathcal{C} \mid F_n\}$, the probability of success given a specific free tree F_n. Note that

$$\mathbf{P}\{\mathcal{C}\} = \mathbf{E}_{F_n}\big\{\mathbf{P}\{\mathcal{C} \mid F_n\}\big\}, \tag{6}$$

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

2.3. Estimating the root

We are now ready to prove the first significant result. Since $\left|\mathrm{Aut}(F_n)\right|$ does not depend on the choice of root, this boils down to maximising the quantity $\mathrm{Prob}(u)\prod_v \deg_u(v)!$. 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 chapter, we will assume that $0/0 = 0$, capturing the cases where both p_i and p_{i-1} are equal to zero.

Theorem 2.1. *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 maximise R_i. More specifically, defining*

$$\mathcal{M} = \sup_{j\geq 1}\left\{R_j : p_j \neq 0 \text{ and there exists } u \in F_n \text{ such that } \deg^*(u) = j\right\}, \tag{7}$$

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

$$\Omega = \left\{u \in F_n : \deg^*(u) = i,\ R_i = \mathcal{M}\right\}. \tag{8}$$

Letting $\mathbf{P}\{\mathcal{C} \mid F_n\}$ be the probability that this MLE *returns the correct root of F_n, we have*

$$\mathbf{P}\{\mathcal{C} \mid F_n\} = \begin{cases} 1, & \text{if } \mathcal{M} = \infty; \\ \mathcal{M}/\sum_{v\in F_n} R_{\deg^*(v)}, & \text{if } \mathcal{M} < \infty. \end{cases} \tag{9}$$

Proof. The probability that any node $u \in F_n$ is the root is given by the formula (5). Thus the goal is to pick a node u that maximises $\mathrm{Prob}(u)\prod_v \deg_u(v)!$. 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

$$\begin{aligned}
\text{Prob}(u) \prod_v \deg_u(v)! &= \prod_{j=0}^{\infty} p_j^{N_j^u} \prod_v \deg_u(v)! \\
&= \prod_{j=0}^{\infty} p_j^{N_j^u} (j!)^{N_j^u} = \prod_j (j! p_j)^{N_j^u} \\
&= (i! p_i)^{N_i} \big((i-1)! p_{i-1}\big)^{N_{i-1}^u} \prod_{j \notin \{i, i-1\}} (j! p_j)^{N_j^u} \qquad (10) \\
&= (i! p_i)^{N_{i+1}^* + 1} \big((i-1)! p_{i-1}\big)^{N_i^* - 1} \prod_{j \notin \{i, i-1\}} (j! p_j)^{N_{j+1}^*} \\
&= \frac{i p_i}{p_{i-1}} \prod_{j=0}^{\infty} (j! p_j)^{N_{j+1}^*}.
\end{aligned}$$

The infinite product in the last line is the same for all u, so we need only maximise 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 maximise the required ratio (see the limit of d-ary trees as $d \to \infty$ in the following section), there are two cases for the probability of success of this MLE.

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 is the only node in the candidate set Ω and our strategy determines the root correctly with probability $P\{\mathcal{C} \mid F_n\} = 1$. If, on the other hand, $\mathcal{M} < \infty$, then since the probability of any node of degree i being the root is proportional to R_i, normalising over all nodes in the free tree F_n, we obtain

$$\mathbf{P}\{\mathcal{C} \mid F_n\} = \mathcal{M} \Big/ \sum_{v \in F_n} R_{\deg^*(v)}. \qquad (11)$$

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

2.4. Applications to d-ary and Cayley trees

Theorem 2.1 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 Catalan trees in the introduction to this chapter. 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 generalises to d-ary trees of any size.

Recall that we can generate an n-node d-ary tree uniformly at random by generating a conditional Galton–Watson tree with a Binomial$(d, 1/d)$ offspring distribution. Here we have

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

for every $i \in \{0, \ldots, d\}$, whence

$$R_i = \frac{ip_i}{p_{i-1}} = i\binom{d}{i}\binom{d}{i-1}^{-1}\frac{1}{d}\cdot\frac{d}{d-1} = \frac{d-i+1}{d-1}. \tag{12}$$

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

$$R_i \Big/ \sum_v R_{\deg^*(v)} = \frac{d-i+1}{\sum_v (d-\deg^*(v)+1)} = \frac{d-i+1}{nd-(2n-2)+n} = \frac{d-i+1}{(d-1)n+2}. \tag{13}$$

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

$$\mathbf{P}\{\mathcal{C}\} = \mathbf{P}\{\mathcal{C} \mid F_n\} = \frac{d}{(d-1)n+2}. \tag{14}$$

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

$$\frac{ip_i}{p_{i-1}} = \frac{i \cdot e \cdot (i-1)!}{e \cdot i!} = 1, \tag{15}$$

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 $1/n$. Of course, this should come as no surprise since rooted Cayley trees are sampled by first picking a free tree uniformly at random and then choosing a root uniformly at random.

In both of these cases, $\mathbf{P}\{\mathcal{C} \mid F_n\}$ only depends on n, and we thus have $\mathbf{P}\{\mathcal{C}\} = \mathbf{P}\{\mathcal{C} \mid F_n\}$, lending to easy analysis of the MLE. This will not be true in all cases, and in the next section we will employ some of the tools from Chapter 1 to analyse $\mathbf{P}\{\mathcal{C}\}$ for more complex offspring distributions.

2.5. The probability of correctness

We begin by setting up a few definitions to better deal with the two cases mentioned in Theorem 2.1, in the limit as $n \to \infty$. Using this notation, we reformulate our maximum likelihood estimator for the root, and compute its expected probability of correctness $\mathbf{P}\{\mathcal{C}\}$.

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, since 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 $\mathcal{S}$. Note that $i = 1$ is never special, since $p_0 > 0$. We group all non-special integers i into equivalence classes $\{J_k\}_{k\geq 1}$ according to the equivalence

$$i \sim j \quad \text{if and only if} \quad \frac{ip_i}{p_{i-1}} = \frac{jp_j}{p_{j-1}}. \tag{16}$$

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

$$N_{J_k} = \sum_{i \in J_k} N_i^*. \tag{17}$$

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.

i) 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, $\mathbf{P}\{\mathcal{C} \mid F_n\}\,\mathbf{1}_{S_n} = \mathbf{1}_{S_n}$.

ii) 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. If this case arises, select a node uniformly at random from the class J_λ, where

$$\lambda = \arg\max_{k \notin \mathcal{S}} \{R_{J_k} : N_{J_k} > 0\}. \tag{18}$$

This maximum is well-defined, since there are at most n nonempty equivalence classes. In this case,

$$\mathbf{P}\{\mathcal{C} \mid F_n\}\mathbf{1}_{S_n^c} = \frac{R_\lambda}{\sum_k N_{J_k} R_{J_k}} \mathbf{1}_{S_n^c}. \tag{19}$$

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 use the notion of Kesten's limit tree described in the first chapter.

Theorem 2.2. *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*

$$\lim_{n\to\infty} n \cdot \mathbf{P}\{\mathcal{C}\} = \sup_{i\geq 1} \frac{ip_i}{p_{i-1}}. \tag{20}$$

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. We have

$$\mathbf{P}\{\mathcal{C}\} = \mathbf{E}\big\{\mathbf{P}\{\mathcal{C} \mid F_n\}\big\} = \mathbf{E}\left\{\frac{R_\lambda}{\sum_k N_{J_k} R_{J_k}}\right\} \leq \sup_{i\geq 1} R_i\, \mathbf{E}\left\{\frac{1}{\sum_k N_{J_k} R_{J_k}}\right\}, \tag{21}$$

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 1.7. That lemma in this case gives us

$$\mathbf{E}\{n/W_n \mid |T| = n\} \to 1, \tag{22}$$

since here $\gamma = 1$. We thus conclude that

$$\limsup_{n\to\infty} n\, \mathbf{P}\{\mathcal{C}\} \leq \sup_{i\geq 1} R_i. \tag{23}$$

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

$$\mathbf{P}\{N_i^* = 0\} \to 0 \tag{24}$$

as $n \to \infty$.

Note that by Lemma 1.4, for any conditional Galton–Watson tree corresponding to the free tree of size n rooted at a node u, for all i, $N_i/np_i \to 1$ 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 $i \geq 1$,

$$\begin{aligned}
\mathbf{P}\{N_i^* = 0\} &= \mathbf{P}\{N_i^* = 0, \deg^*(u) \notin \{i, i-1\}\} \\
&\qquad + \mathbf{P}\{N_i^* = 0, \deg^*(u) = i-1\} + \mathbf{P}\{N_i^* = 0, \deg^*(u) = i\} \\
&= \mathbf{P}\{N_{i-1}^u = 0 \mid \deg^*(u) \notin \{i, i-1\}\}\, \mathbf{P}\{\deg^*(u) \notin \{i, i-1\}\} \\
&\qquad + \mathbf{P}\{N_{i-1}^u - 1 = 0 \mid \deg^*(u) = i-1\}\, \mathbf{P}\{\deg^*(u) = i-1\},
\end{aligned} \tag{25}$$

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 brevity of notation.

For the lower bound, we must consider two cases. We shall declare the situation in which the supremum $\sum_{i\geq 1} R_i$ is *finite* to be case (i), and the complementary case, in which it is *infinite*, to be case (ii). 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 $R \in \mathbf{R}$ 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

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

of indices of equivalence classes with a corresponding ratio at least 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,

$$\mathbf{P}\{\lambda \notin \mathcal{J}\} = \mathbf{P}\Big\{\bigcap_{\ell\in\mathcal{J}} N_{J_\ell} = 0\Big\} \leq \mathbf{P}\{N_j^* = 0\}, \tag{26}$$

which approaches 0 as $n \to \infty$. We can thus bound the probability of success from below by

$$\begin{aligned}\mathbf{P}\{\mathcal{C}\} &= \mathbf{E}\big\{\mathbf{P}\{\mathcal{C} \mid F_n\}\big\}\\ &\geq \mathbf{E}\left\{\mathbf{1}_{[\lambda\in\mathcal{J}]}\,\frac{R_{J_\lambda}}{\sum_k N_{J_k} R_{J_k}}\right\}\\ &\geq \frac{R}{n(1+\epsilon)}\,\mathbf{E}\left\{\mathbf{1}_{[\lambda\in\mathcal{J}]}\,\mathbf{1}_{\left[\sum_k N_{J_k} R_{J_k} \leq n(1+\epsilon)\right]}\right\}\\ &\geq \frac{R}{n(1+\epsilon)}\left(1 - \mathbf{P}\{\lambda\notin\mathcal{J}\} - \mathbf{P}\Big\{\sum_k N_{J_k} R_{J_k} > n(1+\epsilon)\Big\}\right).\end{aligned} \tag{27}$$

As $n \to \infty$, we have that $\mathbf{P}\{\lambda \notin \mathcal{J}\} \to 0$ 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 1.7, we also have $\mathbf{P}\{\sum_k N_{J_k} R_{J_k}/n > 1+\epsilon\} \to 0$. Thus, in both cases (i) and (ii), the sum of terms in the parentheses approaches 1 as $n \to \infty$.

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

$$\liminf_{n\to\infty} n\,\mathbf{P}\{\mathcal{C}\} \sup_{i\geq 1} R_i, \tag{28}$$

and we have equality in the limit.

In case (ii),

$$\liminf_{n\to\infty} n\,\mathbf{P}\{\mathcal{C}\} \geq R \tag{29}$$

for any arbitrarily large choice of R. We thus have

$$\lim_{n\to\infty} n\,\mathbf{P}\{\mathcal{C}\} = \infty = \sup_{i\geq 1} R_i, \tag{30}$$

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 2.2 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 2.4. In most other cases, the MLE does better, asymptotically speaking, than choosing uniformly at random.

Although the limit of $n\,\mathbf{P}\{\mathcal{C}\}$ may be infinite, the following lemma shows that $n\,\mathbf{P}\{\mathcal{C}\}$ is always $o(n)$ if no special integer exists. 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 2.3. *Let F_n be a random free tree of size n corresponding to a critical conditional Galton–Watson tree with offspring distribution p_i. Let $\mathcal{S}$ 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*

$$\lim_{n \to \infty} \mathbf{P}\{\mathcal{C} \cap S_n^c\} = 0. \tag{31}$$

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

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 1.6 to describe the maximum degree. Next, we define $\kappa = \sup_{i \geq 1, i \notin \mathcal{S}} p_i/p_{i-1} < \infty$. Letting $\lambda \notin \mathcal{S}$ be the class chosen by the MLE, the best ratio can be bounded by

$$R_\lambda \leq \kappa(M_n + 1) \leq 2\kappa M_n. \tag{32}$$

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 defined in Section 1.4, we have

$$\sum_k N_{J_k} R_{J_k} = W_n + \frac{D p_D}{p_{D-1}} - \frac{(D+1) p_{D+1}}{p_D}, \tag{33}$$

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

$$\sum_k N_{J_k} R_{J_k} \geq W_n - (D+1)\kappa. \tag{34}$$

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

$$\begin{aligned} \mathbf{P}\{E_n^c\} &\leq \mathbf{P}\big\{W_n < 2\kappa\sqrt{n}\big\} + \mathbf{P}\big\{D + 1 \geq \sqrt{n}\big\} \\ &\leq 4\kappa^2 n\,\mathbf{E}\big\{(1/W_n)^2\big\} + \frac{\mathbf{E}\{D+1\}}{\sqrt{n}} \\ &= O\left(\frac{1}{n}\right) + \frac{\sigma^2 + 2 + o(1)}{\sqrt{n}}, \end{aligned} \tag{35}$$

where we used the fact that $\mathbf{E}\big\{1/(W_n)^2 \mid |T| = n\big\} = O(1/n^2)$ by Lemma 1.7, 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 Lemma 1.7, $W_n/\gamma n \to 1$ in probability given $|T| = n$, and $\mathbf{E}\big\{\gamma n/W_n \mid |T| = n\big\} \to 1$ as n tends to infinity. The probability of correctness of the MLE can thus be bounded by

$$\begin{aligned}
\mathbf{P}\{\mathcal{C} \cap S_n^c\} &\leq \mathbf{P}\{\mathcal{C} \cap S_n^c \cap E_n\} + \mathbf{P}\{E_n^c\} = \mathbf{P}\{\mathcal{C} \cap S_n^c \cap E_n\} + o(1) \\
&= \mathbf{E}_{F_n}\big\{\mathbf{P}\{\mathcal{C} \mid F_n\}\, \mathbf{1}_{S_n^c \cap E_n}\big\} + o(1) \\
&= \mathbf{E}\left\{\frac{R_\lambda}{\sum_k N_{J_k} R_k}\, \mathbf{1}_{S_n^c \cap C_n}\right\} + o(1) \\
&\leq \mathbf{E}\left\{\frac{2\kappa M_n}{W_n/2} \;\middle|\; |T| = n\right\} + o(1) \\
&\leq 4\kappa\sqrt{\mathbf{E}\left\{M_n^2 \mid |T| = n\right\} \mathbf{E}\left\{1/{W_n}^2 \mid |T| = n\right\}} + o(1).
\end{aligned} \tag{36}$$

To bound $\mathbf{E}\left\{M_n^2 \mid |T| = n\right\}$, let A once again denote the event that $\sum_{i=1}^n \xi_i = n - 1$. We have

$$\begin{aligned}
\mathbf{E}\left\{M_n^2 \mid |T| = n\right\} = \frac{\mathbf{E}\left\{M_n^2\, \mathbf{1}_A\right\}}{\mathbf{P}\{A\}} &\leq \frac{n^2\, \mathbf{P}\{M_n \geq n^{7/8}\} + n^{7/4}\, \mathbf{P}\{A\}}{\mathbf{P}\{A\}} \\
&\leq \Theta\big(n^{5/2}\big)\, \mathbf{P}\{M_n \geq n^{7/8}\} + n^{7/4}
\end{aligned} \tag{37}$$

and proceed by applying the union bound to obtain

$$\begin{aligned}
\mathbf{E}\left\{M_n^2 \mid |T| = n\right\} &\leq n\Theta\big(n^{5/2}\big) \sum_{i \geq n^{7/8}} p_i + n^{7/4} \\
&\leq \Theta\big(n^{7/2}\big) \sum_{i \geq 1} \frac{i^2 p_i}{n^{7/4}} + n^{7/4} \\
&= \Theta\big(n^{7/4}\big),
\end{aligned} \tag{38}$$

where the last equality follows from the fact that $\sigma^2 < \infty$. Substituting everything into (36), we have

$$\mathbf{P}\{\mathcal{C} \cap S_n^c\} = 2\kappa\sqrt{O\big(n^{7/4}\big) O\big(1/n^2\big)} = O\left(\frac{1}{n^{1/8}}\right). \quad \blacksquare \tag{39}$$

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 non-special case, since there is now the possibility of stumbling upon a node that must be the root.

Theorem 2.4. *Fix a random free tree F_n of size n corresponding to a Galton–Watson tree with offspring distribution p_i. Let $\mathcal{S}$ denote the set of special integers and suppose that $\mathcal{S} \neq \emptyset$, $0 < \sigma^2 < \infty$, and $\sup_{i \notin \mathcal{S}} p_i/p_{i-1} < \infty$. The probability of the* MLE *being correct satisfies*

$$\lim_{n\to\infty} \mathbf{P}\{\mathcal{C}\} = \sum_{i\in\mathcal{S}} ip_i + o(1). \tag{40}$$

Proof. The special integers $i \in \mathcal{S}$ satisfy $p_i \neq 0$ and $p_{i-1} = 0$. Recall from case (i) of Theorem 2.2 that if there exists a node in the free tree with some special degree $i \in \mathcal{S}$, then there can only be one such node: $\sum_{i\in\mathcal{S}} N_i^* \leq 1$. Thus we can split $\mathbf{P}\{\mathcal{C}\}$ into two cases: Let S_n and S_n^c be defined as in the previous lemma. Then

$$\begin{aligned}\mathbf{P}\{\mathcal{C}\} &= \mathbf{E}\big\{\mathbf{P}\{\mathcal{C} \mid F_n\}\big\} \\ &= \mathbf{E}\big\{\mathbf{P}\{\mathcal{C} \mid F_n\}\,\mathbf{1}_{S_n}\big\} + \mathbf{P}\{\mathcal{C}\,\mathbf{1}_{S_n^c}\}\end{aligned} \tag{41}$$

The first term here is simply $\mathbf{P}\{S_n\}$, since the MLE satisfies $\mathbf{P}\{\mathcal{C} \mid F_n\}\,\mathbf{1}_{S_n} = \mathbf{1}_{S_n}$. As stated in the proof of Theorem 2.2, a conditional Galton–Watson tree converges locally to Kesten's limit tree as $n \to \infty$. Thus, the existence of a $u \in F_n$ with $\deg^*(u) \in \mathcal{S}$ is the event that a random conditional Galton–Watson tree has root of degree $i \in \mathcal{S}$, which occurs with probability $\sum_{i\in\mathcal{S}} ip_i + o(1)$. Then, noting that $\mathbf{P}\{\mathcal{C} \cap S_n^c\} = o(1)$ by Lemma 2.3, we have

$$\mathbf{P}\{\mathcal{C}\} = \sum_{i\in S} ip_i + o(1). \quad \blacksquare \tag{42}$$

Comparing this result with Theorem 2.2, 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 $n\,\mathbf{P}\{\mathcal{C}\}$ 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 $\mathbf{P}\{\mathcal{C}\}$ itself to approach a nonzero constant.

2.6. 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.

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 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. This argument generalises to all Flajolet t-ary trees for $t \geq 2$.

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

$$R_i = \frac{ip_i}{p_{i-1}} = i \tag{43}$$

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 2.2, we conclude that $n\,\mathbf{P}\{\mathcal{C}\}$ approaches 2 as n gets large, so $\mathbf{P}\{\mathcal{C}\} \sim 2/n$.

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

$$R_i \Big/ \sum_v R_{\deg^*(v)} = \frac{i/2}{\sum_v \deg^*(v)/2} = \frac{i}{2(n-1)}. \tag{44}$$

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 1.6. For an upper bound, we have

$$\mathbf{P}\{M_n \geq x\} \leq \bigl(1 + o(1)\bigr) n\,\mathbf{P}\{\xi \geq x\} \sim n/2^x \tag{45}$$

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 $a_n \to \infty$ as $n \to \infty$.) Likewise, we can derive the lower bound

$$\mathbf{P}\{M_n \leq x\} \leq \bigl(\beta + o(1)\bigr) \exp\bigl(-n\,\mathbf{P}\{\xi \geq x\}\bigr) \sim \beta \exp(-n/2^{x+1}) \tag{46}$$

for the constant β given by Lemma 1.6 and this goes to 0 provided that $x = \log_2 n - \omega(1)$. In other words,

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

and

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

i.e., $M_n / \log_2 n \to 1$ in probability. This means that for a planted plane tree,

$$\mathbf{P}\{\mathcal{C}\} = \frac{\mathbf{E}\{M_n\}}{2(n-1)} \sim \frac{\log_2 n}{2n}. \tag{47}$$

Table 1

THE PROBABILITY OF CORRECTNESS OF THE MLE
FOR SOME FAMILIES OF TREES

Family	Distribution	MLE	$\mathbf{P}\{\mathcal{C}\}$
d-ary	Binomial$(d, 1/d)$	Leaf	$\dfrac{k}{(k-1)n+2}$
Cayley	Poisson(1)	Choose uniformly	$1/n$
Full binary	Uniform$\{0, 2\}$	Degree 2	1
Planted plane	Geometric$(1/2)$	Maximize degree	$\dfrac{\mathbf{E}\{M_n\}}{2(n-1)} \sim \dfrac{\log_2 n}{2n}$
Motzkin	Uniform$\{0, 1, 2\}$	Degree 2	$\dfrac{2+o(1)}{n}$

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

$$p_i = \frac{\theta}{(i+1)^\alpha}, \tag{48}$$

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

$$\sum_{i=1}^{\infty} \frac{N_i^* R_i}{n} \to \sum_{i=1}^{\infty} p_{i-1} \frac{i p_i}{p_{i-1}} = 1 \tag{49}$$

in probability, and thus

$$\left| \mathbf{P}\{\mathcal{C} \mid F_n\} - \frac{M_n}{n} \right| \leq f(M_n, n) \tag{50}$$

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

$$\mathbf{P}\{\mathcal{C}\} \sim \frac{\mathbf{E}\{M_n\}}{n}. \tag{51}$$

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

CHAPTER THREE

LEAF MULTIPLICITY

3.1. Introduction

EQUIVALENCE BETWEEN two distinct mathematical objects is a far-reaching concept in mathematics. When two structures are similar, one may define a relation under which they are regarded as one and the same. The term "multiplicity" is often used to indicate the extent to which an object is, in some sense, not structurally unique (or how often it is repeated in a suitably-defined multiset). Towards a concept of the multiplicity of a node in a tree, consider the small example depicted in Fig. 3.1.

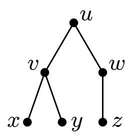


Fig. 3.1. A rooted tree, in which the pair x and y are similar, but x and z are not.

Definitions and notation. Consider a tree T rooted at a node u. For a node v in the tree, we let T_v denote the subtree rooted at v. Let v and w be nodes in the tree and let $v = v_1, v_2, \ldots, v_n = u$ and $w = w_1, w_2, \ldots, w_m = u$ be the paths from v and w, respectively, to the root. We say that v and w are *identical* and write $v \equiv w$ if the paths have the same length and T_{v_j} and T_{w_j} are isomorphic as rooted ordered trees for $1 \leq j \leq n$.

It is clear that $\equiv$ defines an equivalence relation on the set of nodes in the tree, so we may now define the *multiplicity* $\sigma(v)$ of a node v to be the size of the equivalence class $[v]$ under the relation. For example, in Fig. 3.1, the nodes x and y each have multiplicity 2 and z has multiplicity 1. The *leaf multiplicity* (or simply *multiplicity*, when no confusion can arise) $S(T)$ of a rooted tree T is the maximum value of $\sigma(v)$, taken over all nodes v of T. The name "leaf multiplicity" is motivated by the fact that the function σ increases monotonically on any path from the root, so that $S(T)$ remains the same when the maximum is only computed over the set of *leaves* of T.

Note that $\equiv$ is not the only structural equivalence relation one can define on the set of nodes in a tree, and thus σ is only one of many possible notions of leaf multiplicity. Towards the end of this chapter, we will explore an alternate definition μ of multiplicity in a rooted tree that extends to the notion of multiplicity μ_{F} that we saw in Chapter 2. We will then discuss the relationship between σ, μ, and μ_{F}.

Rényi entropy. It will be convenient to simplify our notation with some information-theoretic definitions. Letting $p_i = \mathbf{P}\{\xi = i\}$, for $\alpha > 1$ we define the *Rényi entropy of order* α [46] (see also [32]) to be the value

$$\mathrm{H}_\alpha(\xi) = \frac{1}{\alpha - 1} \log_2 \frac{1}{\sum_{i\geq 0} {p_i}^\alpha}. \tag{1}$$

As $\alpha \to 1$, this approaches the *binary (Shannon) entropy* [48]

$$\mathrm{H}(\xi) = \sum_{i\geq 0} p_i \log_2 \frac{1}{p_i}.$$

Since ξ will be fixed throughout the chapter, for brevity we will let $\mathrm{H}_\alpha = \mathrm{H}_\alpha(\xi)$ and $\mathrm{H} = \mathrm{H}(\xi)$.

Fix an offspring distribution ξ with mean 1 and nonzero finite variance; let T_n be a conditional Galton–Watson tree of size n with this offspring distribution. The leaf multiplicity $S(T_n)$ of this tree is a random variable, and will be denoted by S_n (this should not be confused with the *event* S_n from Chapter 2. The main result of this chapter gives bounds on S_n that are obeyed asymptotically in probability.

Theorem 3.1. *Let ξ be an offspring distribution with $\mathbf{E}\{\xi\} = 1$ and $\mathbf{V}\{\xi\} \in (0, \infty)$. If S_n is the multiplicity of a conditional Galton–Watson tree of size n with offspring distribution ξ, then letting*

$$\gamma = \max_{k\geq 2} {p_0}^k {p_k}^{k/(k-1)}, \tag{2}$$

we have for all $\epsilon > 0$,

$$\mathbf{P}\left\{ S_n \geq (1-\epsilon) \frac{\log_2 n}{\log_2(1/\gamma)} \right\} \to 1$$

as $n \to \infty$, and under the further assumption that $\mathbf{E}\{2^\xi\} < \infty$, we have the upper bound

$$\mathbf{P}\left\{ S_n \leq (1+\epsilon) \frac{2\log_2 n}{\mathrm{H}_2} \right\} \to 1$$

as $n \to \infty$, where H_2 is the Rényi entropy of order 2 of ξ. ▮

This theorem will be proved as two separate lemmas in the next section.

3.2. Asymptotics of the leaf multiplicity

In this section we derive asymptotic upper and lower bounds on S_n. Before we begin, observe that if $p_{\max} = \max_{i\geq 0} p_i$ and $1 < \alpha < \beta < \infty$, we have the inequalities

$$e^{-\mathrm{H}} \leq \Big(\sum_{i\geq 0} {p_i}^\alpha\Big)^{1/(\alpha-1)} \leq \Big(\sum_{i\geq 0} {p_i}^\beta\Big)^{1/(\beta-1)} \leq p_{\max} \tag{3}$$

and

$$p_{\max} \le \Big(\sum_{i\ge 0} {p_i}^{\beta}\Big)^{1/\beta} \le \Big(\sum_{i\ge 0} {p_i}^{\alpha}\Big)^{1/\alpha} \le 1. \tag{4}$$

Defining $\mathrm{H}_\infty = \log_2(1/p_{\max})$, we have the equivalent chain of inequalities

$$\mathrm{H} \ge \mathrm{H}_\alpha \ge \mathrm{H}_\beta \ge \mathrm{H}_\infty \ge \frac{\beta-1}{\beta}\mathrm{H}_\beta \ge \frac{\alpha-1}{\alpha}\mathrm{H}_\alpha \ge 0. \tag{5}$$

Moreover, because we have assumed that $\mathbf{V}\{\xi\} \ne 0$, we have the strict inequality

$$\Big(\sum_{i\ge 0} {p_i}^{k}\Big)^{1/k} < \Big(\sum_{i\ge 0} {p_i}^{2}\Big)^{1/2} \tag{6}$$

for all $k > 2$.

First we prove the upper bound for S_n.

Lemma 3.2. *Let ξ be an offspring distribution with mean 1 and nonzero finite variance σ^2. Suppose further that $\mathbf{E}\{2^\xi\}$ is finite. If S_n is the multiplicity of a conditional Galton–Watson tree of size n with offspring distribution ξ, then*

$$\mathbf{P}\{S_n > (1+\epsilon)2\log_2 n/\mathrm{H}_2\} \to 0 \tag{7}$$

for all $\epsilon > 0$, where H_2 is the Rényi entropy of order 2 of the random variable ξ.

Proof. For $1 \le i \le n$, let ξ_i denote the degree of the ith node in preorder in the tree T_n. For all $1 \le t < n$, the partial sum $\sum_{i=1}^t \xi_i > t-1$ and $\sum_{i=1}^n \xi_i = n-1$. We will concentrate on the least common ancestor of the nodes in the largest equivalence class of T_n. This node, call it w, has the property that the nodes in the equivalence class belong to $k \ge 2$ different subtrees rooted at the children of w. The node w has (random) degree D, which we will deal with by summing over all possible degrees d. Let $\mathcal{A}_{wk}$ denote the collection of all subsets of size k of the children of w (naturally, this collection is empty if w has fewer than k children). For $x > 0$, a node w, integers $2 \le k \le d$, and a set $A \in \mathcal{A}_{wk}$, we let $E(x,w,k,A)$ be the event that all the nodes in A are identical and their subtree sizes are at least x/k. Now for integers $s \ge x/k$, we let $E'(x,k,d,s,A)$ be the event that a uniformly random node w of the tree T_n has degree d and the leftmost k children of w are identical, with subtrees of size s. We have, by the union bound,

$$\begin{aligned}\mathbf{P}\{S_n \ge x\} &\le \mathbf{P}\Big\{\bigcup_{w\in T_n}\bigcup_{k\ge 2}\bigcup_{A\in\mathcal{A}_{wk}} E(x,w,k,A)\Big\} \\ &\le n\sum_{k\ge 2}\sum_{d\ge k}\binom{d}{k}\sum_{s\ge x/k}\mathbf{P}\big\{E'(x,k,d,s,A)\big\}.\end{aligned} \tag{8}$$

Supposing that w is the jth node in preorder, $E'(x,k,d,s,A)$ is the event that $\xi_j = d$, $(\xi_{j+1},\dots,\xi_{j+s})$ forms a tree, and $(\xi_{j+rs+1,\dots,j+rs+s}) = (\xi_{j+1},\dots,\xi_{j+s})$

for all $1 \le r < k$. Let us say that an integer j is "good" if these conditions hold when addition on the indices is done modulo n. Clearly, there are more good j than j satisfying the above conditions. Let G be the event that an index j chosen uniformly at random from $\{1, \dots, n\}$ is "good"; let B be the event that $(\xi_2, \dots, \xi_s)$ forms a tree and $(\xi_{rs+2}, \dots, \xi_{(r+1)s+1}) = (\xi_{j+1}, \dots, \xi_{j+s})$ for all $1 \le r < k$. By the cycle lemma,

$$\begin{aligned}
&\mathbf{P}\bigl\{G \mid (\xi_1, \dots, \xi_n) \text{ forms a tree}\bigr\} \\
&\qquad = \mathbf{P}\Bigl\{G \Bigm| \sum_{i=1}^{n} \xi_i = n-1\Bigr\} \\
&\qquad = \frac{\mathbf{P}\bigl\{\xi_1 = d,\, B,\, \sum_{i=1}^{n} \xi_i = n-1\bigr\}}{\mathbf{P}\bigl\{\sum_{i=1}^{n} \xi_i = n-1\bigr\}} \\
&\qquad = \frac{\mathbf{P}\bigl\{\xi_1 = d,\, B,\, \sum_{i=\lfloor 1+ks\rfloor+1}^{n} \xi_i = (n-1) - d - k(s-1)\bigr\}}{\mathbf{P}\bigl\{\sum_{i=1}^{n} \xi_i = n-1\bigr\}},
\end{aligned}$$

so letting

$$R = \frac{\mathbf{P}\bigl\{\sum_{i=1}^{n-(1-ks)} \xi_i = \bigl(n - (1-ks) - 1\bigr) + (k+1-d)\bigr\}}{\mathbf{P}\{\sum_{i=1}^{n} \xi_i = n-1\}},$$

we have

$$\mathbf{P}\bigl\{G \mid (\xi_1, \dots, \xi_n) \text{ forms a tree}\bigr\} = p_d\, \mathbf{P}\{B\} R. \tag{9}$$

Letting $\lambda = \gcd\{i : i \ge 1, p_i > 0\}$, Kolchin's theorem states that uniformly in y,

$$\mathbf{P}\Bigl\{\sum_{i=1}^{n} \xi_i = n-y\Bigr\} = \begin{cases} \lambda/\bigl(\sqrt{2\pi n}\sigma^2\bigr)e^{-y^2/2n\sigma^2} + o\bigl(1/\sqrt{n}\bigr), & \text{if } n \bmod \lambda = 0; \\ 0, & \text{if } n \bmod \lambda \ne 0. \end{cases} \tag{10}$$

As the $o(1)$ term does not depend on y, we find that

$$R = \sqrt{\frac{n-1}{n-(1-ks)-1+(k+1-d)}}\, \exp\biggl(-\frac{(1-ks+d-k)^2}{2\bigl(n-(1-ks+d-k)\bigr)}\sigma^2\biggr) + o(1),$$

where the $o(1)$ term depends only on n. Assuming that $ks + d \le n/2$, we have $R \le \sqrt{2} + o(1)$. Hence

$$\mathbf{P}\bigl\{G \mid (\xi_1, \dots, \xi_n) \text{ forms a tree}\bigr\} \le \bigl(\sqrt{2} + o(1)\bigr)p_d\, \mathbf{P}\{B\} \tag{11}$$

whenever $ks + d \le n/2$. We now compute a bound on $\mathbf{P}\{B\}$. We have

$$\mathbf{P}\{B \mid \xi_2, \dots, \xi_{1+s}\} = (p_{\xi_2} \cdots p_{\xi_{1+s}})^{k-1}$$

and therefore, by independence of the ξ_i,

$$\begin{aligned} \mathbf{P}\{B\} &= \mathbf{E}\bigl\{(p_{\xi_2}\cdots p_{\xi_{1+s}})^{k-1}\,\mathbf{1}_{[(\xi_2,\ldots,\xi_{1+s})\text{ forms a tree}]}\bigr\} \\ &\le \prod_{i=2}^{1+s}\mathbf{E}\bigl\{(p_{\xi_i})^{k-1}\bigr\} = \Bigl(\sum_{i\ge0}{p_i}^k\Bigr)^s. \end{aligned} \tag{12}$$

We can now combine all of these bounds. Substituting everything into (8), we have

$$\begin{aligned} \mathbf{P}\{S_n \ge x\} &\le n\sum_{k\ge2}\sum_{d\ge k}\binom{d}{k}\sum_{s\ge x/k}(\sqrt{2}+o(1))p_d\Bigl(\sum_{i\ge0}{p_i}^k\Bigr)^s \\ &\le \bigl(\sqrt{2}+o(1)\bigr)n\sum_{k\ge2}\sum_{d\ge k}p_d\binom{d}{k}\Bigl(\sum_{i\ge0}{p_i}^k\Bigr)^{x/k}\frac{1}{1-\sum_{i\ge0}{p_i}^k}. \end{aligned}$$

Since the inequality (6) was strict, there exists $0<\theta<1$ such that

$$\begin{aligned} \mathbf{P}\{S_n \ge x\} &\le \frac{\sqrt{2}+o(1)}{1-\sum_{i\ge0}{p_i}^2}\biggl(n\sum_{d\ge2}p_d\binom{d}{2}\Bigl(\sum_{i\ge0}{p_i}^2\Bigr)^{x/2} \\ &\qquad\qquad + n\sum_{k\ge3}\sum_{d\ge k}p_d\binom{d}{k}\Bigl(\sum_{i\ge0}{p_i}^2\Bigr)^{x/2}\theta^x\biggr) \\ &\le \frac{\sqrt{2}+o(1)}{1-\sum_{i\ge0}{p_1}^2}n(\sigma^2+1)\Bigl(\sum_{i\ge0}{p_i}^2\Bigr)^{x/2} \\ &\qquad\qquad + n\Bigl(\sum_{i\ge0}{p_i}^2\Bigr)^{x/2}\theta^x\sum_{k\ge3}\sum_{d\le k}p_d\binom{d}{k}. \end{aligned} \tag{13}$$

Since

$$\sum_{d\ge2}p_d\sum_{k=3}^{d}\binom{d}{k} \le \sum_{d\ge3}p_d 2^d \le \mathbf{E}\{2^\xi\},$$

we have

$$\mathbf{P}\{S_n \ge x\} \le n\frac{\sqrt{2}(\sigma^2+1)}{1-\sum_{i\ge0}{p_i}^2}\Bigl(\sum_{i\ge0}{p_i}^2\Bigr)^{x/2}\bigl(1+o(1)\bigr), \tag{14}$$

provided that $\mathbf{E}\{2^\xi\}<\infty$. Setting

$$x = (1+\epsilon)\frac{2\log_2 n}{\log_2\bigl(1/\sum_{i\ge0}{p_i}^2\bigr)} = (1+\epsilon)\frac{2\log_2 n}{\mathrm{H}_2},$$

we find that $\mathbf{P}\{S_n \ge x\}\to 0$ as $n\to\infty$. ▮

The next lemma presents a lower bound for S_n.

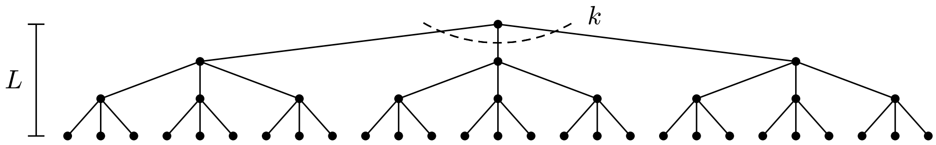


Fig. 3.2. The construction in the proof of Lemma 3.3, with $k = L = 3$.

Lemma 3.3. *Let ξ be an offspring distribution with mean 1 and nonzero finite variance σ^2. If S_n is the multiplicity of a conditional Galton–Watson tree of size n with offspring distribution ξ, then*

$$\mathbf{P}\left\{S_n < (1-\epsilon)\frac{\log_2 n}{\log_2(1/\gamma)}\right\} \to 0 \tag{15}$$

for all $0 < \epsilon < 1$, where $\gamma = \max_{k\geq 2} {p_0}^k {p_k}^{k/(k-1)}$.

Proof. Consider a complete k-ary tree of height L. This tree has k^L leaves and $1 + k + \cdots + k^{L-1} = (k^L - 1)/(k-1)$ internal nodes, all of degree k. The probability that an unconditional Galton–Watson tree takes this shape is

$${p_0}^{k^L} {p_k}^{(k^L-1)/(k-1)};$$

call this probability q. For any real number x, the statement $S_n < x$ implies that no node in the tree can have the given k-ary tree as a subtree for any $k^L \geq x$, as the multiplicity of the k-ary tree is k^L. Fix $k \geq 2$ for now, let L be the first integer for which $k \geq x$, and let $y = k^L$. Observe that $y \leq kx$. Denote the size of the k-ary tree by $z = y + (y-1)/(k-1)$.

We now consider the indices $1, 1+z, 1+2z, 1+3z, \ldots$ in $\{1, \ldots, n-z\}$. Let Y_i be the event (and ${Y_i}^c$ its complement) that $(\xi_i, \ldots, \xi_{i+z-1})$ defines precisely the k-ary tree, where i is in the set of indices defined above, which has size $\lfloor (n-z)/z \rfloor$. Note that

$$\mathbf{P}\{S_n < x\} \leq \mathbf{P}\{S_n < y\} = \mathbf{P}\left\{\bigcap_{i=1}^{n-z} {Y_i}^c \;\middle|\; (\xi_1, \ldots, \xi_n) \text{ defines a tree}\right\}.$$

By the cycle lemma, the probability that $(\xi_1, \ldots, \xi_n)$ defines a tree is $\Theta(n^{3/2})$, so

$$\begin{aligned}
\mathbf{P}\{S_n < x\} &\leq \Theta(n^{3/2})\left\{\bigcap_{i=1}^{n-z} {Y_i}^c\right\} \\
&= \Theta(n^{3/2})\,\mathbf{P}\{{Y_i}^c\}^{\lfloor (n-z)/z \rfloor} \\
&= \Theta(n^{3/2})(1-q)^{\lfloor (n-z)/z \rfloor}
\end{aligned}$$

$$\begin{aligned}
&\le \Theta(n^{3/2}) \exp\biggl(-\biggl\lfloor \frac{n-z}{z} \biggr\rfloor {p_0}^y {p_k}^{(y-1)/(k-1)}\biggr) \qquad (16)\\
&\le \Theta(n^{3/2}) \exp\biggl(-\biggl\lfloor \frac{n-z}{z} \biggr\rfloor {p_0}^{kx} {p_k}^{(kx-1)/(k-1)}\biggr)\\
&\le \Theta(n^{3/2}) \exp\biggl(-\Omega(1)\biggl\lfloor \frac{n-z}{z} \biggr\rfloor \bigl({p_0}^k {p_k}^{(k-1)/(k-1)}\bigr)^x\biggr)\\
&\le \Theta(n^{3/2}) \exp\biggl(-\Omega(1)\biggl\lfloor \frac{n-z}{z} \biggr\rfloor \gamma^x\biggr).
\end{aligned}$$

Substituting $(1-\epsilon)\log_2 n/\log_2(1/\gamma)$ for x, and noting that $z = \Theta(\log n)$, we observe that this bound tends to 0. ∎

3.3. The maximal leaf-degree

Let T_n be a random critical Galton–Watson tree of size n. We let ξ_u be the degree of the node u and let λ_u be the number of children of u that are leaves in T_n, i.e., the *leaf-degree* of u. We denote by L_n the random variable $\max_{u\in T_n} \lambda_u$; it is clear that the multiplicity S_n satisfies $M_n \ge L_n$. The next lemma shows that when the tail of the offspring distribution ξ decays at a rate slower than exponential, the ratio $L_n/\log n \to \infty$ in probability. So while our condition in the upper bound that $\mathbf{E}\{2^\xi\}$ be finite might have seemed somewhat artificial at first glance, we essentially cannot do without it.

Theorem 3.4. *Let $\mathbf{E}\{\xi\} = 1$, $\mathbf{V}\{\xi\} = \sigma^2 \in (0,\infty)$, and suppose that $\mathbf{E}\{\rho^\xi\} = \infty$ for every $1 < \rho < \infty$. Let L_n be the maximal leaf-degree in T_n, the Galton–Watson tree induced by ξ, of size n. Then*

$$\frac{L_n}{\log n} \to \infty$$

in probability along a subsequence, as $n \to \infty$.

Proof. We argue by coupling T_n with Kesten's limit tree T_∞ Let $\tau(T_n,k)$ and $\tau(T_\infty,k)$ denote the truncations of T_n and T_∞, respectively. We let k_n be a sequence to be defined when the time is ripe, and we couple $\tau(T_n,k_n)$ and $\tau(T_\infty,k_n)$ such that

$$\mathbf{P}\bigl\{\tau(T_n,k_n) \ne \tau(T_\infty,k_n)\bigr\} = \mathrm{TV}\bigl(\tau(T_n,k_n),\tau(T_\infty,k_n)\bigr) \to 0.$$

To show that $L_n/\log n \to \infty$ in probability, it suffices to show this for L_n', the maximal leaf-degree among all marked nodes of $\tau(T_\infty,k_n)$ at distance $< k_n$ from the root. Let $\zeta_0,\zeta_1,\ldots,\zeta_{k_n-1}$ be the degrees of the marked nodes in $\tau(T_\infty,k_n)$, indexed by their distance from the root, let λ_i be the leaf-degree corresponding to ζ_i. Now, fix a constant c and let A_i be the event that $\lambda_i \le c\log n$; we have

$$\begin{aligned}
\mathbf{P}\{L_n' \le c\log n\} &\le \mathbf{P}\biggl\{\bigcap_{i=0}^{k_n-1} A_i\biggr\}\\
&= \mathbf{P}\{A_0\}^{k_n-1} \qquad (17)\\
&\le \exp\bigl(-(k_n-1)\,\mathbf{P}\{\lambda_0 > c\log n\}\bigr).
\end{aligned}$$

Setting $k_n = \lceil n^{1/3} \rceil + 1$, we have

$$\mathbf{P}\{L'_n \le c \log n\} \le \exp\bigl(-n^{1/3}\,\mathbf{P}\{\lambda_0 > c\log n\}\bigr). \tag{18}$$

Note that $\lambda_0 \sim \text{Binomial}(\zeta_0 - 1, p_0)$, so that $\mathbf{P}\{\lambda_0 \le p_0\zeta_0/2 \mid \zeta_0\} \le 1/2$ for ζ_0 large enough, by the law of large numbers. Therefore, for n large enough, we have

$$\mathbf{P}\{\lambda_0 > c\log n\} \ge \mathbf{P}\bigl\{\lambda_0 \ge \frac{p_0\zeta_0}{2} > c\log n\bigr\} \ge \frac{1}{2}\,\mathbf{P}\biggl\{\zeta_0 > \frac{2c}{p_0}\log n\biggr\}. \tag{19}$$

To conclude the proof, we must show that $n^{1/3}\,\mathbf{P}\{\zeta_0 > 2c\log n/p_0\} \to \infty$ along a subsequence of n. Note that if $\mathbf{E}\{\rho^\xi\} = \infty$, then $\int_0^\infty \mathbf{P}\{\rho^\xi > x\}\,dx = \infty$, and thus

$$\sum_{\ell=1}^{\infty} 2^\ell\,\mathbf{P}\biggl\{\xi > \frac{\ell}{\log_2\rho}\biggr\} \ge \sum_{\ell=1}^{\infty} 2^\ell\,\mathbf{P}\{\rho^\xi > 2^\ell\} \ge \sum_{\ell=1}^{\infty}\int_{2^\ell}^{2^{\ell+1}} \mathbf{P}\{\rho^\xi > x\}\,dx = \infty,$$

and consequently, $\mathbf{P}\{\xi > \ell/\log_2\rho\} \ge \ell^{-2}2^{-\ell}$ for infinitely many $\ell \in \mathbf{N}$. As

$$\mathbf{P}\biggl\{\zeta > \frac{\ell}{\log_2\rho}\biggr\} \ge \frac{\ell}{\log_2\rho}\,\mathbf{P}\biggl\{\xi > \frac{\ell}{\log_2\rho}\biggr\},$$

we see that

$$\mathbf{P}\biggl\{\zeta > \frac{\ell}{\log_2\rho}\biggr\} \ge \frac{1}{\log_2\rho\cdot\ell 2^\ell} \tag{20}$$

for infinitely many ℓ. Setting $\ell = (2c/p_0)\log n\log_2\rho$, we have,

$$n^{1/3}\,\mathbf{P}\biggl\{\zeta > \frac{2c}{p_0}\log n\biggr\} \ge n^{1/3}\cdot\frac{1}{2^{2c\log n\log_2\rho/p_0}}\cdot\frac{1}{\log_2\rho\cdot 2c\log n\log_2\rho/p_0} \tag{21}$$

for infinitely many n provided that

$$\frac{2c}{p_0}\log 2\log_2\rho \le \frac{1}{6},$$

which is possible by making $\rho > 1$ small enough. Thus, for every $c > 0$,

$$\limsup_{n\to\infty}\mathbf{P}\{L'_n > c\log n\} = 1,$$

which is what we wanted to show. ∎

Note that if for every $\rho < 1$, $p_n > \rho^n$ for all n large enough, then $L_n/\log n \to \infty$ in probability (instead of just along a subsequence).

3.4. Examples

In this section we examine several important families of trees in the Galton–Watson context, and give explicit asymptotic upper and lower bounds for the multiplicity. In each case, the two important parameters will be

$$\gamma = \max_{k\geq 2} {p_0}^k {p_k}^{k/(k-1)} \qquad \text{and} \qquad \mathrm{H}_2 = \log_2 \frac{1}{\sum_{i\geq 0} {p_i}^2}.$$

We must also verify that $\mathbf{E}\{2^\xi\}$ is finite, if the upper bound is to hold. In particular, this latter condition always holds if ξ is bounded. A summary of this section is displayed in Table 2.

Full binary trees. These are trees in which every node must have exactly zero or two children, and arise from the distribution $p_0 = p_2 = 1/2$. We compute $\gamma = 1/16$ and $\mathrm{H}_2 = 1$, so that

$$(1-\epsilon)\frac{\log_2 n}{4} \leq S_n \leq (1+\epsilon)2\log_2 n \tag{22}$$

asymptotically in probability. Because the multiplicity in a full binary tree must be a power of 2, in essence this means that there exists a sequence of integers (a_n) such that

$$\mathbf{P}\big\{S_n \in \{2^{a_n}, 2^{a_n+1}, 2^{a_n+2}, 2^{a_n+3}\}\big\} \to 1.$$

In other words, in general one cannot improve the ratio between the upper and lower bounds in Theorem 3.1 to a factor of less than $8+\epsilon$.

Flajolet *t*-ary trees. Full binary trees are a special case of a Flajolet t-ary tree for $t = 2$. For general t, we have

$$\begin{aligned} \gamma &= {p_0}^t {p_t}^{t/(t-1)} \\ &= \left(1-\frac{1}{t}\right)^t \left(\frac{1}{t}\right)^{t/(t-1)} \\ &= \exp\big(-1 + o_t(1) - \log t\big), \end{aligned} \tag{23}$$

so $\log_2(1/\gamma) = \log_2 e + \log_2 t + o(1)$ as $t \to \infty$. On the other hand,

$$\mathrm{H}_2 = \log_2 \frac{1}{{p_0}^2 + {p_t}^2} = \log_2 \frac{1}{1 - 2/t + 2/t^2}, \tag{24}$$

so $\mathrm{H}_2 \sim 2\log_2 e/t$ as $t \to \infty$. This means that as t gets large, the ratio between the upper and lower bound grows as $t \log t$.

Cayley trees. These trees arise from a Poisson(1) distribution, where $p_i = 1/(e \cdot i!)$ for $i \geq 0$. We verify first that

$$\mathbf{E}\{2^\xi\} = \sum_{i=0}^{\infty} \frac{2^i}{ei!} = e < \infty,$$

Table 2

LEAF MULTIPLICITIES OF CERTAIN FAMILIES OF TREES

Family	Lower bound	Upper bound
Full binary (Uniform$\{0,2\}$)	$\frac{\log_2 n}{4}$	$2\log_2 n$
Flajolet t-ary ($p_0 = 1 - 1/t;\ p_t = 1/t$)	$\frac{\log_2 n}{\log_2 e + \log_2 t + o_{t\to\infty}(1)}$	$\sim_{t\to\infty} t\log n$
Cayley (Poisson(1))	$\frac{\log_2 n}{2 + 4\log_2 e}$	$\frac{2\log_2 n}{\log_2(e^2/(I_0(2))}$
Catalan (Binomial$(2, 1/2)$)	$\log_{256} n$	$\frac{2\log_2 n}{\log_2(8/3)}$
Binomial (Binomial$(d, 1/d)$)	$\frac{\log_2 n}{2 - \log_2((1-1/d)^{4d-2})}$	$\frac{2\log_2 n}{\log_2(e^2/(I_0(2)) + o_{d\to\infty}(1)}$
Motzkin (Uniform$\{0,1,2\}$)	$\log_{81} n$	$2\log_3 n$
Planted plane (Geometric$(1/2)$)	$\log_{256} n$	—

and then work out that $\gamma = 1/(4e^4)$. Letting

$$I_0(z) = \sum_{i=0}^{\infty} \frac{(i^2/4)^k}{i! \cdot \Gamma(z+1)} = \frac{1}{\pi}\int_0^{\pi} e^{z\cos\theta}\, d\theta$$

be the modified Bessel function of the first kind (see [1], p. 376), we find that

$$\sum_{i=0}^{\infty} {p_i}^2 = \frac{1}{e^2}\sum_{i=0}^{\infty}\frac{1}{(i!)^2} = \frac{1}{e^2} I_0(2), \tag{25}$$

meaning that $\mathrm{H}_2 = 2\log_2 e - \log_2\bigl(I_0(2)\bigr)$. Putting everything together, the lower and upper bounds in probability for S_n are, respectively,

$$\frac{\log_2 n}{2 + 4\log_2 e} \approx \frac{\log_2 n}{7.771} \tag{26}$$

and

$$\frac{2\log_2 n}{2\log_2 e - \log_2\bigl((1/\pi)\int_0^{\pi} e^{2\cos\theta}\, d\theta\bigr)} \approx \frac{\log_2 n}{0.8483}. \tag{27}$$

Catalan trees. When we set $p_0 = p_2 = 1/4$ and $p_1 = 1/2$, we obtain the family of Catalan trees. Recall that there is a one-to-one correspondence between Catalan trees on n nodes and full binary trees on $2n+1$ nodes, since one obtains a full binary tree from a Catalan tree by adding artificial external nodes to every empty slot, and this procedure is reversed by removing all leaves from a full binary tree. It is easy to see that the leaf multiplicity of a full binary tree is exactly double the multiplicity of its corresponding Catalan tree. By plugging in $d = 2$ above, the lower bound given by Lemma 3.3 is $\log_2 n/8$, which makes sense since the correspondence with full binary trees tells us that the lower bound on the Catalan trees should be similar to $\log_2(2n+1)/8$. We calculate $\mathrm{H}_2 = \log_2(8/3)$ and the upper bound is $2\log_2 n/\log_2(8/3)$, so the ratio between the upper and lower bounds is $16/\log_2(8/3)$.

Binomial trees. Catalan trees are a special case of a d-ary tree for $d \geq 2$, corresponding to a Binomial$(d, 1/d)$ distribution. We compute

$$\gamma = (p_0 p_2)^2 = \left(\left(\frac{d-1}{d}\right)^d \cdot \frac{d(d-1)}{2} \cdot \frac{(d-1)^{d-2}}{d^d}\right)^2 = \frac{1}{4}\left(1 - \frac{1}{d}\right)^{4d-2}. \tag{28}$$

Note that taking the limit $d \to \infty$, the Binomial$(d, 1/d)$ distributions approach a Poisson(1) distribution. Thus we see from our earlier discussion on the Cayley trees that $\mathrm{H}_2 = \log_2\big(e^2/(I_0(2))\big) + o_{d\to\infty}(1)$. This gives the respective lower and upper bounds

$$\frac{\log_2 n}{2 - (4d-2)\log_2(1 - 1/d)} \qquad \text{and} \qquad \frac{2\log_2 n}{\log_2\big(e^2/(I_0(2))\big) + o_{d\to\infty}(1)}. \tag{29}$$

The lower bound tends to $\log_2 n/(2 + 4\log_2 e)$ as $d \to \infty$, matching the lower bound we obtained for Cayley trees above.

Motzkin trees. These trees correspond to the distribution with $p_0 = p_1 = p_2 = 1/3$. We easily compute $\gamma = 1/81$ and $\mathrm{H}_2 = \log_2 3$, which yields an asymptotic lower bound of $\log_2 n/(\log_2 81) = \log_{81} n$ and an asymptotic upper bound of $2\log_2 n/\log_2 3 = 2\log_3 n$. The ratio between the upper and lower bounds is 8.

Planted plane trees. These are trees with ordered children, so that each can be embedded in the plane in a unique way. They correspond to a Geometric(1/2) distribution, with $p_i = 1/2^{i+1}$ for all i. We find that $\gamma = 1/256$, so we have the asymptotic lower bound $S_n \geq \log_2 n/8$. Unfortunately, we have $\mathbf{E}\{2^\xi\} = \sum_{i\geq 0} 1/2 = \infty$, so Lemma 3.2 cannot be applied to give an upper bound here. Recall from Lemma 1.6 that the maximal degree Δ_n of T_n satisfies $\Delta_n/\log_2 n \to 1$ in probability. However, this does not imply that $S_n = O(\log n)$ in probability.

3.5. Automorphic multiplicity

The multiplicity of a tree does not have a natural extension to unrooted trees, because whether or not two nodes are identical depends crucially on their position in relation to a distinguished root node u. In this section we briefly investigate an alternate notion of multiplicity that does extend nicely to free trees. It arises in the problem of root estimation in Galton–Watson trees described in Chapter 2.

We briefly recall some definitions. Let T be a rooted tree. By disregarding the parent-child directions of the edges, we obtain a free tree T_F. Conversely, if we start with a free tree T_F and any node u, we can define a *rooting of* T_F *at* u to be the rooted tree T_u obtained by fixing u as the root. This does not give rise to a unique tree in general, because children of a given node may hang on the wall in an arbitrary left-to-right order, but our new notion of multiplicity will treat all of these possible ordered trees the same.

Let $\mathrm{Aut}(T_\mathrm{F})$ be the group of all graph automorphisms of T_F, that is, bijections f from the set of vertices T_F to itself such that for vertices u and v, $f(u)$ is adjacent to $f(v)$ whenever u is adjacent to v. We can then define an *automorphism* of T_u to be a graph automorphism of T_F such that the root u stays fixed. By a slight abuse of notation, we denote the set of these rooted-tree automorphisms by $\mathrm{Aut}(T_u)$; formally this is the stabilizer subgroup $\mathrm{Stab}(u)$ of $\mathrm{Aut}(T_\mathrm{F})$. We will say that two nodes v and w in T_u are *congruent* and write $v \sim_u w$ if v and w belong to the same orbit under the action of $\mathrm{Aut}(T_u)$. This means that there exists an element f of $\mathrm{Aut}(T_u)$ such that $f(v) = w$. It is clear that this gives us an equivalence relation on the set of all nodes of T_u, and the *automorphic multiplicity* of a node v, denoted $\mu(u, v)$, is the size of the equivalence class of v under this relation. Since any node can be mapped to itself under an automorphism, $\mu(u, v) \geq 1$ for all v.

In fact, one can define the relation $\sim_u$, and consequently the function μ, purely in terms of the relation $\equiv$. We have $v \sim_u w$ if and only if there exists a permutation for every node in T_u such that applying each permutation to the left-to-right ordering of its respective node's children results in a tree in which $v \equiv w$. The analogue of S in this setting is the *automorphic (leaf) multiplicity* $M(T)$ of a rooted tree T. If o is the tree's root, then $M(T)$ is the maximum value of $\mu(o, v)$, taken over all nodes v in the tree.

Fig. 3.3. Different leaf multiplicities but the same automorphic leaf multiplicity.

Fig. 3.3 illustrates the distinction between the automorphic and non-automorphic multiplicity. We have $S(T_1) = M(T_1) = 4$, since the two non-leaf children of the root have identical (and therefore congruent) subtrees. In T_2, on the other hand, these subtrees are congruent but not identical, so that $M(T_2) = 4$ but the non-automorphic multiplicity of T_2 is only 2.

This definition is still somewhat at odds with the notion of multiplicity that arises in the root estimation problem from Chapter 2. In that setting, one considers all graph automorphisms of the free tree, not just ones that fix the root. We will call the size of the orbit of a node under this larger action the *free multiplicity* μ_F and if two nodes u and v are congruent under an arbitrary graph automorphism, then we write $u \sim_{\mathrm{F}} v$ and say that the two nodes are *free-congruent*. (The notation μ_{F} corresponds to the notation M in Chapter 2.). We also let $M_{\mathrm{F}}(T)$ denote the *free (leaf) multiplicity*, the maximum value of μ_{F} over all nodes in the free tree T_{F}.

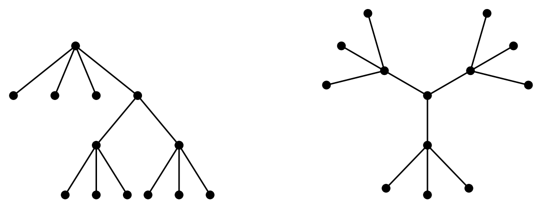


Fig. 3.4. A rooted tree T with $M(T) = 6$ and $M_{\mathrm{F}}(T) = 9$.

Fig. 3.4 shows the relation between the automorphic multiplicity of a rooted tree and the free multiplicity its free-tree counterpart. Note that $M(T) \leq M_{\mathrm{F}}(T)$ for any rooted tree T, since we have $\mu(u) \leq \mu_{\mathrm{F}}(u)$ for every node u. We shall spend the rest of this section showing that this inequality can more or less be reversed. First, we need three lemmas, the latter two of which appeared in a preliminary version of [9]. They were removed from the paper in a revision as they were unnecessary for the root estimation problem, but they will be useful to us here, so we have granted them a reprieve. For the next three lemmas, when we say "multiplicity", we mean "free multiplicity". First, we give a new lemma, whose proof was shown to us by J. Saks.

Lemma 3.5. *If u and v are adjacent nodes in a finite free tree T, then either $\mu_{\mathrm{F}}(u)$ is an integer multiple of $\mu_{\mathrm{F}}(v)$ or the other way around.*

Proof. We may reduce to the case where one of u or v is a leaf. This is because if neither is a leaf, then it is not in the orbit of any leaf by graph automorphism, so we can remove all the leaves from the tree T without changing either of $\mu_{\mathrm{F}}(u)$ or $\mu_{\mathrm{F}}(v)$. (The anxious reader may worry that $\mu_{\mathrm{F}}(v) = \mu(l, v)$ for some leaf $l \neq v$, but then we also have $\mu_{\mathrm{F}}(v) = \mu(l', v)$ where l' is the only neighbour of l.) This is done finitely many times since T is finite and always contains at least one leaf.

Now without loss of generality, suppose u is the leaf and v is its unique neighbour. By the orbit-stabiliser theorem,

$$\left|\mathrm{Aut}(T)\right| = \mu_{\mathrm{F}}(u)\left|\mathrm{Stab}(u)\right| = \mu_{\mathrm{F}}(v)\left|\mathrm{Stab}(v)\right|, \tag{30}$$

where stabilisers are taken with respect to the group $\mathrm{Aut}(T)$. Every automorphism fixing u must permute its neighbours, but since u only has one neighbour, we see that $\mathrm{Stab}(u) \subseteq \mathrm{Stab}(v)$. Hence, denoting the index of a subgroup H of

G by $[G : H]$, we have

$$\begin{aligned}\mu_{\mathrm{F}}(u) &= \frac{\mu_{\mathrm{F}}(v)\bigl|\mathrm{Stab}(v)\bigr|}{\bigl|\mathrm{Stab}(u)\bigr|} \\ &= \frac{\mu_{\mathrm{F}}(v)\bigl[\mathrm{Stab}(v) : \mathrm{Stab}(u)\bigr]\bigl|\mathrm{Stab}(u)\bigr|}{\bigl|\mathrm{Stab}(u)\bigr|} \\ &= \mu_{\mathrm{F}}(v)\bigl[\mathrm{Stab}(v) : \mathrm{Stab}(u)\bigr],\end{aligned} \tag{31}$$

proving the lemma. ▮

The next lemma formalises the intuitive notation that in a free tree, the multiplicities are in some sense smaller towards the centre of the tree.

Lemma 3.6. *Let u — v — w be neighbouring nodes in a free tree T with v being the middle node. Then v cannot have strict maximal free multiplicity among the three nodes; that is, $\mu_{\mathrm{F}}(v) \le \mu_{\mathrm{F}}(u)$ or $\mu_{\mathrm{F}}(v) \le \mu_{\mathrm{F}}(w)$.*

Proof. Suppose for contradiction that $\mu_{\mathrm{F}}(v) > \mu_{\mathrm{F}}(u)$ and $\mu_{\mathrm{F}}(v) > \mu_{\mathrm{F}}(w)$. Then, for each of the pairs of neighbours u — v and v — w, the multiplicity of one of the nodes must be an integer multiple of the multiplicity of the other, by the previous lemma.

So there must be integers $r, s > 1$ such that

$$\mu_{\mathrm{F}}(v) = r\mu_{\mathrm{F}}(w) \quad \text{and} \quad \mu_{\mathrm{F}}(v) = s\mu_{\mathrm{F}}(u). \tag{32}$$

The situation is illustrated in Fig. 3.5. Since $\mu_{\mathrm{F}}(v) = s\mu_{\mathrm{F}}(u)$, in the v-rooted tree u must have $s-1$ children, each of which is in the orbit of v and thus is the the root of a subtree isomorphic to B. Similarly, since $\mu_{\mathrm{F}}(v) = r\mu_{\mathrm{F}}(w)$, w must have $r-1$ child subtrees isomorphic to A.

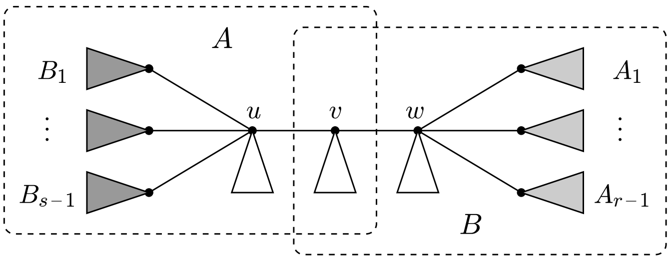


Fig. 3.5. Three adjacent nodes and their subtrees.

We note that in order to satisfy the $r, s > 1$ requirements, we must have

$$|A| \ge (s-1)|B| + 2 \quad \text{and} \quad |B| \ge (r-1)|A| + 2, \tag{33}$$

where the additional $+2$ terms come respectively from nodes u and v (for $|A|$) or v and w (for $|B|$). This implies that

$$|A| \ge (s-1)(r-1)|A| + 2s,$$

which is impossible if $|A| \geq 1$ and $r, s > 1$. The contradiction tells us that v cannot have strict maximal multiplicity among the three nodes. ▌

We have established that if we embed a free tree into the (x, y)-plane and then lift the nodes up by setting each node's z-coordinate to its multiplicity, then the result is a convex, spidery bowl or valley. This is illustrated in Fig. 3.6.

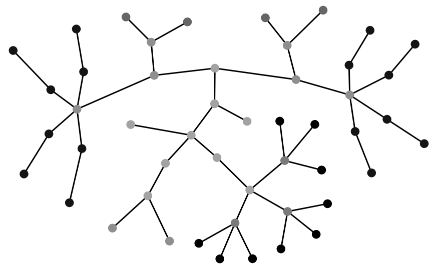


Fig. 3.6. Darker shades of grey indicate higher multiplicities in this free tree.

On a path between any two endpoints, the multiplicities decrease monotonically towards the centre of the tree before increasing monotonically towards the endpoint. There is a central connected core of nodes of minimal multiplicity and we are able to show that this minimal multiplicity cannot be greater than 2.

Lemma 3.7. *If $F = (V, E)$ is a finite free tree, then the node of minimal multiplicity in F has multiplicity 1 or 2.*

Proof. The proof is by contraposition. Let $u \in V(F)$ be a node of minimal multiplicity and suppose $\mu_{\mathrm{F}}(u) > 2$. Let C_u be the orbit of u. There is a subtree F' whose endpoints are the members of C_u; since $m > 2$ and the graph is connected, there is necessarily at least one node $v \in F' \setminus C_u$. By Lemma 3.6, we have $\mu_{\mathrm{F}}(v) \leq \mu_{\mathrm{F}}(u)$ but by minimality of $\mu_{\mathrm{F}}(u)$, we know that $\mu_{\mathrm{F}}(v) = \mu_{\mathrm{F}}(u)$. So we can repeat the argument with C_v to find that the tree is infinite (at each step we are removing $\mu_{\mathrm{F}}(u)$ nodes from the free tree, but the process never terminates).

Note that this argument does not work when $\mu_{\mathrm{F}}(u) = 2$ because F' may simply consist of two nodes connected by one edge. ▌

Theorem 3.8. *Let T be a rooted tree with n nodes; let $M(T)$ and $M_{\mathrm{F}}(T)$ be the automorphic multiplicity and free multiplicity of T, respectively. We have the inequality*

$$M_{\mathrm{F}}(T) \leq 2M(T),$$

and this bound is the best possible, in general.

Proof. Suppose first that $n \geq 3$. Let v be a leaf of maximal automorphic multiplicity in T_{F}, and let $[v]$ denote the set of nodes that are free-congruent to v (so $\big|[v]\big| = M_{\mathrm{F}}(T)$). By Lemma 3.7, a node s of minimal automorphic multiplicity either has $\mu_{\mathrm{F}}(s) = 1$ or $\mu_{\mathrm{F}}(s) = 2$, and since we assumed that $n \geq 3$, we can require that s not be a leaf.

If $\mu_{\mathrm{F}}(s) = 1$, then $M(T_s) = M_{\mathrm{F}}(T)$, since any automorphism of T_{F} already fixes s. The nodes in $[v]$ all lie in some subtrees of s, and without loss of generality, we may assume that they do not all lie in the same subtree, since if s' is the only child of s whose subtree contains nodes of $[v]$, we can reroot the tree T_s at s' instead without changing the maximum automorphic multiplicity. There are $d \geq 2$ children of the root whose subtrees contain elements of $[v]$; each one contains an equal proportion of these nodes, so d divides $M_{\mathrm{F}}(T)$. If we reroot the tree at any node outside these subtrees, then the automorphic multiplicity of the tree does not change. If, on the other hand, we choose a node in one of these subtrees, then there are still $(d-1)M_{\mathrm{F}}(T)/d$ leaves that can still be shuffled amongst themselves, so the maximum automorphic multiplicity is $(d-1)M_{\mathrm{F}}(T)/d \geq M_{\mathrm{F}}(T)/2$.

If $\mu_{\mathrm{F}}(s) = 2$, there is a node s' that is free-congruent to s, and there is mirror symmetry in the graph. This means that there is a way to split the graph along an edge such that the two sides have the exact same shape, one contains s, and the other contains s'. The side containing s has $M_{\mathrm{F}}(T)/2$ members of $[v]$; call this half $[v]_s$ and the other half $[v]_{s'}$. When the tree is rooted at s, we find that $M(T_s) = M_{\mathrm{F}}(T)/2$, since any two members of $[v]_s$ can be exchanged and any two members of $[v]_{s'}$ can be exchanged (but exchanges cannot happen between the two subtrees). And rerooting the tree at an arbitrary node, it is clear that the automorphic multiplicity of the tree will not decrease.

When $n = 1$ the statement is trivial, and taking $n = 2$ shows that the bound is the best possible, because if T is the tree with a root and a single (leaf) child, then $M_{\mathrm{F}}(T) = 2$ and $M(T) = 1$. ▮

This theorem tells us that the asymptotics of the free multiplicity are the same as the asymptotics of the automorphic multiplicity, up to a fudge factor of 2. Because congruence of two nodes is immediately implied by their being identical under $\equiv$, we have $S(T) \leq M(T)$ for all rooted trees T. Thus if $M_n = M(T_n)$ and $F_n = M_{\mathrm{F}}(T_n)$, where T_n is a conditional Galton–Watson tree of size n, then if γ is as defined in Lemma 3.3, in probability the lower bound

$$F_n \geq M_n \geq (1-\epsilon)\frac{\log_2 n}{\log_2(1/\gamma)} \tag{34}$$

on the automorphic and free multiplicities holds with probability tending to 1. A strengthening of this bound alongside a corresponding upper bound would be a significant contribution.

CHAPTER FOUR

THE INDEPENDENCE NUMBER

4.1. Introduction

THE INDEPENDENCE NUMBER is a fundamental graph invariant that arises often in computational complexity theory and the analysis of algorithms. In a graph $G = (V, E)$, a subset $S \subseteq V$ of vertices is said to be an *independent set* if no two elements of S are adjacent. The dual notion is that of a *vertex cover*, namely a subset $C \subseteq V$ such that every edge in G has an endpoint in C. The *independence number* $I(G)$ of G is defined to be the size of the largest independent set in G. In this chapter, we concern ourselves with the case $G = T$, a random tree in the Galton–Watson model. In recent years, analysis of the independence number of trees has been carried out for various other random models. C. Banderier, M. Kuba, and A. Panholzer studied various families of simply-generated trees [5], and a recent paper of M. Fuchs, C. Holmgren, D. Mitsche, and R. Neininger considers random binary search trees as well as random recursive trees [25].

Because every tree T is bipartite, the independence number $I(T)$ is always at least $|T|/2$ (we take the larger element of the bipartition). Recall that a vertex set S is a *vertex cover* of T if every edge of T intersects a vertex in S. Letting $V(T)$ denote the size of a minimum-cardinality vertex cover, we have the formula $n = V(T) + I(T)$. In a tree, there always exists a maximum-cardinality independent set that includes all of the leaves, and the following algorithm, which will be the starting point of our discussion, uses this fact to find an independent set of maximum size. Note that this is only one of many possible algorithms that accomplishes this task.

Algorithm I (*Independent set*). Given a directed tree T, this algorithm computes a maximum-cardinality independent set A of vertices.

I1. [Initialize.] Set $A \leftarrow \emptyset$.

I2. [Compute leaves and parents.] Let $L(T)$ be the set of leaves of T, that is, the set of vertices with out-degree 0. Let $P(T)$ be the set of parents of nodes in $L(T)$.

I3. [Update.] Set $A \leftarrow A \cup L(T)$ and $T \leftarrow T \setminus L(T) \setminus P(T)$. (At this stage, T may now be a forest.)

I4. [Loop?] If $T = \emptyset$, halt and output A; otherwise, return to step I2. ▌

Algorithm I repeatedly peels away leaves and their parents to arrive at what we shall call the *layered independent set*. We refer to $L(T)$ as layer 0, to $P(T)$ as layer 1, to $L\big(T \setminus L(T) \setminus P(T)\big)$ as layer 2, and so on. In this manner, each node u gets assigned a *peel number* $\rho(u)$, the layer number of the set to which it belongs. The peel number $\rho(T)$ of a tree T is the peel number of the root of T. We also let $m(T)$ denote the maximum of the peel numbers of vertices in T; this quantity is twice the number of loops that Algorithm I undergoes before termination, rounded up. Note that all the peel numbers can be computed by postorder traversal of the tree in time $O(|T|)$, and then the layered independent set is simply the collection of all nodes with even peel number.

A quantity related to the peel number is the *leaf height* $\lambda(u)$ of a node $u \in T$. It is the length of the path to the nearest leaf in the (fringe) subtree rooted at u. The leaf height $\lambda(T)$ of a tree T is the maximal leaf height of any node in T. The fact that $\rho(u) = k$ implies that there is a leaf at depth k from the root, so $\lambda(u) \leq \rho(u)$ for all nodes u in a tree. It is also easily seen that for any tree T, $\lambda(T) \leq \rho(T) \leq m(T)$. A small example is given in Fig. 4.1; note that for nodes with few children or small subtrees, the two quantities are quite similar. One corollary of our main results is that under certain conditions, this phenomenon persists as n gets large, that is, the peel number and leaf height have the same order of asymptotic growth.

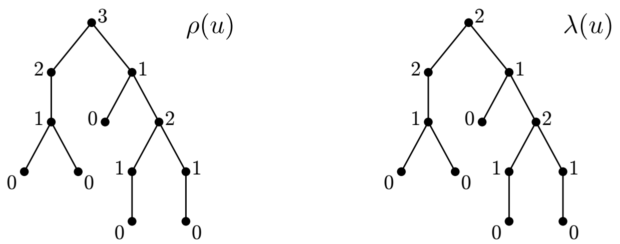


Fig. 4.1. Peel numbers (left) and leaf heights (right) of nodes.

The leaf height goes by the name *protection number* in the literature and has enjoyed some recent attention. With this usage, a node whose minimal distance from any leaf is k is called *k-protected* and a 2-protected node is often simply said to be *protected*. In this chapter we say that a node has leaf height k, which we believe is more illustrative than saying it is k-protected. The number of nodes with leaf height ≥ 2 was examined by G.-S. Cheon and L. W. Shapiro for planted plane trees, Motzkin trees, full binary trees, Catalan trees, and ternary trees [15]; by T. Mansour [43] for k-ary trees; by R. R. X. Du and and H. Prodinger for digital search trees [20]; by H. M. Mahmoud and M. D. Ward for binary search trees [41] and for random recursive trees [42]; and by L. Devroye and S. Janson, who considered simply generated trees and also unified some earlier results regarding binary search trees and random recursive trees. Nodes with leaf height > 2 were studied in binary search trees by M. Bóna [8] and in planted plane trees by K. Copenhaver [17]. In the setting of simply generated trees and Pólya trees, the leaf height of the root as well as the leaf height of a node chosen uniformly at random was studied in [27].

The main results of this chapter characterize the asymptotic behaviour of the independence number $I_n = I(T_n)$, the maximum peel number $M_n = m(T_n)$, and the maximum leaf height $L_n = \lambda(T_n)$ for a Galton-Watson tree T_n, which is conditioned on having n nodes. We also include distributional properties of closely related statistics, such as the peel number and leaf height of the root of an unconditional Galton-Watson tree, as well as the leaf height L'_n of the root of and the leaf height L''_n of a node chosen uniformly at random in a conditonal Galton–Watson tree.

4.2. Asymptotics of the independence number

We begin by studying unconditional Galton–Watson trees. Recall that the generating function $f(s)$ of an offspring distribution ξ is the infinite series $\mathbf{E}\{s^\xi\}$, which converges absolutely when $0 \le s \le 1$. We can thus differentiate to obtain $f'(s) = \mathbf{E}\{\xi s^{\xi-1}\}$. A quantity that will play a key role in our story is q, the unique solution in $(0,1)$ of $q = f(1-q)$.

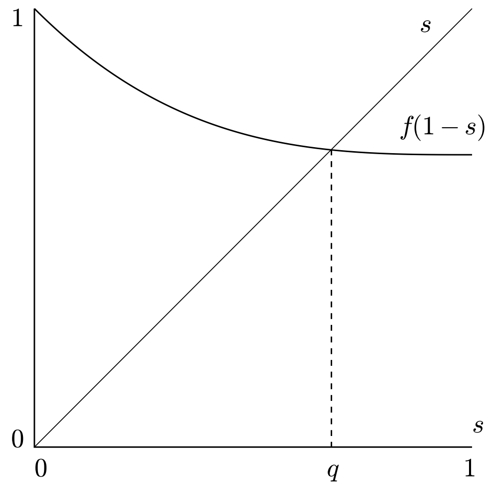


Fig. 4.2. The parameter q satisfying $q = f(1-q)$.

Lemma 4.1. *Let ξ be an offspring distribution with $0 < \mathbf{E}\{\xi\} \le 1$ and nonzero variance, and let $f(s) = \mathbf{E}\{s^\xi\}$. The probability that the root of a Galton–Watson tree T with this distribution belongs to the layered independent set is q, which belongs to the interval $(1/2, 1)$.*

Proof. Note that q is the probability that all the children of the root are not in the layered independent set. By the recursive definition of a Galton–Watson tree, we have

$$q = \sum_{i \ge 0} p_i (1-q)^i = f(1-q). \tag{1}$$

Since $f(s) = \mathbf{E}\{s^\xi\}$ is nondecreasing, $f(1-s)$ is nonincreasing and s is increasing, hence there is a unique solution to $s = f(1-s)$ in the compact interval $[0,1]$.

Of course, q cannot be 1 since $\mathbf{P}\{\xi = 0\} \neq 0$. The fact that $f(s) > s$ for all $x \in (0,1)$ implies that $q = f(1-q) > 1-q$, hence $q > 1/2$. ▮

This lemma is essentially known (see, e.g., Banderier et al. [5]).

Examples. There is a well-known connection between certain families of trees and conditioned Galton–Watson trees. In each of the following cases, sampling a conditional Galton–Watson tree T_n with the given distribution is equivalent to uniformly sampling a tree of size n from the respective tree family.

i) In *Flajolet's t-ary tree*, every node is either a leaf or has t children (see [24], p. 68). This corresponds to the distribution with $p_0 = 1-1/t$ and $p_t = 1/t$, so we can compute q numerically by finding the unique solution to the equation

$$q = 1 - \frac{1}{t} + \frac{(1-q)^t}{t}. \tag{2}$$

in the interval $(1/2, 1)$. In the case $t = 2$ of full binary trees, we find that $q = 2 - \sqrt{2} \approx 0.585786$, and since the $(1-q)^t/t$ term is very small for larger values of t, q is approximately $1 - 1/t$ for large t.

ii) To obtain a random rooted *Cayley tree*, we set $p_i = (i!e)^{-1}$ for all $i \geq 0$. Since $f(s) = e^{s-1}$, we have $qe^q = 1$, which we can invert in terms of the *Lambert W function*. Concretely, we have

$$q = W(1) = \left(\int_{-\infty}^{\infty} \frac{dt}{(e^t - t)^2 + \pi^2} \right)^{-1} - 1 \approx 0.567143, \tag{3}$$

which is also known as the *omega constant*.

iii) *Planted plane trees* correspond to the distribution $p_i = 1/2^{i+1}$ for $i \geq 0$. In this case, $f(s) = 1/(2-s)$, yielding the equation $q^2 + q - 1 = 0$, whose solution in the correct range is $q = 1/\varphi \approx 0.618034$. (The golden ratio $\varphi = 1.618034$ is the more famous solution to this quadratic equation).

iv) *Motzkin trees*, also known as unary-binary trees, are trees in which every non-leaf node has either one tree or two children. This corresponds to the distribution $p_0 = p_1 = p_2 = 1/3$ and $p_i = 0$ for all $i \geq 3$. So we have $q = \big(1 + (1-q) + (1-q)^2\big)/3$ and we have $q = 3 - \sqrt{6} \approx 0.550510$.

v) A *binomial tree* of order d can be thought of as a tree in which every node has d "slots" for its children, some of which may be filled. Thus a node can have r children in $\binom{d}{r}$ different ways, for $0 \leq r \leq d$. This corresponds, fittingly, to a binomial offspring distribution, where

$$p_i = \binom{d}{i} \left(\frac{1}{d}\right)^i \left(1 - \frac{1}{d}\right)^{d-i}, \tag{4}$$

for $0 \leq i \leq d$, and $p_i = 0$ otherwise. For this distribution, we have $f(s) = (s/n + 1 - 1/n)^n$, meaning that

$$q = \left(1 - \frac{1}{d} + \frac{1-q}{d}\right)^d = \left(1 - \frac{q}{d}\right)^d. \tag{5}$$

For large d, this tends to the omega constant. An important case is $d = 2$, which produces a random *Catalan tree*; it can be readily computed that $q = 4 - 2\sqrt{3} \approx 0.535898$ for these trees.

The following theorem shows the link between q and the size of the largest independent set in a conditioned Galton–Watson tree.

Theorem 4.2. *Let ξ be an offspring distribution with $0 < \mathbf{E}\{\xi\} \leq 1$ and $\mathbf{V}\{\xi\} \neq 0$, and let $f(s) = \mathbf{E}\{s^\xi\}$. The independence number $I_n = I(T_n)$ of a Galton–Watson tree, conditioned on having n nodes, satisfies*

$$\frac{I_n}{n} \to q$$

in probability as $n \to \infty$, where q is the unique solution in $(1/2, 1)$ of the equation $q = f(1-q)$.

Proof. For a vertex u, we let Γ_u denote the set of children of u and let

$$g(u) = \begin{cases} 1, & \text{if the peel number of } u \text{ is even;} \\ 0, & \text{otherwise} \end{cases}. \tag{6}$$

Note that the recursive function

$$G(u) = g(u) + \sum_{v \in \Gamma_u} G(v) \tag{7}$$

is exactly the independence number of the subtree rooted at u. Since g is bounded, we can apply a result of S. Janson ([33], Theorem 1.3) to find that for a conditional Galton–Watson tree T_n with root u,

$$\frac{I_n}{n} = \frac{G(u)}{n} \to \mathbf{E}\big\{f(T)\big\} = \mathbf{E}\big\{g(u)\big\} = q \tag{8}$$

in probability as $n \to \infty$. ▌

Note that examples (ii), (iii), and the Catalan case agree with explicit computations given in [5]. For simply generated trees, that paper, which uses singularity analysis, derives the constant q in a different manner, proves the stronger statement $\mathbf{E}\big\{I(T_n)\big\} = qn + O(1)$, and also gives a formula for the variance in terms of the degree-weight generating function. In particular, they show there exists a constant ν depending on the family of trees such that the variance is $\nu n + O(1)$.

4.3. Minimum-size s-path covers

This section represents a brief digression, and will not be related to our remaining results, though it discusses the natural generalisation of Algorithm I and is related to the open problem we give at the end of the chapter. As mentioned in the introduction, the size $V(T)$ of the minimal vertex cover of a tree T with n nodes has size $n - I(T)$, where $I(T)$ is the independence number. In particular, Algorithm I outputs a minimum-cardinality vertex cover alongside the maximum-cardinality independent set; it is the set of all nodes with odd peel number. We now tackle a more general notion of vertex covers. For an integer $s \geq 2$, an *s-path vertex cover* of a rooted tree T is a subset C of vertices such that any directed path of length $s-1$ in the tree contains a vertex in C. Thus the common-or-garden vertex cover corresponds to $s = 2$. (The off-by-one quirk in the definition goes away if we measure a path not by its length, but instead by its *order*, that is, the number of vertices it contains.) Note that only directed paths are considered; so two children of the same node are *not* connected by a path of length 2.

One might be tempted to generalize our earlier observations by claiming that the set of nodes with peel number congruent to $s-1$ modulo s is a minimal s-path vertex cover. This is not true! Consider a tree in which the root has two children, and one of the children has itself one child. Then no node has peel number equal to 2, but of course, the minimal 3-path vertex cover consists of the root. Towards a correct generalisation, consider the fact that if in every loop of Algorithm I, we removed all subtrees of height exactly 1, then the roots of these removed subtrees are precisely the vertices with odd peel number. Thus we arrive at an algorithm for computing a minimal s-path vertex cover.

Algorithm P (*Compute s-path vertex cover*). Given $s \geq 2$, and a rooted tree T, this algorithm computes a minimal s-path vertex cover C.

P1. [Initialize.] Set $C \leftarrow \emptyset$.

P2. [Done?] If there are no subtrees with height at least $s-1$, we output C and terminate.

P3. [Prune a subtree.] Let v be a node in T such that the subtree T_v rooted at v has height exactly $s-1$. We set $C \leftarrow C \cup \{v\}$ and set $T \leftarrow T \setminus T_v$ (this may now be a forest). Return to step P2. ▮

Note that if the original tree had height less than $s-1$, the algorithm outputs the empty set, which is a valid cover, since there are no paths of length $s-1$ in the tree. The fact that this algorithm actually does output a minimum-size vertex cover is proved in [11], and it is also remarked that the algorithm can be made to run in $O(|T|)$ time.

Let $V_s(T)$ denote the size of the minimum s-path vertex cover of a Galton–Watson tree T. To determine this random quantity, we will have to determine the probability that a node is added to the set C in Algorithm P. We will say that a vertex $v \in T$ is "marked" if Algorithm P adds it to the cover C. The following lemma gives necessary and sufficient conditions for the root of a tree to be marked.

Lemma 4.3. *The root u of a tree is marked if and only if there exists a path of length $s-1$ from the root that contains no marked vertices (other than the root).*

Proof. Suppose that the root u is marked. This means that in the final iteration of Algorithm P, after all other marked nodes have been removed, the tree has height $s-1$. This means that some unmarked node v is at depth $s-1$, and no node is marked on the path to this node. (This happens when v is a leaf or all children of v are marked, since if a child of v is unmarked, then we have an unmarked path of length s in the tree and the algorithm would have to mark some node on this path before marking the root.) Conversely, if such a path exists, then Algorithm P will be in this state in the final iteration of the loop and will therefore mark the root. ∎

This observation can be used to derive a functional equation for the probability that a node in an unconditional Galton–Watson tree is marked, as the following lemma shows.

Lemma 4.4. *Let T be a Galton–Watson tree with offspring distribution ξ satisfying $\mathbf{E}\{\xi\} \le 1$. Let $f(z) = \mathbf{E}\{\xi^z\}$ be the generating function of the distribution and let*

$$g(z,q) = 1 - f\bigl(q + (1-q)z\bigr) \tag{9}$$

The probability q_s that the root of the tree is in the minimum s-path vertex cover produced by Algorithm P satisfies

$$q_s = g(g(\cdots(g(0,q_s)\cdots,q_s),q_s),q_s), \tag{10}$$

where the function g is iterated $s-1$ times.

Proof. For $1 \le j < s$, let E_j be the event that in an unconditional Galton–Watson tree, there is a path of length j from the root that contains no marked vertices (except possibly the root). Thus $q_s = \mathbf{P}\{E_{s-1}\}$. Restating things slightly, E_j is the probability that there exists an unmarked child v of the root in whose subtree E_{j-1} is true. If the degree of the root is i, then the probability that all the children of the root fail to have this property is $\bigl(q_s + (1-q_s)\,\mathbf{P}\{E_{j-1}\}\bigr)^i$, so for $1 < j < s$,

$$\mathbf{P}\{E_j\} = \sum_{i\ge 0} p_i\bigl(1-(q_s+(1-q_s)\,\mathbf{P}\{E_{j-1}\})^i\bigr) = 1-f\bigl(q_s+(1-q_s)\,\mathbf{P}\{E_{j-1}\}\bigr). \tag{11}$$

Note that $\mathbf{P}\{E_1\}$ is simply the probability $1-f(q_s) = g(0,q_s)$ that one of the children of the root is unmarked, so unravelling the above equation proves the lemma. ∎

Note that when $s=2$, $q_s = 1-q$, where q is the solution to $z = f(1-z)$ we studied earlier. By a recursive computation analogous to the one we performed for the independence number, we find that if $V_s(T_n)$ denotes the minimum size of an s-path vertex cover of the conditional Galton–Watson tree T_n, then as $n \to \infty$,

$$\frac{V_s(T_n)}{n} \to q_s, \tag{12}$$

in probability. The function g given by Lemma 4.4 is rather unwieldy, so we cannot hope to find neat closed forms for the limit of $V_s(T_n)$ like we did for I_n in many special cases. However, we can, in principle, use g to numerically approximate the s-path vertex cover number for arbitrary distributions satisfying $\mathbf{E}\{\xi\} \le 1$.

***An application to random ideals.** We close this section by mentioning a link between these results and certain notions in commutative algebra. We will give all definitions necessary to describe the connection, but with minimal detail, as they will not be used outside this subsection. The concepts here are treated by most introductory commutative algebra textbooks (e.g., [22]). Let k be an arbitrary field and let $R = k[x_1, x_2, \ldots, x_n]$ be the polynomial ring on n variables with coefficients in k. A *squarefree monomial ideal* is an ideal generated by monomials of the form $\prod_{i \in S} x_i$ for some $S \subseteq \{1, 2, \ldots, n\}$. Regarding a rooted tree T as a directed graph on the vertex set $\{1, 2, \ldots, n\}$ with edges leading away from the root, the *path ideal* $I_s(T)$, for $s \ge 2$ is the ideal generated by the set of monomials

$$\big\{x_{i_1} x_{i_2} \cdots x_{i_s} : (i_1, i_2, \ldots, i_s) \text{ is a path of length } s \text{ in } T\big\}.$$

When $s = 2$, this is called an *edge ideal*.

Recall that a proper ideal $\mathfrak{p}$ of a ring R is *prime* if $ab \in \mathfrak{p}$ implies that $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. The *Krull dimension* $\dim R$ of a ring R is the supremum of lengths of chains of primes contained in R. If I is an ideal of R, the dimension $\dim I$ of I is defined to be the dimension of the quotient ring R/I. This definition is intended to reflect the fact that if S is the set of all points $(a_1, a_2, \ldots, a_n)$ in n-dimensional affine space over k such that $f(a_1, a_2, \ldots, a_n) = 0$ for all $f \in I$, then S has the same dimension as I. The *codimension* (or *height*) of a prime ideal $\mathfrak{p}$ is the supremum of lengths of chains descending from $\mathfrak{p}$; for instance, $\mathfrak{p} = (x_2, x_5, x_7)$ has codimension 3 because we have the chain $\mathfrak{p} \supset (x_2, x_5) \supset (x_2)$. More generally, the codimension of an ideal I is defined to be the minimum codimension of a prime containing I. For a polynomial ring R over a field and any ideal I of R, $\dim R = \dim I + \operatorname{codim} I$.

The main connection between path ideals and s-path vertex covers of a tree T is that the codimension of $I_s(T)$ is exactly the minimum size of an s-path vertex cover. To see this, note that every squarefree monomial ideal admits a primary decomposition into an intersection of prime ideals generated by subsets of the variables in the following manner. For every generating element that is expressible as ab for a and b relatively prime, we can write $I = \big(I + (a)\big) \cap \big(I + (b)\big)$; $I + (a)$ is the result of adding the element a to the generating set of I. This is best illustrated by an example. Let $I_2(T) = (x_1x_2, x_1x_3, x_2x_4, x_2x_5, x_3x_6)$ be the edge ideal of a certain tree T (the reader may want to draw the tree and play along). We can decompose $I_2(T) = (x_2, x_3) \cap (x_1, x_2, x_6) \cap (x_1, x_3, x_4, x_5) \cap (x_1, x_4, x_5, x_6)$. The prime ideal of minimal dimension that contains $I_2(T)$ is (x_2, x_3), which corresponds to the fact that $\{x_2, x_3\}$ is a vertex cover of the graph. In the case when $s = 2$, the maximum size of an independent set in T thus corresponds to the dimension of $I_2(T)$. This does not hold for $s > 2$, but we still have $\operatorname{codim} I_s(T) = V_s(T)$. As an example, for the tree T from the example above, the largest prime in the primary decomposition of $I_3(T)$ is (x_1), since every path of length 2 meets the root x_1.

There has been much recent work in determining various algebraic properties of random monomial ideals. For example, a 2019 paper of J. A. de Loera et al. characterised thresholds for the Krull dimension of random ideals arising from an Erdős-Rényi-type model [39]. Although path ideals of trees have been studied, notably in an M.Sc. thesis of J. He [30], to our knowledge no one has studied properties of path ideals of trees picked uniformly at random from certain families (which is equivalent to conditional Galton–Watson trees). We have shown in this section that under such a model, one can compute the codimension and Krull dimension of the random path ideal using only the generating function of the Galton–Watson offspring distribution.

4.4. Distribution of the peel number

Let r_i denote the probability that the root of an unconditional Galton–Watson tree has peel number i. In this section, we shall compute the distribution $(r_i)_{i\geq 0}$. It will also be convenient to set $r_i = 0$ when i is negative. We will establish the notation

$$r_i^+ = \sum_{j\geq i} r_j \qquad \text{and} \qquad r_i^- = \sum_{j=0}^{i} r_j\,. \tag{13}$$

There will be some asymmetry for odd and even i, so let us write $r_i^{+\text{odd}}$ for the subsum of r_i^+ consisting of odd terms and $r_i^{+\text{even}}$ for the subsum of r_i^+ consisting of even terms. Defining $r_i^{-\text{odd}}$ and $r_i^{-\text{even}}$ similarly, we have, of course, $r_i^{+\text{odd}} + r_i^{+\text{even}} = r_i^+$ and $r_i^{-\text{odd}} + r_i^{-\text{even}} = r_i^-$.

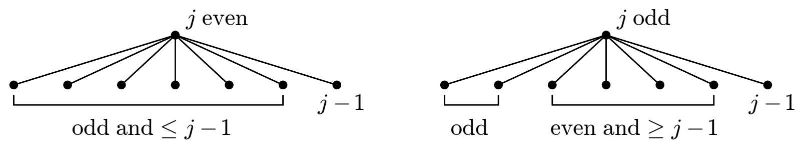


Fig. 4.3. Children of nodes with even and odd peel numbers.

Clearly, $r_0 = p_0$. For even indices j, all children must have an odd peel number at most $j-1$ and at least one must have peel number $j-1$. Thus, if ξ is the number of children at the root, then for $i \geq 1$,

$$r_{2i} = \mathbf{E}\bigl\{\bigl(r_{2i-1}^{-\text{odd}}\bigr)^\xi - \bigl(r_{2i-3}^{-\text{odd}}\bigr)^\xi\bigr\} = f\bigl(r_{2i-1}^{-\text{odd}}\bigr) - f\bigl(r_{2i-3}^{-\text{odd}}\bigr). \tag{14}$$

For odd indices j, all the children of the root with even peel number must have peel number at least $j-1$, and at least one must have peel number $j-1$. Since

$$\sum_{i\geq 0} r_{2i-1} = 1-q \qquad \text{and} \qquad \sum_{i\geq 0} r_{2i} = q, \tag{15}$$

we find that for $i \geq 1$,

$$r_{2i-1} = \mathbf{E}\big\{\big(1-q+r_{2i-2}^{+\text{even}}\big)^{\xi} - \big(1-q+r_{2i}^{+\text{even}}\big)^{\xi}\big\}. \tag{16}$$

The following lemma describes r_i for large i.

Lemma 4.5. *Let r_i be the probability that the root of an unconditional Galton–Watson tree with offspring distribution $\xi \sim (p_i)_{i\geq 0}$ has peel number equal to i. As $i \to \infty$, we have*

$$r_i = f'(1-q)^{i+o(i)}. \tag{17}$$

Proof. In the even case, we have

$$\begin{aligned} r_{2i} &= f\big(r_{2i-1}^{-\text{odd}}\big) - f\big(r_{2i-3}^{-\text{odd}}\big) \\ &\sim r_{2i-1} \sum_{j\geq 0} jp_j \big(r_{2i-3}^{-\text{odd}}\big)^{j-1} \\ &= r_{2i-1} f'\big(r_{2i-3}^{-\text{odd}}\big), \end{aligned} \tag{18}$$

which, since $r_{2i-3}^{-\text{odd}} \to 1-q$, is asymptotic to $r_{2i-1}f'(1-q)$. Similarly, we have

$$r_{2i-1} = f\big(r_{2i-2}^{+\text{even}} + 1 - q\big) - f\big(r_{2i}^{+\text{even}} + 1 - q\big) \sim r_{2i-2} \sum_{j\geq 0} jp_j (1-q)^{j-1}, \tag{19}$$

which is also asymptotic to $r_{2i-2}f'(1-q)$. ∎

If N_i is the number of nodes in the ith layer for our algorithm, then Aldous's theorem (Theorem 1.4) implies that for every fixed i,

$$\frac{N_i}{n} \to r_i \tag{20}$$

in probability. The number of nodes in the layers decreases at the indicated rate, namely $f'(1-q)$. As $q \in (1/2, 1)$, we have

$$p_1 = f'(0) < f'(1-q) < f'\left(\frac{1}{2}\right) \leq \mathbf{E}\left\{\frac{1}{2^{\xi}}\right\}. \tag{21}$$

The next section will need the event that the maximum peel number in an unconditional tree occurs at the root. We have the following lemma.

Lemma 4.6. *Let T be an unconditional Galton–Watson tree with offspring distribution ξ. Let R be the peel number of the root of such a tree and let M be the maximum peel number of any node in the tree. Let q be the solution to $q = f(1-q)$, where f is the reproduction generating function of this distribution. Then*

$$\tau_i := \mathbf{P}\{R = M = i\} = f'(1-q)^{i+o(i)} \tag{22}$$

as $i \to \infty$.

Proof. The fact that $\tau_i \le \mathbf{P}\{R = i\} = r_i = f'(1-q)^{i+o(i)}$ means that we only have to worry about finding a lower bound. To that end, consider the ξ children of the root (each the root of unconditional Galton–Watson trees), with peel numbers $R_1, \dots, R_\xi$ and maximum peel numbers $M_1, \dots, M_\xi$. We consider the odd and even cases separately.

When i is odd, the event that $R = M = i$ is implied by the event that there exists some $1 \le j \le \xi$ with $R_j = M_j = i-1$ and for all $k \ne j$, we have R_j odd and $M_j \le i$. Therefore, when i is odd, we have, by the inclusion-exclusion inequality,

$$\tau_i \ge \mathbf{E}\bigl\{\xi \cdot \tau_{i-1}\, \mathbf{P}\{R \text{ odd}, M \le i\}^{\xi-1}\bigr\} - \mathbf{E}\left\{\binom{\xi}{2} \cdot {\tau_{i-1}}^2\, \mathbf{P}\{R \text{ odd}, M \le i\}^{\xi-2}\right\}. \tag{23}$$

Note that $\mathbf{P}\{R \text{ odd}, M > i\} = o(1)$ as $i \to \infty$, and so

$$\begin{aligned} \tau_i &\ge \mathbf{E}\bigl\{\xi \cdot \tau_{i-1}\bigl(1 - q - o(1)\bigr)^{\xi-1}\bigr\} - \mathbf{E}\left\{\binom{\xi}{2} \cdot {\tau_{i-1}}^2\bigl(1-q-o(1)\bigr)^{\xi-2}\right\} \\ &= \tau_{i-1} f'\bigl(1-q-o(1)\bigr) - \frac{{\tau_{i-1}}^2}{2} f''(1-q) \\ &\ge \tau_{i-1} f'\bigl(1-q-o(1)\bigr) - \frac{{\tau_{i-1}}^2 \sigma^2}{2}. \end{aligned} \tag{24}$$

When i is even, the event that $R = M = i$ is a subset of the event that there exists some $1 \le j \le \xi$ with $R_j = M_j = i-1$ and for all $k \ne j$, we have R_j odd, $R_j \le i-2$, and $M_j \le i$. With another application of the inclusion-exclusion inequality and by a similar argument as in the odd case, we have

$$\begin{aligned} \tau_i &\ge \mathbf{E}\bigl\{\xi \cdot \tau_{i-1}\, \mathbf{P}\{R \le i-2,\, M \le i,\, R \text{ odd}\}^{\xi-1}\bigr\} \\ &\qquad - \mathbf{E}\left\{\binom{\xi}{2} {\tau_{i-1}}^2\, \mathbf{P}\{R \le i-2,\, M \le i,\, R \text{ odd}\}^{\xi-2}\right\} \\ &\ge \tau_{i-1} f'\bigl(1-q-o(1)\bigr) - \frac{{\tau_{i-1}}^2\sigma^2}{2}. \end{aligned} \tag{25}$$

In both the odd and even cases, we see that $\tau_i \ge f'(1-q)^{i+o(i)}$, completing the proof. ∎

Using this result, we can give the following property of the distribution of the maximum peel number.

Lemma 4.7. *The maximum peel number M in an unconditional Galton–Watson tree satisfies*

$$\mathbf{P}\{M \geq i\} = f'(1-q)^{i/2+o(i)}, \tag{26}$$

as $i \to \infty$, where f is the reproduction generating function.

Proof. As before, let $R_1, \ldots, R_\xi$ denote the peel numbers of children of the root and let $M_1, \ldots, M_\xi$ denote the maximum peel numbers in their respective subtrees. Let $\mu_i = \mathbf{P}\{M = i\}$, $\mu_i^- = \mathbf{P}\{M \leq i\}$, and $\mu_i^+ = \mathbf{P}\{M \geq i\}$. The event that $M \geq i$ is implied by the event

$$\max_{1 \leq j \leq \xi} M_j \geq i$$

or

$$\max_{1 \leq j \leq \xi} M_j < i \text{ and there is some } 1 \leq j \leq \xi \text{ with } R_j = M_j = i-1,$$

which we shall call B. Thus, letting E_j be the event that $R_j = M_j = i-1$, we have

$$\mathbf{P}\{M \geq i\} \geq \mathbf{P}\Big\{\max_{1 \leq j \leq \xi} M_j \geq i\Big\} + \mathbf{P}\Big\{\max_{1 \leq j \leq \xi} M_j < i, \bigcup_{j=1}^{\xi} E_j\Big\}. \tag{27}$$

Note first that

$$\mathbf{P}\Big\{\max_{1 \leq j \leq \xi} M_j \geq i\Big\} = 1 - \mathbf{E}\big\{(1-\mu_i^+)^\xi\big\} = 1 - f(1-\mu_i^+). \tag{28}$$

By taking the Taylor series expansion of f around 1, we have

$$f(1-s) = f(1) - sf'(1) + \frac{s^2}{2} f''(\theta) \tag{29}$$

for some $1 - s \leq \theta \leq 1$, so that

$$f(1-s) = 1 - s + \frac{s^2}{2} f''(\theta) \leq 1 - s + \frac{s^2}{2}\sigma^2 \tag{30}$$

and

$$\mathbf{P}\Big\{\max_{1 \leq j \leq \xi} M_j \geq i\Big\} \geq \mu_i^+ - \frac{{\mu_i^+}^2 \sigma^2}{2}. \tag{31}$$

Next, by the union bound, we have

$$\begin{aligned}
\mathbf{P}\Big\{\max_{1 \leq j \leq \xi} M_j < i, \bigcup_{j=1}^{\xi} E_j\Big\} &= \mathbf{P}\Big\{\bigcup_{j=1}^{\xi} E_j\Big\} - \mathbf{P}\Big\{\bigcup_{j=1}^{\xi} E_j, \max_{1 \leq j \leq \xi} M_j \geq i\Big\} \\
&= 1 - \mathbf{E}\big\{(1-\tau_{i-1})^\xi\big\} - \mathbf{E}\big\{\xi(\xi-1)\tau_{i-1}\mu_i^+\big\} \\
&= 1 - f(1-\tau_{i-1}) - \sigma^2 \tau_{i-1}\mu_i^+ \\
&\geq \tau_{i-1} - \frac{{\tau_{i-1}}^2\sigma^2}{2} - \sigma^2\tau_{i-1}\mu_i^+.
\end{aligned} \tag{32}$$

Collecting these bounds back into (27), we have

$$\mu_i^+ \ge \mu_i^+ + \tau_{i-1} - \frac{\mu_i^{+2}\sigma^2}{2} - \frac{\tau_{i-1}{}^2\sigma^2}{2} - \sigma^2\tau_{i-1}\mu_i^+ \tag{33}$$

and therefore

$$\frac{\sigma^2}{2}\mu_i^{+2} \ge \tau_{i-1}(1-\sigma^2\mu_i^+) - \frac{\tau_{i-1}^2\sigma^2}{2}. \tag{34}$$

Let $\phi(x)$ be a decreasing function that is $o(1)$ as $x \to \infty$. We combine (34) with Lemma 4.6 to conclude that

$$\frac{\sigma^2}{2}\mu_i^{+2} \ge \tau_{i-1}\big(1-\phi(i)\big) = f'(1-q)^{i+o(i)}. \tag{35}$$

To bound μ_i^+ from above, we observe that since the event that $M \ge i$ is a subset of the event B, we have

$$\begin{aligned} \mu_i^+ &\le 1 - f(1-\mu_i^+) + \mathbf{E}\big\{1-(1-\tau_{i-1})^\xi\big\} \\ &\le \mu_i^+ - \frac{\mu_i^{+2}}{2}\big(\sigma^2+o(1)\big) + \tau_{i-1} - \frac{\tau_{i-1}{}^2}{2}\big(\sigma^2+o(1)\big), \end{aligned} \tag{36}$$

and therefore

$$\frac{\sigma^2+o(1)}{2}\mu_i^{+2} \le \tau_{i-1} = f'(1-q)^{i+o(i)}, \tag{37}$$

which is what we need. ∎

4.5. Asymptotics of the peel number

We are now ready to prove an asymptotic result for the peel number of T_n. Our proof uses Kesten's tree T_∞.

Theorem 4.8. *Let M_n be the peel number of T_n, a conditional Galton–Watson tree on n nodes with offspring distribution ξ satisfying $\mathbf{E}\{\xi\} \le 1$ and $\mathbf{P}\{\xi = 1\} \ne 1$. Then*

$$\frac{M_n}{\log n} \to \frac{1}{\log\big(1/f'(1-q)\big)} \tag{38}$$

in probability, where f is the generating function of ξ.

Proof. For any tree t, let $h(t)$ denote its height and recall that we use the notation $m(t)$ for the maximum peel number. For the lower bound, we employ Kesten's limit tree T_∞. Let S_k denote the set of nodes of T_∞ that are children of nodes on the spine of $\tau(T_\infty, k)$ (i.e. nodes that are marked in the construction of T_∞ up to depth k). Let

$$\alpha_n = \left\lfloor \frac{\sqrt{n}}{\log^2 n} \right\rfloor \qquad \text{and} \qquad \beta_n = \left\lfloor \frac{\sqrt{n}}{\log n} \right\rfloor.$$

By the same result of [36] and [50] that we used before, we can find a coupling of $\tau(T_n, \beta_n)$ and $\tau(T_\infty, \beta_n)$ such that

$$\mathbf{P}\bigl\{\tau(T_n, \beta_n) \neq \tau(T_\infty, \beta_n)\bigr\} = o(1). \tag{39}$$

For every node u in T_∞, let T_u be the subtree of T_∞ rooted at u (this somewhat conflicts with our notation T_n, but it will be clear what we mean because u is a node and n is an integer). Let $M_n = m(T_n)$. Letting E_{ux} denote the event that $m(T_u) \le x$ and F denote the event that

$$\max_{u \in S_{\alpha_n}} h(T_u) \ge \beta_n - \alpha_n,$$

we have

$$\mathbf{P}\{M_n \le x\} \le \mathbf{P}\bigl\{\tau(T_n, \beta_n) \neq \tau(T_\infty, \beta_n)\bigr\} + \mathbf{P}\Bigl\{\bigcap_{u \in S_{\alpha_n}} E_{ux}\Bigr\} + \mathbf{P}\{F\}. \tag{40}$$

We already pointed out that $\mathbf{P}\bigl\{\tau(T_n, \beta_n) \neq \tau(T_\infty, \beta_n)\bigr\} = o(1)$; we bound the other two terms by

$$\begin{aligned} \mathbf{P}\Bigl\{\bigcap_{u \in S_{\alpha_n}} E_{ux}\Bigr\} + \mathbf{P}\{F\} \le \mathbf{P}\Bigl\{|S_{\alpha_n}| \le \frac{\sigma^2 \alpha_n}{2}\Bigr\} &+ \mathbf{P}\Bigl\{|S_{\alpha_n}| \ge \frac{\sigma^2 3\alpha_n}{2}\Bigr\} \\ &+ \mathbf{P}\bigl\{m(T) \le x\bigr\}^{\sigma^2 \alpha_n / 2} \\ &+ \frac{2\sigma^2}{2} \alpha_n \, \mathbf{P}\bigl\{h(T) \ge \beta_n - \alpha_n\bigr\}, \end{aligned} \tag{41}$$

where T is an unconditional Galton–Watson tree. Now,

$$S_{\alpha_n} / (\sigma^2 \alpha_n) \to 1$$

in probability by the weak law of large numbers, as the expected number of children of any node on the spine of T_∞ is $\sigma^2 + 1$. So, the first two terms of (41) tend to zero. Next, we see that

$$\begin{aligned} \mathbf{P}\bigl\{m(T) \le x\bigr\}^{\sigma^2 \alpha_n / 2} &= \bigl(1 - \mathbf{P}\bigl\{m(T) > x\bigr\}\bigr)^{\sigma^2 \alpha_n / 2} \\ &\le \exp\biggl(-f'(1-q)^{x/2 + o(x)} \frac{\sigma^2 \alpha_n}{2}\biggr), \end{aligned} \tag{42}$$

which tends to zero if $x = (1-\epsilon) \log n / \log\bigl(1/f'(1-q)\bigr)$. For the final term, we have, by Kolmogorov's theorem (see, e.g. [40] or [3])

$$\frac{3\sigma^2 \alpha_n}{2} \mathbf{P}\bigl\{h(T) > \beta_n - \alpha_n\bigr\} \sim \frac{3\sigma^2}{2} \cdot \frac{2\alpha_n}{\sigma^2(\beta_n - \alpha_n)} \sim \frac{3\alpha_n}{\beta_n} \sim \frac{3}{\log n}, \tag{43}$$

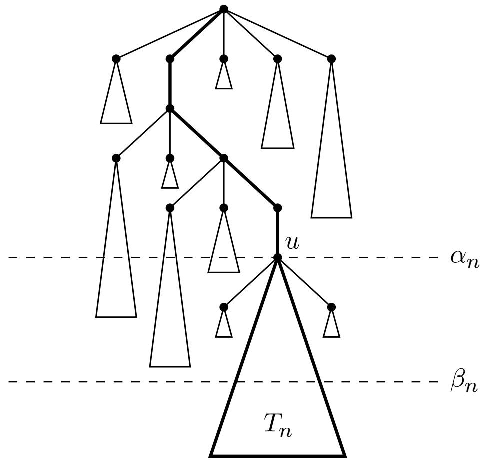


Fig. 4.4. The proof uses Kesten's infinite tree T_∞ for both bounds.

which goes to zero. We have shown that

$$\mathbf{P}\left\{ M_n < (1-\epsilon)\frac{\log n}{\log\bigl(1/f'(1-q)\bigr)} \right\} \to 0 \tag{44}$$

for all $\epsilon > 0$.

For the upper bound we will again work with T_∞, truncated to level β_n, but also require some further auxiliary definitions. Let u^* denote the unique node on the spine of T_∞ at distance α_n from the root of T_n. Let T_n^* be the subtree rooted at u^* in T_n. Let S be the set of children of nodes on the spine at distance $\le \alpha_n$ from the root. We then define

$$M_n' = \max_{u\in S} \rho(u) \qquad \text{and} \qquad M_n'' = \max_{u\in S} m(T_u).$$

Next, we let S^* denote the set of nodes u on the spine with the property that all of u's non-spine children have an odd peel number. Let Y_n be the maximal number of *consecutive* nodes on the spine that are in S^*. Lastly, we let Y_n^* denote the number of consecutive nodes going down the spine, starting at the parent of u^*, whose non-spine children all have an odd peel number. Assuming that $\tau(T_n,\beta_n) = \tau(T_\infty,\beta_n)$, we have the inequality

$$\begin{aligned} M_n &\le \max\bigl(m(T_n^*), \rho(u^*) + Y_n^*, M_n' + Y_n, M_n''\bigr) \\ &\le \max\bigl(m(T_n^*) + Y_n^*, 2M_n', 2Y_n, M_n''\bigr). \end{aligned} \tag{45}$$

To explain this, we note that nodes in S^* have a peel number that is one more than the peel numbers of their children on the spine. Nodes on the spine that are not in S^* have a peel number that is at most one more than the maximum peel number of any of their non-spine children (and this is bounded from above by M_n').

Let $\epsilon > 0$ be given and let $x = (1+\epsilon)\log n/\log\bigl(1/f'(1-q)\bigr)$. We have

$$\begin{aligned}\mathbf{P}\bigl\{m(T_n) \geq x\bigr\} \leq \mathbf{P}\bigl\{\tau(T_n, \beta_n) \neq \tau(T_\infty, \beta_n)\bigr\} + \mathbf{P}\Bigl\{\max_{u \in S} h(T_u) \geq \beta_n - \alpha_n\Bigr\} \\ + \mathbf{P}\Bigl\{Y_n^* \geq \sqrt{\log n}\Bigr\} + \mathbf{P}\{Y_n \geq x/2\} + \mathbf{P}\{M_n' \geq x/2\} \\ + \mathbf{P}\{M_n'' \geq x\} + \mathbf{P}\Bigl\{m(T_n^*) \geq x - \sqrt{\log n}\Bigr\}.\end{aligned} \tag{46}$$

As noted in our proof of the lower bound, the first two terms are $o(1)$, so we have reduced our task to showing that the latter five terms are also $o(1)$.

Let ζ be the offspring distribution of nodes on the spine (recall that $\mathbf{P}\{\zeta = i\} = ip_i$). For a node on the spine, the probability that it is in S^* is

$$\mathbf{E}\bigl\{(1-q)^{\zeta - 1}\bigr\} = \sum_{i \geq 0} ip_i(1-q)^{i-1} = f'(1-q). \tag{47}$$

Thus, Y_n^* is a geometric random variable with parameter $1 - f'(1-q)$, and hence

$$\mathbf{P}\Bigl\{Y_n^* \geq \sqrt{\log n}\Bigr\} = o(1). \tag{48}$$

Also, Y_n is bounded from above in distribution by the maximum of α_n independent Geometric$\bigl(f'(1-q)\bigr)$ random variables, so that

$$\mathbf{P}\{Y_n \geq x/2\} \leq \alpha_n f'(1-q)^{x/2} = o(1). \tag{49}$$

Next,

$$\mathbf{P}\{M_n' \geq x/2\} \leq \mathbf{E}\bigl\{|S|\bigr\}\,\mathbf{P}\{R \geq x/2\} = \sigma^2 \alpha_n f'(1-q)^{x/2 + o(x)} = o(1) \tag{50}$$

and

$$\mathbf{P}\{M_n'' \geq x\} \leq \mathbf{E}\bigl\{|S|\bigr\}\,\mathbf{P}\{M \geq x\} = \sigma^2 \alpha_n f'(1-q)^{x/2 + o(x)} = o(1). \tag{51}$$

This leaves us with the final term of (46). Observe that $|T_n^*| = n - \alpha_n - \sum_{u \in S} |T_u|$, which is at most $n - \max_{u \in S} |T_u|$. Thus,

$$\begin{aligned}\mathbf{P}\Bigl\{|T_n^*| \geq n - \frac{n}{\log^5 n}\Bigr\} &\leq \mathbf{P}\Bigl\{\max_{u \in S} |T_u| \leq \frac{n}{\log^5 n}\Bigr\} \\ &= \mathbf{E}\Bigl\{\mathbf{P}\Bigl\{|T| \leq \frac{n}{\log^5 n}\Bigr\}^{|S|}\Bigr\} \\ &\leq \mathbf{P}\Bigl\{|S| \leq \frac{\sigma^2 \alpha_n}{2}\Bigr\} + \Bigl(1 - \mathbf{P}\Bigl\{|T| > \frac{n}{\log^5 n}\Bigr\}\Bigr)^{\sigma^2 \alpha_n / 2}.\end{aligned} \tag{52}$$

Since, as noted earlier, $|S|/(\sigma^2\alpha_n) \to 1$ in probability and since $\mathbf{P}\{|T| \geq n\} = \Theta(1/\sqrt{n})$, we have

$$\begin{aligned}\mathbf{P}\Big\{|T_n^*| \geq n - \frac{n}{\log^5 n}\Big\} &\leq o(1) + \exp\bigg(-\Theta\Big(\frac{\log^{5/2} n}{\sqrt{n}}\Big)\frac{\sigma^2}{2}\alpha_n\bigg) \\ &\leq o(1) + \exp\Big(-\Theta\Big(\sqrt{\log n}\Big)\Big),\end{aligned} \tag{53}$$

which is $o(1)$. So

$$\begin{aligned}&\mathbf{P}\Big\{m(T_n^*) \geq x - \sqrt{\log n}\Big\} \\ &\qquad \leq \max_{1\leq k\leq n-n/\log^5 n} \mathbf{P}\Big\{m(T_n^*) \geq x - \sqrt{\log n} \;\Big|\; |T_n^*| = k\Big\} \\ &\qquad\qquad + \mathbf{P}\Big\{|T_n^*| \geq n - \frac{n}{\log^5 n}\Big\}.\end{aligned} \tag{54}$$

Noting that given $|T_n^*| = k$, T_n^* is again a conditional Galton–Watson tree, and letting F_k be the event that there exists a node $v \in T_n$ with $|T_v| \leq k$ and $m(T_v) \geq x - \sqrt{\log n}$, we see that

$$\mathbf{P}\Big\{m(T_n^*) \geq x - \sqrt{\log n} \;\Big|\; |T_n^*| = k\Big\} \leq \mathbf{P}\Big\{F_{n-n/\log^5 n}\Big\}. \tag{55}$$

Now for $v \in T_n$ define $t(v)$ to be the subtree of v in the shifted preorder degree sequence $\xi_v, \xi_{v+1}, \ldots, \xi_n, \xi_1, \ldots \xi_{v-1}$. Let $\rho(v)$ be the peel number of the root v of $t(v)$ and let G_{vx} denote the event that $\rho(v) \geq x - \sqrt{\log n}$ and $|t(v)| \leq n/\log^5 n$. We have

$$\mathbf{P}\{M_n \geq x\} \leq \mathbf{P}\Big\{\bigcup_{v\in T_n} G_{vx}\Big\} + o(1). \tag{56}$$

Note that $\max_{v\in T_n;|t(v)|\leq n/\log^5 n} \rho(v)$, is invariant under the cyclic shift of the preorder degree sequence. This rotational invariance, by the cycle lemma, shows that

$$\mathbf{P}\Big\{\bigcup_{v\in T_n} G_{vx}\Big\} = \frac{\mathbf{P}\{\bigcup_{v\in T_n} G_{vx},\, \sum_{1\leq i\leq n}(\xi_i - 1) = -1\}}{\mathbf{P}\{\sum_{1\leq i\leq n}(\xi_i - 1) = -1\}}, \tag{57}$$

where on the right-hand side, all probabilities are with respect to an i.i.d. sequence $\xi_1, \ldots, \xi_n$. We bound (57) from above by

$$\mathbf{P}\Big\{\bigcup_{v\in T_n} G_{vx}\Big\} \leq n \cdot \frac{\mathbf{P}\Big\{\rho(1) \geq x - \sqrt{\log n},\, |t(1)| \leq n/\log^5 n,\, \sum_{i=1}^n \xi_i = n-1\Big\}}{\mathbf{P}\{\sum_{i=1}^n \xi_i = n-1\}}. \tag{58}$$

By conditioning on the size of $t(1)$, we obtain the further bound

$$\mathbf{P}\Big\{\bigcup_{v\in T_n} G_{vx}\Big\} \leq n \cdot \mathbf{P}\Big\{\rho(1) \geq x - \sqrt{\log n}\Big\} \cdot \frac{\sup_{n/\log^5 n\leq k\leq n} \mathbf{P}\{\sum_{i=1}^k \xi_i = k\}}{\mathbf{P}\{\sum_{i=1}^n \xi_i = n-1\}}.$$

By Kolchin's estimate, the fraction is $\Theta(1)$, therefore,

$$\mathbf{P}\Big\{\bigcup_{v\in T_n} G_{vx}\Big\} \le nf'(1-q)^{x+o(x)},$$

which goes to 0 if $x = (1+\epsilon)\log n/\log\big(1-f'(1-q)\big)$. ∎

If, instead of removing leaves and parents at each step, we only remove leaves, then it is clear that the number of rounds needed to delete all nodes is simply the height of the tree. The height of random binary trees was studied by P. Flajolet and A. Odlyzko, who showed that in this case, $H_n/\sqrt{n}$ converges in law to a theta distribution [23]. Earlier, it was shown by N. G. de Bruijn, D. E. Knuth, and S. O. Rice that the expected height of a random planted plane tree is $\sqrt{\pi n} + O(1)$. It is interesting that deleting only leaves from T_n at each step requires $\Theta\big(\sqrt{n}\big)$ rounds of deletion, but deleting leaves and their parents causes the number of rounds to decrease to $\Theta(\log n)$.

Examples. We apply Theorem 4.8 to calculate explicit asymptotics of the maximum peel number for the various families of trees mentioned in Chapter 1.

i) *Flajolet's t-ary trees:* We have $f'(1-q) = 1-q$ and thus $M_n/\log n \to 1/\log(1/(1-q))$ in probability. As t gets large, q approaches $1-1/t$, so that the limit of $M_n/\log n$ is approximately $1/\log t$ for large t. For the case of full binary trees when $t = 2$, recall that $q = 2-\sqrt{2}$ and thus $M_n/\log n \to -1/\log(\sqrt{2}-1)$ in probability.

ii) *Cayley trees:* In this case, $f'(1-q) = e^{-q}$ and hence $M_n/\log n \to 1/q$ in probability; we know from earlier that $q = W(1)$, so $1/q \approx 1.763223$.

iii) *Planted plane trees:* We calculate $f'(1-q) = 1/(q+1)^2$. Recalling that $q = 1/\varphi$ where $\varphi = (\sqrt{5}-1)/2$ is the golden ratio, we have in probability $M_n/\log n \to 1/\varphi^2 \approx 0.381966$.

iv) *Motzkin trees:* The derivative $f'(1-q) = (3-2q)/3$ and substituting $q = 3-\sqrt{6}$ we get $M_n/\log n \to 1/(\log 3 - \log(2\sqrt{6}-3)) \approx 2.186769$.

v) *Binomial trees:* In this case, we have $f'(1-q) = (1-q/d)^{d-1}$. As $d \to \infty$, it is clear to see that $M_n/\log n \to 1/W(1)$ in probability, matching the earlier calculation for Cayley trees above. For the special case $d = 2$ of Catalan trees, we have $f'(1-q) = 1-q/2$ and thus $M_n/\log n \to -1/\log(\sqrt{3}-1)$. This constant is greater than that we obtain for full binary trees above, which is consistent with intuitive reasoning about the maximal peel numbers of these trees.

4.6. Distribution of the leaf height

We now repeat the treatment given in Section 3, but this time for the distribution $(\ell_i)_{i\ge 0}$, where for each $i \ge 0$, ℓ_i denotes the probability that the root of an unconditional Galton–Watson tree has leaf height equal to i. Observe that ℓ_0

is exactly the probability p_0 that the root has no children and in general, for a node u with children Γ_u the leaf height $\lambda(u)$ is

$$\lambda(u) = \begin{cases} 0, & \text{if } u \text{ is a leaf;} \\ \min_{v\in\Gamma_u} \lambda(v) + 1, & \text{otherwise.} \end{cases} \tag{59}$$

We define ℓ_i^+ and ℓ_i^- analogously to r_i^+ and r_i^-:

$$\ell_i^+ = \sum_{j\geq i} \ell_j \qquad \text{and} \qquad \ell_i^- = \sum_{j=0}^{i} \ell_j\,; \tag{60}$$

since $(\ell_i)_{i\geq 0}$ defines a distribution, $\ell_{i+1}^+ + \ell_i^- = 1$ for every $i \geq 0$. Letting E_i be the event that all the children of the root have leaf height at least i, we have, for $i \geq 1$, $\ell_i = \mathbf{P}\{E_{i-1}\} - \mathbf{P}\{E_i\}$. We can then compute

$$\ell_1 = 1 - \mathbf{E}\big\{(1-\ell_0)^\xi\big\} = 1 - f(1-\ell_0) = 1 - f(1-p_0). \tag{61}$$

and, in general, for $i \geq 1$,

$$\begin{aligned} \ell_{i+1} &= \mathbf{P}\{E_i\} - \mathbf{P}\{E_{i+1}\} \\ &= \mathbf{E}\big\{(\ell_i^+)^\xi\big\} - \mathbf{E}\big\{(\ell_{i+1}^+)^\xi\big\} \\ &= f(\ell_i^+) - f(\ell_{i+1}^+) \\ &= f(1-\ell_{i-1}^-) - f(1-\ell_i^-). \end{aligned} \tag{62}$$

By convexity of f, we see that ℓ_i is nonincreasing, and this formula provides a fast method to compute the ℓ_i recursively. The following lemma describes the behaviour of ℓ_i as i gets large.

Lemma 4.9. *Let T be a Galton–Watson tree with offspring distribution $\xi \sim (p_i)_{i\geq 0}$. If $p_1 \neq 0$, then $\ell_i = (p_1 + o(1))^i$. Otherwise if $p_1 = 0$ and $\kappa = \min\{i > 1 : p_i \neq 0\}$, then*

$$\log \ell_i = \Theta(\kappa^i) \tag{63}$$

as $i \to \infty$.

Proof. The recursive formula above is our starting point. Expanding f as a power series, for $i \geq 1$ we have, by our choice of κ,

$$\begin{aligned}
\ell_{i+1} &= \sum_{j=0}^{\infty} p_j\big((\ell_i^+)^j - (\ell_{i+1}^+)^j\big) \\
&= 0 + p_1(\ell_i^+ - \ell_{i+1}^+) + \sum_{j \geq \kappa} p_j\big((\ell_i^+)^j - (\ell_{i+1}^+)^j\big) \\
&= p_1\ell_i + p_\kappa \ell_i\big((\ell_i^+)^{\kappa-1} + (\ell_i^+)^{\kappa-2}(\ell_{i+1}^+)^1 + \ldots + (\ell_{i+1}^+)^{\kappa-1}\big) \\
&\qquad + \sum_{j>\kappa} p_j\big((\ell_i^+)^j - (\ell_{i+1}^+)^j\big). \qquad (64) \\
&\leq p_1\ell_i + \kappa p_\kappa \ell_i (\ell_i^+)^{\kappa-1} + \sum_{j>\kappa} j p_j \ell_i (\ell_i^+)^{j-1} \\
&\leq p_1\ell_i + \ell_i(\ell_1^+)^{\kappa-1}\Big(\sum_{j \geq \kappa} j p_j\Big) \\
&= p_1\ell_i + \ell_i(1-p_0)^{\kappa-1}(1-p_1).
\end{aligned}$$

Letting $\alpha = p_1 + (1-p_1)(1-p_0)^{\kappa-1} < 1$, we have $\ell_{i+1} \leq \ell_i\alpha$. Hence $\ell_{i+1} \leq \ell_1\alpha^i$ and therefore $\ell_i \to 0$ as $i \to \infty$.

Let $\epsilon > 0$ and pick n_ϵ large enough such that $\ell_i^+ \leq \epsilon$ for all $i \geq n_\epsilon$. When $p_1 \neq 0$, we have $\ell_i p_1 \leq \ell_{i+1} \leq \ell_i(p_1 + \epsilon^{\kappa-1})$, so we immediately conclude that $\ell_i = \big(p_1 + o(1)\big)^i$. If $p_1 = 0$, then

$$\begin{aligned}
\ell_{i+1} &\leq \kappa p_\kappa \ell_i (\ell_i^+)^{\kappa-1} + \sum_{j>\kappa} j p_j \ell_i (\ell_i^+)^{\kappa-1}(\ell_i^+)^{j-\kappa} \\
&\leq \kappa p_\kappa \ell_i (\ell_i^+)^{\kappa-1}\Big(1 + \sum_{j>\kappa} j p_j \epsilon^{j-\kappa}\Big) \\
&\leq \kappa p_\kappa \ell_i (\ell_i^+)^{\kappa-1}\left(1 + \frac{\epsilon}{1-\epsilon}\right) \qquad (65) \\
&\leq \kappa p_\kappa \ell_i (\ell_i)^{\kappa-1}\Big(\sum_{j=0}^{\infty} \alpha^j\Big)^{\kappa-1}\left(\frac{1}{1-\epsilon}\right) \\
&= \frac{\kappa p_\kappa}{(1-\alpha)^{\kappa-1}(1-\epsilon)} {\ell_i}^{\kappa}.
\end{aligned}$$

From this and the fact that $\ell_i \to 0$, we see that for some positive constants c_1, $c_2 < 1$, and c_3,

$$\ell_i \leq c_1 {c_2}^{\kappa^{i-c_3}} \qquad (66)$$

for all $i \geq c_3$. We also have

$$\ell_{i+1} \geq \kappa p_\kappa \ell_i (\ell_i^+)^{\kappa-1} \geq \kappa p_\kappa {\ell_i}^k, \qquad (67)$$

so that for some positive constants c_1', $c_2' < 1$, and c_3',

$$\ell_i \geq c_1' {c_2'}^{\kappa^{i-c_3'}} \qquad (68)$$

for all $i \ge c_3'$. This proves that $\log \ell_i = \Theta(\kappa^i)$. We finish the proof by noting that ℓ_i^+ can be bounded in a similar manner. ∎

4.7. Asymptotics of the leaf height

In this section, we will describe the asymptotic behaviour of the leaf height of a tree T_n (recall that this is the maximum of $\lambda(v)$, taken over all the nodes $v \in T_n$). The result depends on whether p_1 is zero or nonzero, and we have split this into two statements, since the proofs of each are rather involved.

Theorem 4.10. *Let L_n be the leaf height of T_n, a conditional Galton–Watson tree on n nodes with offspring distribution $\xi \sim (p_i)_{i\ge 0}$. If $p_1 \ne 0$, then*

$$\frac{L_n}{\log n} \to \frac{1}{\log(1/p_1)} \tag{69}$$

in probability.

Proof. Let Y_n be the length of the longest string, oriented away from the root, of nodes of degree one in T_n. Clearly $L_n \ge Y_n$, so we will first show that for $\epsilon > 0$,

$$\mathbf{P}\big\{Y_n < (1-\epsilon)\log n/\log(1/p_1)\big\} \to 0.$$

Now, $\mathbf{P}\{Y_n < x\}$ is the probability that a string of 1s appears in the preorder degree sequence of the tree $(\xi_1, \xi_2, \dots, \xi_n)$, given that the sequence is of length n and that the sequence does, in fact, define a tree; as we have used previously, this latter probability is $\Theta(n^{-3/2})$, from Dwass [21]. So letting $Y_n(\xi_1, \xi_2, \dots, \xi_n)$ be the length of the longest subsequence of 1s in the preorder degree sequence, we have

$$\mathbf{P}\{Y_n < x\} = \Theta(n^{-3/2})\,\mathbf{P}\big\{Y_n(\xi_1, \xi_2, \dots, \xi_n) < x\big\}. \tag{70}$$

We divide the sequence into n/x subsequences of length x each and let E_i be the event that the ith subsequence *does not* consist only of 1s. Then

$$\mathbf{P}\big\{Y_n(\xi_1, \xi_2, \dots, \xi_n) < x\big\} = \mathbf{P}\left\{\bigcup_{i=1}^{n/2} E_i\right\} = \mathbf{P}\{E_i\}^{n/2}. \tag{71}$$

Since $\mathbf{P}\{E_i\} = 1 - {p_1}^x$ for all i,

$$\mathbf{P}\big\{Y_n(\xi_1, \xi_2, \dots, \xi_n) < x\big\} = (1 - {p_1}^x)^{n/x} \le \exp\left(-\frac{n{p_1}^x}{x}\right). \tag{72}$$

When $x = (1-\epsilon)\log n/\log(1/p_1)$, this is equal to $\exp(-n^\epsilon/x)$, so

$$\begin{aligned}\mathbf{P}\big\{L_n < (1-\epsilon)\log n/\log(1/p_1)\big\} &\le \mathbf{P}\big\{Y_n < (1-\epsilon)\log n/\log(1/p_1)\big\} \\ &\le \Theta(n^{3/2})e^{-\Theta(n^\epsilon/\log n)},\end{aligned} \tag{73}$$

which goes to 0 as $n \to \infty$.

To tackle the upper bound, it will be helpful for us to reorder the degrees into level (also called breadth-first) ordering and to consider the following random variable. Arrange the level-order degree sequence $\xi_1, \xi_2, \ldots, \xi_n$ in a cycle, and let Z_n be the longest string of consecutive non-zero numbers in this cyclic ordering. Clearly the probability that no sub-cycle of length x of this ordering consists only of zeroes is at most $n(1-p_0)^x$. So letting A be the event that $(\xi_1, \xi_2, \ldots, \xi_n)$ defines a tree, we can crudely bound $\mathbf{P}\{Z_n \ge x\}$ by

$$\mathbf{P}\{Z_n \ge x\} = \frac{\mathbf{P}\{Z_n \ge x,\, A\}}{\mathbf{P}\{A\}} \le \Theta(n^{3/2}) n (1-p_0)^x = \Theta(n^{5/2})(1-p_0)^x, \quad (74)$$

which goes to zero if $c > (5/2)/\log\bigl(1/(1-p_0)\bigr)$ and x is set to $c \log n$. By symmetry, if Z_n is the longest string of nonzeroes in the *preorder* listing, then the same result holds, that is,

$$\mathbf{P}\biggl\{ Z_n \ge \frac{3 \log n}{\log\bigl(1/(1-p_0)\bigr)} \biggr\} \to 0. \quad (75)$$

Note that $L_n \le Z_n$. Now for $1 \le \Delta \le n$ let $L_n(\xi_1, \xi_2, \ldots, \xi_\Delta)$ be the smallest leaf height if we start constructing a tree using degrees $\xi_1, \xi_2, \ldots, \xi_\Delta$, in preorder. Two situations can occur: either $\xi_1, \xi_2, \ldots, \xi_\Delta$ defines at least one tree in a possible forest, or $\xi_1, \xi_2, \ldots, \xi_\Delta$ defines an incomplete tree. In the former case, $L_n(\xi_1, \xi_2, \ldots, \xi_\Delta)$ is the leaf height of the first completed tree; in the latter, set $L_n(\xi_1, \xi_2, \ldots, \xi_\Delta) = 0$. Note that if $Z_n \le \Delta$, then $L_n(\xi_1, \xi_2, \ldots, \xi_\Delta) \le \Delta$. For a sequence $\xi_1, \xi_2, \ldots, \xi_n$ of degrees, we define

$$L_{ni} = L_n(\xi_i, \xi_{i+1} \ldots, \xi_{i+\Delta-1}), \quad (76)$$

where addition in the indices is taken modulo n. If $Z_n \le \Delta$, note that $L_n \le \max_{1 \le i \le n} L_{ni}$. So we have

$$\begin{aligned} \mathbf{P}\{L_n > x\} &\le \mathbf{P}\{L_n > x,\, Z_n \le \Delta\} + \mathbf{P}\{Z_n > \Delta\} \\ &\le \mathbf{P}\Bigl\{ \max_{1 \le i \le n} L_{ni} > x,\, Z_n \le \Delta \Bigr\} + \mathbf{P}\{Z_n > \Delta\} \\ &\le \mathbf{P}\Bigl\{ \max_{1 \le i \le n} L_{ni} > x \Bigr\} + \mathbf{P}\{Z_n > \Delta\}. \end{aligned} \quad (77)$$

The second term is $o(1)$ if we pick $c > (5/2)/\log\bigl(1/(1-p_0)\bigr)$ as before and set $\Delta = c \log n$. In the first term, the maximum is invariant under rotations of

$(\xi_1, \xi_2, \ldots, \xi_n)$, so we use may employ the cycle lemma to obtain

$$\begin{aligned}\mathbf{P}\Big\{\max_{1\le i\le n} L_{ni} > x\Big\} &= \frac{\mathbf{P}\big\{\max_{1\le i\le n} L_{ni} > x,\ \sum_{i=1}^n \xi_i = n-1\big\}}{\mathbf{P}\big\{\sum_{i=1}^n \xi_i = n-1\big\}}\\ &\le \frac{\mathbf{P}\big\{\max_{1\le i\le n} L_{ni} > x,\ \sum_{i=\Delta+1}^n \xi_i = n-1-\sum_{i=1}^{\Delta} \xi_i\big\}}{\Theta(n^{-1/2})}\\ &\le O(n^{3/2}) \sup_\ell \mathbf{P}\Big\{L_{n1} > x,\ \sum_{i=\Delta+1}^n \xi_i = \ell\Big\}\\ &= O(n^{3/2}) \cdot \mathbf{P}\{L_{n1} > x\} \cdot \sup_\ell \mathbf{P}\Big\{\sum_{i=\Delta+1}^n \xi_i = \ell\Big\}.\end{aligned} \tag{78}$$

Rogozin's inequality tells us that

$$\sup_\ell \mathbf{P}\Big\{\sum_{i=\Delta+1}^n \xi_i = \ell\Big\} \le \frac{\gamma}{\sqrt{1-\Pi}} \cdot \frac{1}{\sqrt{n-\Delta}}, \tag{79}$$

where γ is a universal constant and $\Pi = \sup_j p_j$. So if $L(T)$ is the leaf height of the root of an unconditional Galton–Watson tree, then

$$\begin{aligned}\mathbf{P}\{L_n > x\} &\le O(n)\,\mathbf{P}\{L_{n1} > x\}\\ &\le O(n)\,\mathbf{P}\{L(T) > x\}\\ &\le O(n)\ell_x^+\\ &\le O(n)\big(p_1 + o(1)\big)^x,\end{aligned} \tag{80}$$

which goes to 0 when $x = (1+\epsilon)\log n/\log(1/p_1)$. ∎

The next theorem handles the other case, in which p_1 is zero.

Theorem 4.11. *Let L_n be the leaf height of T_n, a conditional Galton–Watson tree on n nodes with offspring distribution $\xi \sim (p_i)_{i\ge 0}$. Let $\kappa = \min\{i > 1 : p_i \ne 0\}$. If $p_1 = 0$, then*

$$\frac{L_n}{\log\log n} \to \frac{1}{\log\kappa} \tag{81}$$

in probability.

Proof. Let $\epsilon > 0$, let A be the event that $(\xi_1, \xi_2, \ldots, \xi_n)$ forms a tree and let $L_n(\xi_1, \xi_2, \ldots, \xi_n)$ be as in the previous lemma. If $L(T)$ is the leaf height of the root of an unconditional Galton–Watson tree T, we have, by the cycle lemma,

$$\begin{aligned}\mathbf{P}\{L_n \ge x\} &= \frac{\mathbf{P}\big\{L_n(\xi_1, \xi_2, \ldots, \xi_n) \ge x,\ A\big\}}{\mathbf{P}\{A\}}\\ &\le \Theta(n^{3/2})\,\mathbf{P}\big\{L_n(\xi_1, \xi_2, \ldots, \xi_n) \ge x\big\}\\ &\le \Theta(n^{3/2})n\,\mathbf{P}\{L(T) \ge x\}\\ &\le \Theta(n^{5/2})c_1 c_2^{\kappa^{x-c_3}},\end{aligned} \tag{82}$$

for some positive constants c_1, $c_2 < 1$, and c_3. When $x = (1+\epsilon)\log_\kappa \log n$, this is $\Theta(n^{5/2})c_2^{\Theta((\log n)^{1+\epsilon})}$, which goes to 0 as $n \to \infty$.

Let T_∞ be Kesten's infinite tree. Our proof of the lower bound uses the fact that

$$\mathbf{P}\bigl\{\tau(T_\infty, n^{1/3}) \neq \tau(T_n, n^{1/3})\bigr\} \to 0 \tag{83}$$

as $n \to \infty$. Let U be the set of all unconditional Galton–Watson trees T rooted less than $n^{1/4}$ of the way down the spine. Let $h(T)$ denote the height of an unconditional Galton–Watson tree; the probability that one of these trees has height greater than $n^{1/3}/2$ is bounded above by

$$\mathbf{E}\bigl\{|U|\bigr\}\,\mathbf{P}\bigl\{h(T) > n^{1/3}/2\bigr\} \le \sigma^2 n^{1/4}\left(\frac{2+o(1)}{\sigma^2 n^{1/3}/2}\right) \sim \frac{4}{n^{1/12}}, \tag{84}$$

which goes to 0. Here we used the fact that $\mathbf{E}\{\zeta\} = \sigma^2 + 1$ and applied Kolmogorov's result (see [40] and [3]) about the height of a Galton–Watson tree. If all the heights above are at most $n^{1/3}/2$, then T_n^*, the tree obtained by taking the spine up to level $n^{1/4}$ and all hanging unconditional trees up to that point, is a subtree of $\tau(T_\infty, n^{1/3})$, since $n^{1/4} + n^{1/3}/2 < n^{1/3}$.

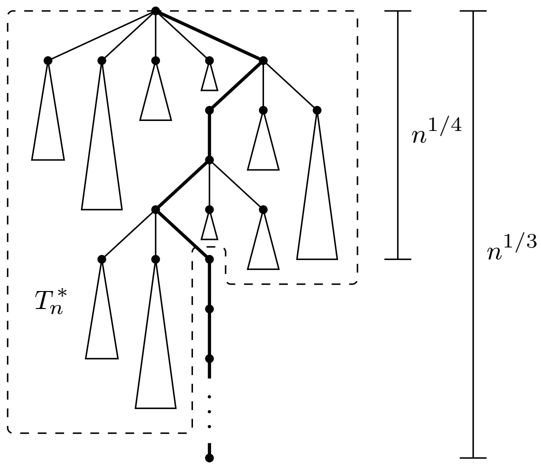


Fig. 4.5. None of the unconditional trees in T_n^* reach level $n^{1/3}$.

For every tree $t \in H$, let E_t be the event that the leaf height of the root of t is less than x and let E_T be the same event for an unconditional Galton–Watson tree (since each t in H is such a tree, there is no real moral distinction between these events). We have

$$\begin{aligned}\mathbf{P}\{L_n < x\} \le \mathbf{P}\bigl\{\tau(T_\infty, n^{1/3}) \neq \tau(T_n, n^{1/3})\bigr\} \\ + \mathbf{P}\bigl\{T_n^* \neq \tau(T_n, n^{1/3})\bigr\} + \mathbf{P}\Bigl\{\bigcap_{t\in H} E_t\Bigr\}\end{aligned}$$

$$\le o(1) + \mathbf{E}\{\mathbf{P}\{E_T\}^{|H|}\} \tag{85}$$
$$\le o(1) + \mathbf{E}\{(1 - c_1' \cdot c_2'^{\kappa^{x-c_3'}})^{|H|}\}$$
$$\le o(1) + \mathbf{E}\{(1 - c_1' \cdot c_2'^{\kappa^{x-c_3'}})^{n^{1/4}\sigma^2/2}\} + \mathbf{P}\left\{|H| < \frac{n^{1/4}\sigma^2}{2}\right\},$$

for some c_1', c_2', c_3' positive, $c_2' < 1$. Take $x = (1-\epsilon)\log_\kappa \log n$. Letting the random variables $\zeta_1, \zeta_2, \ldots, \zeta_n$ be independent and distributed as ζ, we find that

$$\mathbf{P}\{L_n < x\} \le o(1) + \exp\bigl(-\Theta(n^{1/4}e^{-\Theta((\log n)^{1-\epsilon})})\bigr) + \mathbf{P}\left\{\sum_{i=1}^{n^{1/4}}(\zeta_i - 1) < \frac{n^{1/4}\sigma^2}{2}\right\}. \tag{86}$$

Since $\mathbf{E}\{\zeta - 1\} = \sigma^2$, by the weak law of large numbers, this entire expression is $o(1)$. ∎

It is important to note that the node with maximum leaf height in a tree is not usually the root. We have the following result for the distribution of the leaf height of the root of a conditional Galton–Watson tree T_n.

Lemma 4.12. *Let L_n' denote the leaf height of the root of T_n, a conditional Galton–Watson tree of size n and offspring distribution ξ. Let f be the generating function of this distribution. Then the probability distribution of L_n' is given by*

$$\lim_{n\to\infty} \mathbf{P}\{L_n' = i\} = \prod_{j=0}^{i-1} f'(\ell_j^+). \tag{87}$$

Proof. For a node v on the spine of Kesten's infinite tree T_∞, let H^* be the leaf height of the spine-child of v, and $H(1), H(2), \ldots, H(\zeta-1)$ be the leaf heights of the $\zeta - 1$ independent unconditional Galton–Watson trees spawned by v. Then the leaf height of a node v on the spine is

$$1 + \min_{w\in\Gamma_v} \lambda(w) = 1 + \min\bigl(H(1), H(2), \ldots, H(\zeta-1)\bigr). \tag{88}$$

This defines a Markov chain on the positive integers that proceeds up the spine. The state H^* (which is just a positive integer indicating the leaf height of the node on the spine), is taken to the state

$$1 + \min\bigl(H^*, H(1), H(2), \ldots, H(\zeta-1)\bigr) \tag{89}$$

in one step of the Markov chain; here all H_i have distribution $(\ell_i)_{i\ge 0}$. Let H^{**} be the limit stationary random variable of this Markov chain. That the limit exists and that the chain is positive recurrent follows from the fact that at each step, there is a positive probability that the next state is 1. This happens when $\zeta > 1$ and one of $H(i)$ is 0. In fact, H^{**} is the unique solution of the distributional identity

$$H^{**} \overset{\mathcal{L}}{=} 1 + \min\bigl(H^{**}, H(1), H(2), \ldots, H(\zeta-1)\bigr). \tag{90}$$

N. Broutin, L. Devroye, and N. Fraiman showed that under a coalescence condition (satisfied here), the limit of the root value of T_n tends in distribution to the stationary random variable for Kesten's spinal Markov chain [12]. Thus $L'_n \to H^{**}$ in distribution.

We can describe the distribution of H^{**} more explicitly. For convenience, let $\ell_i^{**} = \mathbf{P}\{H^{**} = i\}$. For $i \geq 1$,

$$\ell_i^{**} = \mathbf{P}\{H_j \geq i+1 \text{ for all } 1 \leq j \leq \zeta - 1\}\,\mathbf{P}\{H^{**} \geq i-1\}. \tag{91}$$

This means that

$$\ell_i^{**} = \ell_{i-1}^{**}\,\mathbf{E}\big\{(\ell_{i-1}^{+})^{\zeta-1}\big\} = \ell_{i-1}^{**}\sum_{j\geq 1} jp_j(\ell_{i-1}^{+})^{j-1}, \tag{92}$$

and we can rewrite this in terms of the generating function $f(s)$ of ξ as

$$\ell_i^{**} = \ell_{i-1}^{**} f'(\ell_{i-1}^{+}) = \ell_0^{**}\prod_{j=0}^{i-1} f'(\ell_j^{+}) = \prod_{j=0}^{i-1} f'(\ell_j^{+}), \tag{93}$$

proving the lemma. ∎

Lastly, we can obtain a random variable by taking the leaf height of a node chosen uniformly at random from T_n. Its distribution is asymptotically the same as the leaf height of the root of an unconditional Galton–Watson tree.

Lemma 4.13. *Let L''_n be a random variable obtained by taking the leaf height of a node chosen uniformly at random from a conditional Galton–Watson tree T_n. We have*

$$\lim_{n\to\infty} \mathbf{P}\{L''_n = i\} = \ell_i \tag{94}$$

for all $i \geq 0$.

Proof. By Aldous's theorem, if T_n^* denotes the subtree of T_n rooted at a uniformly selected random node, then for all trees t,

$$\lim_{n\to\infty} \mathbf{P}\{T_n^* = t\} = \mathbf{P}\{T = t\}, \tag{95}$$

where T is the unconditional Galton–Watson tree. The result is immediate. ∎

Examples. We now apply Theorems 4.10 and 4.11 to compute the maximum leaf height (asymptotically in probability) for common families of trees. These results, along with the independence numbers and maximum peel numbers we computed earlier, are collected in Table 3. In the table, W denotes the Lambert function and $\varphi = (\sqrt{5}-1)/2$ is the golden ratio.

i) *Flajolet's t-ary trees:* For $t \geq 2$, we have $p_1 = 0$ here and $\kappa = t$, so we have $L_n/\log\log n \to 1/\log t$ in probability, by Theorem 4.11.

Table 3

ASYMPTOTIC VALUES OF PARAMETERS FOR CERTAIN FAMILIES OF TREES

Family	I_n	M_n	L_n
Full binary (Uniform$\{0,2\}$)	$(2-\sqrt{2})n$	$\dfrac{\log n}{\log(1/(\sqrt{2}-1))}$	$\log_2 \log n$
Flajolet t-ary ($p_0 = 1-1/t$; $p_t = 1/t$)	$\left(1 - \dfrac{1+o_{t\to\infty}(1)}{t}\right)n$	$\sim_{t\to\infty} \log_t n$	$\log_t \log n$
Cayley (Poisson 1))	$W(1)n$	$\log n / W(1)$	$\log n$
Planted plane (Geometric(1/2))	$\dfrac{n}{\varphi}$	$\dfrac{\log n}{\varphi^2}$	$\log_4 n$
Motzkin (Uniform$\{0,1,2\}$)	$(3-\sqrt{6})n$	$\dfrac{\log n}{\log 3 - \log(2\sqrt{6}-3)}$	$\log_3 n$
Catalan (Binomial(2, 1/2))	$(4-2\sqrt{3})n$	$\dfrac{\log n}{\log(1/(\sqrt{3}-1))}$	$\log_2 n$
Binomial (Binomial$(d, 1/d)$)	$\sim_{d\to\infty} W(1)n$	$\sim_{d\to\infty} \dfrac{\log n}{W(1)}$	$\dfrac{\log n}{(d-1)\log(d/(d-1))}$

ii) *Cayley trees:* In this case, $p_1 = 1/e$, so Theorem 4.10 gives us $L_n/\log n \to 1$ in probability.

iii) *Planted plane trees:* For these trees, $p_1 = 1/4$, so we have $L_n/\log n \to 1/\log 4$ in probability, by Theorem 4.10.

iv) *Motzkin trees:* This family has $p_1 = 1/3$ and $L_n/\log_n \to 1/\log 3$ in probability.

v) *Binomial trees:* For a parameter $d \geq 2$, we have $p_1 = (1-1/d)^{d-1}$, which means that $L_n/\log n \to 1/\big((1-d)\log(1-1/d)\big)$ in probability. As $d \to \infty$, the denominator approaches 1, which gives the same leaf height as the case of Cayley trees. In the special case when $d = 2$, we have the Catalan trees, for which $p_1 = 1/2$ and $L_n/\log n \to 1/\log 2$ in probability.

Further directions. The definition of the peel number and our characterisation of its asymptotic growth fully describes the running time of Algorithm I, mentioned in the introduction, which computes the layered independent set. It would be interesting to consider the runtime of the more general Algorithm P, described in Section 4.3. To this end, we define *higher-order peel numbers* as follows. Algorithm P generates an $(r+1)$-path vertex cover by repeatedly deleting subtrees with height exactly r (and marking their roots). If a node u is deleted in the mth iteration of the loop and is at depth i of the subtree that is deleted, then its *peel number of order* r (or r*th order peel number*) is $mk - i$. Note that the loop counter m should start at 1 and we have $0 \leq i \leq r$. By this definition, the peel number we studied in this chapter is simply the first order peel number. To determine the runtime of Algorithm P, one should in principle be able to approach the higher order peel numbers in the same way we approached the case $r = 1$ in Sections 4.3 and 4.4. However, even in this case one had to handle the even and odd cases separately, and we anticipate that the analysis of higher-order peel numbers will require careful reasoning with respect to congruence modulo $r + 1$.

REFERENCES

[1] Milton Abramowitz and Irene Ann Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Washington: U.S. Government Printing Office, 1972).

[2] David Aldous, "Asymptotic fringe distributions for general families of random trees," *Annals of Applied Probability* **1** (1991), 228–266.

[3] Gerold Alsmeyer, *Galton–Watson Processes*, Course notes at the University of Münster, 2008.

[4] Krishna Athreya and Peter Ney, *Branching Processes* (Berlin: Springer Verlag, 1972).

[5] Cyril Banderier, Markus Kuba, and Alois Panholzer, "Analysis of three graph parameters for random trees," *Random Structures and Algorithms* **35** (2009), 42–69.

[6] Irénée-Jules Bienaymé, "De la loi de multiplication et de la durée des familles," *Société Philomathique Paris Extraits* **5** (1845), 37–39.

[7] Irénée-Jules Bienaymé, "Considérations à l'appui de la découverte de Laplace," *Comptes Rendus de l'Académie des Sciences* **37** (1853), 309–324.

[8] Miklós Bóna, "k-protected vertices in binary search trees," *Advances in Applied Mathematics* **53** (2014), 1–11.

[9] Anna May Brandenberger, Luc Devroye, and Marcel Kieren Goh, "Root estimation in Galton–Watson trees," *Random Structures and Algorithms* (2021), 24 pp, to appear.

[10] Anna May Brandenberger, Luc Devroye, Marcel Kieren Goh, and Rosie Y Zhao, "Leaf multiplicity in a Bienaymé–Galton–Watson tree," *Discrete Mathematics and Theoretical Computer Science* **24** (2022), article no. 7.

[11] Boštjan Brešar, Marko Jakovac, Ján Katrenič, Gabriel Semanišin, and Andrej Taranenko, "On the vertex k-path cover," *Discrete Applied Mathematics* **161** (2013), 1943–1949.

[12] Nicolas Broutin, Luc Devroye, and Nicolás Fraiman, "Recursive functions on conditional Galton–Watson trees," *Random Structures and Algorithms* **57** (2020), 304–316.

[13] Sébastien Bubeck, Luc Devroye, and Gábor Lugosi, "Finding Adam in random growing trees," *Random Structures and Algorithms* **50** (2017), 158–172.

[14] Pafnuty Lvovich Chebyshev, "Des valeurs moyennes," *Journal de Mathématiques Pures et Appliquées* **12** (1867), 177–184.

[15] Gi-Sang Cheon and Louis Welles Shapiro, "Protected points in ordered trees," *Applied Mathematics Letters* **21** (2008), 516–520.

[16] Herman Chernoff, "A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics* **23** (1952), 493–507.

[17] Keith Copenhaver, "k-protected vertices in unlabeled rooted plane trees," *Graphs and Combinatorics* **33** (2017), 347–355.

[18] Luc Devroye, "Simulating size-constrained Galton–Watson trees," *SIAM Journal on Computing* **41** (2012), 1–11.

[19] Luc Devroye, Marcel Kieren Goh, and Rosie Y Zhao, "On the peel number and the leaf-height of a Galton–Watson tree," *Combinatorics, Probability and Computing* (2022), 23 pp., to appear.

[20] Rosena Ruo Xia Du and Helmut Prodinger, "On protected nodes in digital search trees," *Applied Mathematics Letters* **25** (2012), 1025–1028.

[21] Meyer Dwass, "The total progeny in a branching process," *Journal of Applied Probability* **6** (1969), 682–686.

[22] David Eisenbud, *Commutative Algebra with a View Toward Algebraic Geometry* (Berlin and New York: Springer-Verlag, 1995).

[23] Philippe Flajolet and Andrew Odlyzko, "The average height of binary trees and other simple trees," *Journal of Computer and System Sciences* **25** (1982), 171–213.

[24] Philippe Flajolet and Robert Sedgewick, *Analytic Combinatorics* (New York: Cambridge University Press, 2009).

[25] Michael Fuchs, Cecilia Holmgren, Dieter Mitsche, and Ralph Neininger, "A note on the independence number, domination number and related parameters of random binary search trees and random recursive trees," *Discrete Applied Mathematics* **292** (2021), 64–71.

[26] Francis Galton and Henry William Watson, "On the probability of extinction of families," *Journal of the Anthropological Institute of Great Britain and Ireland* **4** (1875), 138–144.

[27] Bernhard Gittenberger, Zbigniew Gołębiewski, Isabella Larcher, and Małgorzata Sulkowska, "Protection numbers in simply generated trees and Pólya trees," *arXiv preprint 1904.03519* (2019).

[28] John Haigh, "The recovery of the root of a tree," *Journal of Applied Probability* **7** (1970), 79–88.

[29] Theodore Edward Harris, *The Theory of Branching Processes* (Berlin: Springer-Verlag, 1963).

[30] Jing He, *The path ideal of a tree and its properties*, M.Sc. thesis (Lakehead University, Thunder Bay, Ont., 2007).

[31] Wassily Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association* **58** (1963), 13–30.

[32] Philippe Jacquet and Wojciech Szpankowski, "Entropy computations via analytic depoissonization," *IEEE Transactions on Information Theory* **45** (1999), 1072–1081.

[33] Svante Janson, "Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees," *Random Structures and Algorithms* **48** (2016), 57–101.

[34] Marshall Hall, Jr., *The Theory of Groups* (New York: Macmillan, 1959).

[35] Douglas Peter Kennedy, "The Galton–Watson process conditioned on the total progeny," *Journal of Applied Probability* **12** (1975), 800–806.

[36] Götz Kersting, "On the height profile of a conditioned Galton–Watson tree," *arXiv preprint 1101.3656* (1998).

[37] Harry Kesten, "Subdiffusive behavior of a random walk on a random cluster," *Annales de l'Institut Henri Poincaré* **22** (1986), 425–487.

[38] Valintin Fedorovich Kolchin, *Random Mappings* (New York: Optimisation Software Inc., 1986).

[39] Jesús Antonio de Loera, Sonja Petrović, Lily Silverstein, Despina Stasi, and Dane Wilburne, "Random monomial ideals," *Journal of Algebra* **519** (2019), 440–473.

[40] Russell Lyons and Yuval Peres, *Probability on Trees and Networks* (Cambridge: Cambridge University Press, 2017).

[41] Hosam Mahmoud Mahmoud and Mark Daniel Ward, "Asymptotic distribution of two-protected nodes in random binary search trees," *Applied Mathematics Letters* **25** (2012), 2218–2222.

[42] Hosam Mahmoud Mahmoud and Mark Daniel Ward, "Asymptotic properties of protected nodes in random recursive trees," *Journal of Applied Probability* **52** (2015), 290–297.

[43] Toufik Mansour, "Protected points in k-ary trees," *Applied Mathematics Letters* **24** (2011), 478–480.

[44] Amram Meir and John Wesley Moon, "On the altitude of nodes in random trees," *Canadian Journal of Mathematics* **30** (1978), 997–1015.

[45] Valentin Vladimirovich Petrov, *Sums of Independent Random Variables* (New York: Springer, 1975).

[46] Alfréd Rényi, "On measures of entropy and information," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (1961), 547–561.

[47] Boris Alexeyevich Rogozin, "On an estimate of the concentration function," *Theory of Probability and its Applications* **6** (1961), 94–97.

[48] Claude Elwood Shannon, "A mathematical theory of communication," *Bell System Technical Journal* **27** (1948), 379–423.

[49] Albert Nikolaevich Shiryaev, *Probability* (Berlin: Springer-Verlag, 1984).

[50] Benedikt Stufler, "Local limits of large Galton–Watson trees rerooted at a random vertex," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **55** (2019), 155–183.