Stochastic Models in Probability and Statistics
Vol. 1, No. 1, 2024, pp. 75–104. Research Paper

On sequential growth of trees subject to various labeling constraints: from enumeration to probability theory

Thierry E. Huillet*

Laboratoire de Physique Théorique et Modélisation (CNRS, UMR 8089)
CY Cergy Paris Université, 95302 Cergy-Pontoise, France
Email(s): thierry.huillet@cyu.fr

Abstract. Trees, as loop-free graphs, are fundamental hierarchical structures of Nature. Depending on the way their constitutive atoms are labeled, their growth obeys different sequential dynamics when a new atom is being appended to a current tree, possibly forming a new tree. Randomized versions of the underlying counting problems are shown to lead, in general, to Markovian triangular sequences.

Keywords: Combinatorial probability; Distinguishability; Forests; Markovian triangular sequences; Randomization; Simply generated and increasing trees.

1 Introduction

The purpose of this note is to present explicit and asymptotic methods to count various kinds of topological trees (for which there is no data giving the distance between any two of their nodes). In all cases, the use of generating functions (g.f.'s) is an essential ingredient. Explicit formulas are derived with the help of Lagrange's inversion formula. On the other hand, singularity analysis of g.f.'s leads to asymptotic formulas aiming at describing large trees.

The analysis concerns counting of the following labeled tree structures (with no restriction on the degree of their nodes):

- Simply generated labeled trees, either ordered (plane) or not (Cayley trees).
- Recursive increasing labeled trees, either ordered (plane) or not.
- The height, depth, and number of leaves of such trees.
- A weighted version of such trees.

*Corresponding author

Received: 24 December 2023 / Accepted: 22 April 2024
DOI: 10.22067/smps.2024.45214

- Forests of such trees are also considered.

The various labeling conditions translate various distinguishability possibilities of the nodes (somewhere called here, alternatively vertices or atoms) of the tree (or forest). We will herewith be also concerned by the stochastic randomization of the underlying strictly combinatorial aspects of the counting issues. Relation of enumeration problems of trees and forests to random walks can be found in Pitman (1998).

Focus will be on sequential growth of trees and forests as given, after randomization, by Markov triangular probability sequences. By that, we mean the update of their shapes when the number of its atoms goes from n to $n+1$, in an aggregation process of individuals. Whenever the additional atom connects to one of the internal nodes of the size-n tree, the number of external nodes (leaves) grows by one unit, whereas if the additional atom connects to one of its leaves, the number of its leaves remains constant. Leaves of a tree consist of its boundary, connecting it to the external world. When dealing with forests, the additional atom can either connect to one of the trees of the current forest (aggregation) or generate a new tree (nucleation). Trees are efficient graph structures to fully connect people as the number of edges required to do so is minimum but these are fragile in that the deletion of an edge disconnects the subtree below it from the main tree.

If, by recursion from the root, $\Phi(z)$ solves the functional equation

$$\Phi(z) = ze^{\sum_{k\geq 1}\frac{1}{k}\Phi\left(z^k\right)}, \tag{1}$$

then

$$c_n = [z^n]\,\Phi(z)\,,$$

(the z^n-coefficient in the power series expansion of $\Phi(z)$) is the number of rooted unlabeled size-n trees (see (Wolfram.com) and the references therein). The number of unrooted such trees is given by $c_n^* = [z^n]\,\Phi^*(z) = c_n/n$, where

$$\Phi^*(z) = \Phi(z) - \frac{1}{2}\left[\Phi^2(z) - \Phi\left(z^2\right)\right].$$

The enumeration of such trees with n up to 6 is given in (Wolfram.com). There are no known expressions for c_n (or c_n^*) but, from local singularity analysis, $c_n \sim \beta n^{-3/2} z_c^{-n}$ (respectively, $c_n^* \sim \beta n^{-5/2} z_c^{-n}$) for large n, where $\beta = 0.5349\ldots$ and $1/z_c = 2.955765\ldots$ is defined from $\Phi(z_c) = 1$. Note that the functional equation (1) is also,

$$\Phi(z) = z\prod_{n\geq 1}(1-z^n)^{-c_n}\,,$$

with $\prod_{n\geq 1}(1-z^n)^{-c_n} = e^{\sum_{k\geq 1}\frac{1}{k}\Phi\left(z^k\right)}$ the unordered "exponential" of $\Phi(z)$. Moreover,

$$\Phi(z,u) = ze^{\sum_{k\geq 1}k^{-1}\left[(u-1)z^k + \frac{1}{k}\Phi\left(z^k,u\right)\right]} = z(1-z)^{u-1}\,e^{\sum_{k\geq 1}\frac{1}{k}\Phi\left(z^k,u\right)}$$

is the joint g.f. for the nodes (vertices) and the dangling nodes (culs de sac or leaves) of such trees, with

$$c_{n,k} := \left[z^n u^k\right]\Phi(z,u)$$

the number of unlabeled rooted trees with size-n having k dangling nodes, $k = 2,\dots,n-1$. From Wolfram, $c^*_{6,1} = 1$, $c^*_{6,2} = 1$, $c^*_{6,3} = 2$, $c^*_{6,4} = 2$, and $c^*_{6,5} = 1$. For each such tree, the question of how many ways they can be labeled to yield simple or recursive unrooted trees arises.

Let B_l be a 0-diagonal $n \times n$ matrix with entries in $\{0,1\}$ whose the first row is $(0;1,\dots,1;0,\dots,0)$ (with $n_1 \geq 1$ ones), the second row is $(0;0,\dots,0;1,\dots,1;0,\dots,0)$ (with $n_2 \geq 1$ following 1 's not overlapping the n_1 first ones), ..., whose the lth row is $(0;0,\dots,0;0,\dots,0;\dots;1,\dots,1)$ (with $n_l \geq 1$ final 1's not overlapping the $n_1 + \dots + n_{l-1}$ previous ones). The remaining rows are zero-sum rows, and $n_1 + \dots + n_l = n-1$ is the overall number of edges in the size-n unrooted tree; the nonoverlapping condition in the formation of B_l guarantees that there are no cycles in the tree graph.

If $n_1 \geq n_2 \geq \dots \geq n_l \geq 1$, then there are $p_{n-1,l}$ such matrices B_l, where $p_{n,l}$ is the number of unordered partitions of n into l positive summands, with $p_{n,l}$ obeying the three-term recursion

$$p_{n,l} = p_{n-1,l-1} + p_{n-l,l},$$

with boundary conditions $p_{n,l} = 0$ if $l \leq 0$ or $l > n$. Upon transposition of B_l, the symmetrized matrix

$$A_l = B_l + B'_l$$

is thus, up to permutation of its rows and columns, the incidence matrix of unrooted unlabeled trees with $k = n - l$ dangling nodes. Both the row and column sums vectors of A_l are the same. We conclude that $c^*_{n,k}$, $k = 2,\dots,n-1$, is the number of matrices obtained as PA_lP', where P runs over all permutation $n \times n$ matrices (the equivalence class of A_l).

2 Number of atoms and leaves in a size-n simple tree

We shall distinguish two main types of simply generated (or simple) trees, namely, as follows:

- Ordered (or plane) trees: The reason is that one can draw the tree in the half-plane so that the children of every parent are ordered from left to right, say from the youngest child to the oldest one. Embeddings obtained from cyclic rotations of the sub-trees around the root are not allowed.

Such trees are amenable to the Ulam–Harris–Neveu ordering of their nodes (horizontal ordering holds), and they can be represented as strata with the founder on top and the successive layers below; see Neveu (1986). Given that an individual of the population at generation k has been labeled by vertex $\mathbf{v} = v_1 \dots v_k$ (as a concatenation of k positive integers) and gives birth to $K_{\mathbf{v}} \geq 1$ children, its offspring is labeled by $\mathbf{v}1,\dots,\mathbf{v}K_{\mathbf{v}}$. Each individual at generation k thus gets a concatenated label $\mathbf{v} = v_1 \dots v_k$ for which label $v_1 \dots v_{k-1}$ is one of its mothers, $v_1 \dots v_{k-2}$, is one of its grandmothers, ..., up to $\emptyset$, is the conventional label of the root. Such ordered trees are the combinatorial versions of Galton–Watson trees in the theory of branching processes.

- Unordered (nonplane) trees: If the above condition is not imposed, then such trees are called Cayley trees.

2.1 Nonincreasing (simply generated) trees

By recursion from the root, the number of labeled simply generated (or simple) trees of size n generated by the local g.f. $\phi(z)$ (with nonnegative $\phi_m := [z^m]\,\phi(z)$, $m \geq 0$, $\phi(0) = 1$) is obtained as

$$C_n = n!\,[z^n]\,\Phi(z), \tag{2}$$

where $\Phi(z)$ solves the functional equation $\Phi(z) = z\phi(\Phi(z))$, $\Phi(0) = 0$. Lagrange inversion formula states that for all $n \geq 1$

$$[z^n]\,\Phi(z) = \frac{1}{n}\left[z^{n-1}\right]\phi(z)^n. \tag{3}$$

In enumeration problems from combinatorics, $k!\phi_k$ are assumed to be nonnegative integers.

A more general form of Lagrange inversion formula states that (with $'$ denoting derivative)

$$[z^n]\,h(\Phi(z)) = \frac{1}{n}\left[z^{n-1}\right]\left(h'(z)\,\phi(z)^n\right), \tag{4}$$

for any arbitrary analytic output function h; see Surya and Warnke (2023).

2.1.1 Cayley (unordered) trees Rényi (1959)

For rooted and labeled Cayley trees, $\phi(z) = e^z$. By Lagrange inversion formula,

$$\Phi(z) = \sum_{n\geq 1} \frac{n^{n-1}}{n!} z^n$$

solves $\Phi(z) = ze^{\Phi(z)}$ (with $\phi(z) = e^z$ and $C_n = n^{n-1}$).

Note $z_c = \sup(z > 0 : \Phi(z) < \infty) = 1/e$ with $\Phi(z_c) = 1$, $\Phi'(z_c) = \infty$. The probability to observe a particular size-n tree τ_n among all such size-n trees is $1/c_n$. The tilted probability to observe a tree of size n among all possible trees is

$$\frac{z^n C_n/n!}{\Phi(z)},$$

for those $z \in (0, z_c]$, where $z_c = 1/e$. Tilting is necessary because the overall number of Cayley trees, regardless of their sizes, is infinite.

The joint g.f. of their nodes and leaves reads:

$$\Phi(z,u) = z(u - 1 + e^{\Phi(z,u)}),$$

hence with

$$\begin{aligned} [z^n]\,\Phi(z,u) &= \frac{1}{n}\left[z^{n-1}\right](u-1+e^z)^n \\ &= \frac{1}{n}\left[z^{n-1}\right]\sum_{k=0}^{n}\binom{n}{k}u^k\,(e^z-1)^{n-k}. \end{aligned}$$

Therefore,

$$\begin{aligned} C_{n,k} &:= n!\left[z^n u^k\right]\Phi(z,u) = (n-1)!\binom{n}{k}\left[z^{n-1}\right](e^z-1)^{n-k} \\ &= \frac{n!}{k!}S_{n-1,n-k}, \qquad k = 1,\ldots,n-1, \end{aligned} \tag{5}$$

due to the vertical "g.f. of Stirling numbers of the second kind" $S_{n,k}$:

$$\sum_{n\geq k} S_{n,k}\frac{z^n}{n!} = \frac{(e^z-1)^k}{k!} \Rightarrow [z^n](e^z-1)^k = \frac{k!}{n!}S_{n,k}.$$

We get

$$kC_{n+1,k} = k(n+1)C_{n,k} + (n+1)(n-k+1)C_{n,k-1}, \tag{6}$$

with boundary conditions $C_{n,0} = C_{n,n} = 0$, $C_{n,1} = n!$ and $C_{n,n-1} = n$. In addition,

$$C_n := [z^n]\Phi(z,1) = n^{n-1} = \sum_{k=1}^{n-1} C_{n,k}.$$

Assuming uniform sampling, we have $\mathbf{P}(L_n = k) = \frac{C_{n,k}}{C_n} > 0$, $k = 1,\ldots,n-1$ (otherwise 0): The law of L_n has finite support, varying with n. From (6), with $\mathbf{P}(L_n = k) = \frac{C_{n,k}}{C_n}$, we get

$$k\mathbf{P}(L_{n+1} = k) = \left(\frac{n}{n+1}\right)^{n-1}[k\mathbf{P}(L_n = k) + (n-(k-1))\mathbf{P}(L_n = k-1)]. \tag{7}$$

The latter recursion (6) may be written as

$$\mathbf{P}(L_{n+1} = k) = q_{k,k}^{(n)}\mathbf{P}(L_n = k) + q_{k-1,k}^{(n)}\mathbf{P}(L_n = k-1),$$

defining the (positive) transition coefficients $q_{k,k}^{(n)}$ and $q_{k-1,k}^{(n)}$, which are not transition probabilities. This three-term ("space-time" inhomogeneous) recurrence is therefore not one of standard Markov chains with a usual probability transition matrix. However, it is one of triangular Markov probability sequences whose support varies with n linearly. Note that the ratio $\rho_{k,n} := q_{k-1,k}^{(n)}/q_{k,k}^{(n)} = (n-(k-1))/k$ obeys

$$\begin{aligned} q_{k-1,k}^{(n)}/q_{k,k}^{(n)} &< 1 \text{ if } k > (n+1)/2, \\ q_{k-1,k}^{(n)}/q_{k,k}^{(n)} &> 1 \text{ if } k < (n+1)/2, \end{aligned}$$

translating that L_n is attracted to the central part of its support, whose two endpoints $\{1, n-1\}$ are (asymmetrically) weakly repelling. Introduce the superdiagonal transition matrix

$$Q_{n,n} := \begin{pmatrix} q_{1,1}^{(n)} & q_{1,2}^{(n)} & 0 & \cdots \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & q_{n-1,n-1}^{(n)} & q_{n-1,n}^{(n)} \\ \vdots & \vdots & 0 & q_{n,n}^{(n)} \end{pmatrix},$$

and let $Q_{n-1,n}$ be its $(n-1)\times n$ truncated version. With $\pi_2^L := 1$, taking indeed into account the boundary conditions, the distributions $\pi_n^L := (\mathbf{P}(L_n = k), k \in \{1,\ldots,n-1\})$, $n \geq 2$, satisfy the recursion

$$\pi_{n+1}^L = (\pi_n^L, 0)\, Q_{n,n} = \pi_n^L Q_{n-1,n}, \qquad n \geq 2.$$

Thus, $\pi_n^L = \overleftarrow{\prod}_{m=2,\ldots,n-1} Q_{m-1,m}$, $n \geq 3$ is an integrated form solution of the recursion, as a left product of nested rectangular matrices (starting from $Q_{n-2,n-1}$ to the right, ending up with $Q_{1,2}$ to the left). Because for each $n \geq 1$, π_n^L is a probability vector, we get that π_n^L is orthogonal to the 1-shifted column sum vector $\mathbf{q}_{n-1} - \mathbf{1}$ of $Q_{n-1,n}$ with kth entry $q_{k,k}^{(n)} + q_{k,k+1}^{(n)} - 1$, $k = 1, \ldots, n-1$. This can easily be seen while post-multiplying π_{n+1}^L by the column vector $\mathbf{1}$ and observing $\pi_{n+1}^L \mathbf{1} = \pi_n^L \mathbf{1} = 1$.

Next, the identity (Rényi 1959)

$$\sum_{k=1}^{n-1} C_{n,k} \frac{\binom{x}{n-k}}{\binom{n}{n-k}} = x^{n-1}$$

yields (with $x = n-1$): $\sum_{k=1}^{n-1} k C_{n,k} = n(n-1)^{n-1}$. Also,

$$\partial_u \Phi(z,1) = \frac{z}{1 - \Phi(z,1)},$$

with

$$\begin{aligned} [z^n] \partial_u \Phi(z,1) &= \left[z^{n-1}\right] \frac{1}{1-\Phi(z,1)} = \frac{1}{n-1} \left[z^{n-2}\right] \left(\left(\frac{1}{1-z} \right)' e^z \right) \\ &= n(n-1)^{n-1}, \end{aligned}$$

so that

$$\begin{aligned} \mathbf{E}(L_n) &= \frac{[z^n] \partial_u \Phi(z,1)}{[z^n] \Phi(z,1)} = n \left(1 - \frac{1}{n}\right)^{n-1} \sim n/e, \\ \sigma^2(L_n) &= n(n-1)\left(1 + \frac{2}{n}\right)^{n-1} + n\left(1 - \frac{1}{n}\right)^{n-1} - n^2 \left(1 - \frac{1}{n}\right)^{2(n-1)} \\ &\sim n \frac{e-2}{e^2}. \end{aligned}$$

The variance term is obtained while plugging $x = n-2$ in the identity. The Central Limit Theorem (CLT) therefore holds:

$$\frac{L_n - \mathbf{E}(L_n)}{\sigma(L_n)} \xrightarrow{d} \mathcal{N}(0,1). \tag{8}$$

Rényi (1959) and Steele (1987) rather studies unrooted Cayley trees (there are n^{n-2} such trees) with no major difference with the rooted case here. Trees, in Rényi's writing, consist of the optimal graphs to connect n cities with just one path joining any pair of cities, passing through their most recent common ancestor.

2.1.2 Ordered (plane) trees

For rooted ordered trees, $\phi(z) = 1/(1-z)$. The g.f.

$$\Phi(z) = \frac{1-\sqrt{1-4z}}{2}$$

solves $\Phi(z) = z/(1-\Phi(z))$ (with $\phi(z) = 1/(1-z)$).

Note $z_c = \sup(z > 0 : \Phi(z) < \infty) = 1/4$ with $\Phi(z_c) = 1/2$ and $\Phi'(z_c) = \infty$. $\Phi(z)$ has an algebraic singularity at $z_c = 1/2$.

Here, $C_n = n![z^n]\Phi(z) = \frac{(2n-2)!}{(n-1)!}$ and $c_n = [z^n]\Phi(z) \sim \frac{1}{\sqrt{\pi}} n^{-3/2} 4^{n-1}$ (by singularity analysis of $\Phi(z)$). Note also $C_{n+1} = 2(2n-1)C_n$.

Here, C_n counts the number of labeled size-n trees respecting the Ulam-Harris-Neveu ordering of the children, while c_n counts the number of such size-n trees, dropping the global labeling condition. There are thus $c_n = \frac{(2n-2)!}{n!(n-1)!}$ rooted ordered trees with n atoms.

Let us now come to leaves of such trees:

With $\Phi(z) = \Phi(z,1)$, the joint g.f. of nodes and leaves solves

$$\Phi(z,u) = z(u-1+1/(1-\Phi(z,u))).$$

Hence,

$$\Phi(z,u) = \frac{1+(u-1)z-\sqrt{(1+(u-1)z)^2-4zu}}{2},$$

with

$$\begin{aligned}
[z^n]\Phi(z,u) &= \frac{1}{n}\left[z^{n-1}\right](u+z/(1-z))^n \\
&= \frac{1}{n}\left[z^{n-1}\right]\sum_{k=0}^{n}\binom{n}{k}u^k(z/(1-z))^{n-k} \\
&= \frac{1}{n}\sum_{k=0}^{n}\binom{n}{k}u^k\left[z^{k-1}\right](1-z)^{-(n-k)} \\
&= \frac{1}{n}\sum_{k=0}^{n}\binom{n}{k}\frac{[n-k]_{k-1}}{(k-1)!}u^k \\
&= \frac{1}{n}\sum_{k=0}^{n}\binom{n}{k}\binom{n-2}{k-1}u^k.
\end{aligned}$$

Therefore,

$$C_{n,k} := n!\left[z^n u^k\right]\Phi(z,u) = (n-1)!\binom{n}{k}\binom{n-2}{k-1}, \qquad k = 1,\ldots,n-1. \tag{9}$$

Using (Pascal triangle)

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

and

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$$

entails (with $C_{n,1}=n!$ and $C_{n,n}=0$)

$$C_{n+1,k}=\frac{n\left(n-1\right)}{n-k}C_{n,k}+\frac{n\left(n-1\right)}{k-1}C_{n,k-1}, \tag{10}$$

from which the following three-term recurrence for $\mathbf{P}\left(L_n=k\right)=\frac{C_{n,k}}{C_n}$ can be deduced (with $\mathbf{P}\left(L_n=n\right)=\mathbf{P}\left(L_n=0\right)=0$):

$$\mathbf{P}\left(L_{n+1}=k\right)=\frac{C_{n+1,k}}{C_{n+1}}=\frac{n\left(n-1\right)}{2\left(2n-1\right)}\left[\frac{1}{n-k}\mathbf{P}\left(L_n=k\right)+\frac{1}{k-1}\mathbf{P}\left(L_n=k-1\right)\right]. \tag{11}$$

For such random walks, the next step is either a no-move step or a move-up one. What (11) says is that whenever L_n approaches its extremal values, either 1 (or $n-1$), there is a smaller chance (inversely proportional to this value) that the next connection will result in a no-move step (a move-up step); anti-preferential attachment rule holds. Here, the ratio $\rho_{k,n}:=q_{k-1,k}^{(n)}/q_{k,k}^{(n)}=\left(n-k\right)/\left(k-1\right)$ obeys

$$\begin{aligned} q_{k-1,k}^{(n)}/q_{k,k}^{(n)} &< 1 \text{ if } k>\left(n+1\right)/2,\\ q_{k-1,k}^{(n)}/q_{k,k}^{(n)} &> 1 \text{ if } k<\left(n+1\right)/2, \end{aligned}$$

translating that L_n is again attracted to the central part of its support, whose two endpoints $\{1,n-1\}$ are (asymmetrically) repelling. Here,

$$n!\left[z^n\right]\Phi\left(z,1\right)=\frac{(2n-2)!}{(n-1)!}=\sum_{k=1}^{n-1}C_{n,k}=C_n.$$

Next,

$$\partial_u\Phi\left(z,1\right)=\frac{z}{2}+\frac{z}{2\sqrt{1-4z}}=z\frac{1-\Phi\left(z\right)}{1-2\Phi\left(z\right)},$$

$$n!\left[z^n\right]\partial_u\Phi\left(z,1\right)=n!\left[z^{n-1}\right]\left(\frac{1}{2}+\frac{1}{2\sqrt{1-4z}}\right)=\frac{n}{2}C_n \text{ for } n\geq 2.$$

$$\begin{aligned} \mathbf{E}\left(L_n\right) &= \frac{\left[z^n\right]\partial_u\Phi\left(z,1\right)}{\left[z^n\right]\Phi\left(z,1\right)}=\frac{n}{2},\\ \sigma^2\left(L_n\right) &\sim n/8. \end{aligned}$$

The variance term can be computed from (see Drmota 2009, p. 83)

$$\sigma^2\left(L_n\right)=n!\left[z^n\right]\left[\partial_u^2\Phi\left(z,1\right)+\partial_u\Phi\left(z,1\right)\right]/C_n-\left[n!\left[z^n\right]\partial_u\Phi\left(z,1\right)/C_n\right]^2.$$

A CLT holds.

2.2 Increasing (or recursive) trees

A size-n rooted and increasing Cayley tree has vertices with indices or labels $\{1,\ldots,n\}$ increasing for any path from the root to its leaves. Wherever a new connection is created in this tree, the adjunction of a new node with index $n+1$ will result in a size-$(n+1)$ rooted increasing tree. Increasing trees can in addition be unordered (Cayley) or ordered. Such trees were studied by Bergeron et al. (1992).

2.2.1 Cayley (unordered) increasing trees

The g.f. of unordered, rooted increasing trees solves the ordinary differential equation $\Phi'(z) = e^{\Phi(z)}$, $\Phi(0) = 0$, hence $\Phi(z) = -\log(1-z)$ with $z_c = \sup(z > 0 : \Phi(z) < \infty) = 1$ with $\Phi(z_c) = \infty$. Hence, $\Phi(z)$ has a logarithmic singularity at $z_c = 1$. We get $C_n = (n-1)!$ and $c_n = 1/n$.

The joint generating of nodes and leaves solves:

$$\partial_z \Phi(z,u) = u - 1 + e^{\Phi(z,u)}.$$

Hence,

$$z = \int_0^{\Phi(z,u)} \frac{dz'}{u-1+e^{z'}},$$

with solution

$$\Phi(z,u) = \log\left(\frac{1-u}{1-ue^{z(1-u)}}\right).$$

With $k = 1,\ldots,n-1$, we have

$$\begin{aligned} C_{n,k} &:= n!\left[z^n u^k\right]\Phi(z,u) = (n-1)!\left[z^{n-1}u^k\right]\left(u-1+\frac{1-u}{1-ue^{z(1-u)}}\right) \\ &=: E_{n-1,k}, \qquad k = 1,\ldots,n-1, \end{aligned} \tag{12}$$

a shifted version of the first kind Eulerian triangle, for which

$$C_{n+1,k} = kC_{n,k} + (n-k+1)\,C_{n,k-1}. \tag{13}$$

Also, from

$$\Phi(z,1) = -\log(1-z),$$

$$C_n := [z^n]\,\Phi(z,1) = (n-1)! = \sum_{k=1}^{n-1} C_{n,k}.$$

With $\mathbf{P}(L_n = k) = \frac{C_{n,k}}{C_n}$, $k = 1,\ldots,n-1$, therefore,

$$\mathbf{P}(L_{n+1} = k) = \frac{C_{n+1,k}}{C_{n+1}} = \frac{k}{n}\mathbf{P}(L_n = k) + \left(1-\frac{k-1}{n}\right)\mathbf{P}(L_n = k-1), \tag{14}$$

the dynamics of a conventional inhomogeneous Markov chain with transition probabilities. Multiplying (14) by k (respectively, k^2) and summing over the admissible range of k yields recurrences for $\mathbf{E}(L_n)$ (respectively, $\mathbf{E}\left(L_n^2\right)$) from which

$$\mathbf{E}(L_n) = \frac{[z^n]\,\partial_u\Phi(z,1)}{[z^n]\,\Phi(z,1)} = n/2,$$
$$\sigma^2(L_n) = 7n/12 + 1/3 \sim 7n/12.$$

A CLT holds.

What (14) says is that, whenever L_n approaches its extremal values, either 1 (or $n-1$), there is a large probability (proportional to this value) that the next connection will propel the walker back inside its support. Else, the ratio $\rho_{k,n} := q_{k-1,k}^{(n)}/q_{k,k}^{(n)} = (n-k+1)/k$ also obeys:

$$q_{k-1,k}^{(n)}/q_{k,k}^{(n)} < 1 \text{ if } k > (n+1)/2,$$
$$q_{k-1,k}^{(n)}/q_{k,k}^{(n)} > 1 \text{ if } k < (n+1)/2,$$

translating that the dynamics of L_n is attracted by the central body of its support, whose endpoints remain repelling.

2.2.2 Increasing ordered (plane) trees

The exponential generating function (e.g.f.) $\Phi(z)$ of such ordered, rooted increasing trees solves: $\Phi'(z) = 1/(1-\Phi(z))$, $\Phi(0) = 0$. Hence, $\Phi(z) = 1-\sqrt{1-2z}$ with

$$C_n = n!\,[z^n]\,\Phi(z) = (2n-3)!! = 2^{-(n-1)}(2n-2)!/(n-1)!, \tag{15}$$

and $C_{n+1} = (2n-1)\,C_n$. There are thus 2^{n-1} more of simple ordered n-trees than increasing ordered n-trees. The factor 2^{n-1} is the number of ordered partitions of n into positive summands.

Here, $z_c = \sup(z>0 : \Phi(z) < \infty) = 1/2$ with $\Phi(z_c) = 1$, $\Phi'(z_c) = \infty$.

The joint e.g.f. of atoms and leaves solves:

$$\partial_z\Phi(z,u) = u - 1 + 1/(1-\Phi(z,u))$$

or

$$z = \int_0^{\Phi(z,u)} \frac{dz'}{u-1+1/(1-z')}.$$

With $C(z) = \sum_{n\geq 1}\frac{n^{n-1}}{n!}z^n$ the Cayley function solving $C(z)\,e^{C(z)} = z$,

$$\Phi(z,u) = \frac{u - C\left(ue^{-u}e^{z(1-u)^2}\right)}{1-u}$$

as the inverse function of

$$\int_0^z \frac{dz'}{u-1+1/(1-z')} = \frac{\log(1+(1-u)\,z/u) - (1-u)\,z}{(1-u)^2},$$

(see Bergeron et al. (1992)). The triangular array

$$C_{n,k} := n!\left[z^n u^k\right]\Phi(z,u), \qquad k = 1,\ldots,n-1,$$

constitutes a shifted version of the second kind Eulerian triangle, with

$$C_{n+1,k} = kC_{n,k} + (2n-k)\,C_{n,k-1}. \tag{16}$$

Furthermore,

$$C_n := n!\,[z^n]\,\Phi(z,1) = (2n-3)!! = \sum_{k=1}^{n-1} C_{n,k}.$$

With $\mathbf{P}(L_n = k) = \frac{C_{n,k}}{C_n}$, $k = 1,\ldots,n-1$, therefore,

$$\mathbf{P}(L_{n+1} = k) = \frac{C_{n+1,k}}{C_{n+1}} = \frac{k}{2n-1}\mathbf{P}(L_n = k) + \left(1 - \frac{k-1}{2n-1}\right)\mathbf{P}(L_n = k-1). \tag{17}$$

Multiplying (17) by k and summing over the admissible range of k yields recurrences for $\mathbf{E}(L_n)$ from which

$$\mathbf{E}(L_n) := \frac{[z^n]\,\partial_u \Phi(z,1)}{C_n} = \frac{2n-1}{3}.$$

A CLT holds.

Remark: When conditioning, as above, on the number n of nodes of a tree with g.f. $\Phi(z)$, the law of L_n has bounded support $1,\ldots,n-1$. One may wish to consider the distribution of say N_k, the number of nodes of the tree given its number of leaves is k; but then, the law of N_k has unbounded support $k, k+1, \ldots$. To this end, we first observe that the probability law of N, the number of nodes in a tree with g.f. $\Phi(z)$, is given by the tilting

$$\mathbf{P}(N = n) = \frac{z^n C_n / n!}{\Phi(z)},$$

for any $z < z_c$ ($z \leq z_c$), depending on $\Phi(z_c) = \infty$ ($\Phi(z_c) < \infty$). Tilting is necessary in the randomization process because of the divergence of the series c_n. Then, with $n \geq k$,

$$\begin{aligned}\mathbf{P}(N_k = n) &= \mathbf{P}(L_n = k)\,\frac{z^n C_n / n!}{\Phi(z)} / \sum_{n\geq k} \text{Num} \\ &= \frac{z^n C_{n,k} / n!}{[u^k]\Phi(z,u)},\end{aligned}$$

where $[u^k]\Phi(z,u) = \sum_{n\geq k} z^n C_{n,k}/n!$ is the horizontal g.f. of $\Phi(z,u)$. The law of N_k necessarily depends on z.

3 Root degree

With $\Phi(z,1) = \Phi(z)$ solving $\Phi(z) = z\phi(\Phi(z))$ or $\Phi'(z) = \phi(\Phi(z))$, define

$$R(z,u) = z\phi(u\Phi(z)) \text{ or } R(z,u) = \int_0^z \phi\left(u\Phi\left(z'\right)\right) dz', \tag{18}$$

where u marks the root-degree of the tree in that

$$R_{n,k} := n!\left[u^k z^n\right] R(z,u)$$

is the number of trees with n nodes and root-degree k.

3.1 Nonincreasing trees

3.1.1 Cayley (unordered) trees

- If $R(z,u) = ze^{u\Phi(z)}$, $\partial_u R(z,1) = z\Phi(z) e^{\Phi(z)} = \Phi(z)^2$. By Lagrange inversion theorem, $\mathbf{E}(R_n) = 2\left(1-\frac{1}{n}\right) \sim 2$. More generally, we have

$$\begin{aligned} [z^n] R(z,u) &= \left[z^{n-1}\right] 1/(1-u\Phi(z)) \\ &= \frac{1}{n-1}\left[z^{n-2}\right] \phi(uz)' \phi(z)^{n-1} \\ &= \frac{u}{n-1}\left[z^{n-2}\right] \phi'(uz) \phi(z)^{n-1} \\ &= \frac{u}{n-1}\left[z^{n-2}\right] e^{z(u+n-1)} \\ &= \frac{u(u+n-1)^{n-2}}{(n-1)!}. \end{aligned}$$

Hence,

$$R_{n,k} := n!\left[z^n u^k\right] R(z,u) = n\binom{n-2}{k-1}(n-1)^{n-1-k} \text{ and}$$

$$\mathbf{E}\left(u^{R_n}\right) = \frac{n!\,[z^n] R(z,u)}{C_n} = \frac{nu(u+n-1)^{n-2}}{n^{n-1}} = u\left(1-\frac{1}{n}+\frac{u}{n}\right)^{n-2}, \tag{19}$$

a shifted binomial distribution for which $\mathbf{E}\left(u^{R_n}\right) \underset{n\to\infty}{\to} ue^{-(1-u)}$, the probability generating function (p.g.f.) of a shifted Poisson (1) random variable (r.v.).

3.1.2 Ordered (plane) trees

- If, as for this case, $R(z,u) = z/(1-u\Phi(z))$,

$$\begin{aligned} [z^n] R(z,u) &= \left[z^{n-1}\right] 1/(1-u\Phi(z)) \\ &= \frac{1}{n-1}\left[z^{n-2}\right] \phi(uz)' \phi(z)^{n-1} \\ &= \frac{u}{n-1}\left[z^{n-2}\right] \phi'(uz) \phi(z)^{n-1} \\ &= \frac{u}{n-1}\left[z^{n-2}\right] \frac{1}{(1-uz)^2 (1-z)^{n-1}} \\ &= \frac{1}{n-1}\sum_{k=1}^{n-1} k\binom{2n-k-1}{n-2} u^k, \end{aligned}$$

$$r_{n,k} := \left[z^n u^k\right] R(z,u) = \frac{k}{n-1}\binom{2n-k-1}{n-2},$$

$$
\begin{aligned}
\mathbf{E}\left(u^{R_n}\right) &= \frac{n!\,[z^n]\,R(z,u)}{C_n} = \frac{u}{\binom{2n-2}{n-1}}\left[z^{n-2}\right]\frac{1}{(1-uz)^2(1-z)^{n-1}} \\
&= \frac{1}{\binom{2n-2}{n-1}}\sum_{k=1}^{n-1} k\binom{2n-k-1}{n-2}u^k.
\end{aligned}
$$

We have

$$\partial_u R(z,1) = \frac{\Phi(z)^3}{z} \text{ and}$$

$$
\begin{aligned}
[z^n]\,\partial_u R(z,1) &= \left[z^{n+1}\right]\Phi(z)^3 = \frac{3}{n+1}[z^n]\frac{z^2}{(1-z)^n} \\
&= \frac{3}{n+1}\left[z^{n-2}\right](1-z)^{-n} = \frac{3}{n+1}\binom{2n-3}{n-2},
\end{aligned}
$$

$$\mathbf{E}(R_n) = \frac{[z^n]\,\partial_u R(z,1)}{C_n} = \frac{3n!}{n+1}\binom{2n-3}{n-2}\frac{(n-1)!}{(2n-2)!} = \frac{3n}{2(n+1)} \sim \frac{3}{2}.$$

For simple trees, there is a small (finite) number of root sub-trees (at least one of which must be very large!).

3.2 Increasing (recursive) trees

3.2.1 Cayley (unordered) trees

- For nonplane increasing trees (see Bergeron et al. 1992, p. 40 and Mahmoud et al. 1993), with $\Phi(z) = -\log(1-z)$, we have

$$
\begin{aligned}
\partial_z R(z,u) &= e^{u\Phi(z)} = (1-z)^{-u}, \\
R(z,u) &= \frac{1-(1-z)^{1-u}}{1-u}, \\
n!\,[z^n]\,R(z,u) &= [u]_{n-1}, \\
n!\left[z^n u^k\right]R(z,u) &= |s_{n-1,k}| =: R_{n,k}, \\
\mathbf{E}\left(u^{R_n}\right) &= \frac{[u]_{n-1}}{(n-1)!},
\end{aligned}
$$

$$\mathbf{E}(R_n) = \frac{[z^n]\,\partial_u R(z,1)}{C_n} = H_{n-1} \sim \log n, \tag{20}$$

$$\sigma^2(R_n) \sim \log n.$$

A CLT holds.

3.2.2 Ordered (plane) trees

For plane increasing trees, with $\Phi(z) = 1 - \sqrt{1-2z}$, we have

$$\begin{aligned}\partial_z R(z,u) &= \frac{1}{1-u\Phi(z)} = \frac{1}{1-u+u\sqrt{1-2z}},\\ R(z,u) &= \frac{1-\sqrt{1-2z}}{u} + \frac{1-u}{u^2}\log\left(1-u+u\sqrt{1-2z}\right).\end{aligned}$$

Hence,

$$\begin{aligned}n!\left[z^n u^k\right] R(z,u) &= \frac{(2n-k-3)!}{2^{n-k-1}(n-k-1)!}\\ &\sim \sqrt{\frac{2}{\pi n}} e^{-k^2/(4n)},\end{aligned}$$

and

$$\mathbf{P}(R_n = k) = \frac{n!\left[z^n u^k\right] R(z,u)}{(2n-3)!!} = \frac{1}{(2n-3)!!}\frac{(2n-k-3)!}{2^{n-k-1}(n-k-1)!},$$

with

$$\mathbf{E}(R_n) = \frac{[z^n]\,\partial_u R(z,1)}{C_n} \sim \sqrt{\pi n}. \tag{21}$$

4 Height of a size-n tree (Flajolet and Sedgewick, 2009, pp. 216–217)

The height parameter is an extremal one. The height $H_n := h_{\tau_n}$ of a size-n tree τ_n is the height of one of its nodes at largest distance to the root. It obeys

$$h_{\tau_n} = 1 + \max_{\tau \prec \tau_n} h_\tau,$$

where the maximum is over all root sub-trees τ of the full size-n tree τ_n. Taking into account that the number of trees with n nodes and root-degree k is $R_{n,k} = n! r_{n,k}$, with $C_n(h) = n! c_n(h)$ the number of labeled trees' configurations with size n having height less than h, the following bivariate recurrence holds:

$$c_{n+1}(h+1) = \sum_{k=1}^{n} r_{n,k} \sum_{n_1+\ldots+n_k=n} \prod_{r=1}^{k} c_{n_r}(h).$$

There are $\binom{n-1}{k-1}$ terms in the sum over $n_l \geq 1$, $l = 1, \ldots, k$ (the number of ordered partitions of n into k parts). We have $\mathbf{P}(H_n \leq h) = c_n(h)/c_n$.

Alternatively, let $\Phi_h(z)$ be the g.f. of those simple trees generated by ϕ with height $\leq h$. We have

$$\Phi_{h+1}(z) = z\phi(\Phi_h(z)), \qquad \Phi_0(z) = z,$$

where $n![z^n]\Phi_h(z) =: C_h(n)$ is the number of size-n labeled trees whose height is $\leq h$. With $c_h(n) := C_h(n)/n!$ and $\phi_k = \left[z^k\right]\phi(z)$, it holds that

$$c_{h+1}(n+1) = \sum_{k=1}^{n} \phi_k [z^n] \Phi_h(z)^k$$
$$= \sum_{k=1}^{n} \phi_k \sum_{n_1+\ldots+n_k=n} \prod_{l=1}^{k} c_h(n_l). \tag{22}$$

Therefore, with $\overline{N}_h$ the number of trees whose height is less than or equal to h,

$$\mathbf{P}\left(\overline{N}_h = n\right) = \frac{n![z^n]\Phi_h(z)}{n![z^n]\Phi(z)} = \frac{c_h(n)}{c_n},$$

where $\Phi(z) = \Phi_\infty(z)$. We have $\overline{N}_h = n$ if and only if $H_n \leq n$. From implied linear recurrences and a complex analysis based on Mellin transforms:

- Cayley trees: Rényi and Szekeres (1967) have shown that $(2\pi n)^{-1/2} H_n$ has a nontrivial weak limit. Hence $\mathbf{E}(H_n) \sim \sqrt{2\pi n}$.

- For ordered trees (see De Bruijn et al. 1972): $\mathbf{E}(H_n) \sim \sqrt{\pi n}$.

In both cases, $\mathbf{E}(H_n) \sim \lambda\sqrt{n}$, where $\lambda = (\pi\zeta)^{1/2}$ with $\zeta = \frac{2\phi'(\tau)^2}{\phi(\tau)\phi''(\tau)}$ and $\phi(\tau) = \tau\phi'(\tau)$. Furthermore, $H_n/\mathbf{E}(H_n) \stackrel{d}{\to} X$, where X has theta density, with q-moments $(q > 1)$: $\mathbf{E}(X^q) = q(q-1)\Gamma(q/2)\zeta^q$. See Proposition $VII.16$ of Flajolet and Sedgewick (2009).

The e.g.f. $\Phi_h(z)$ of the number of increasing trees whose height is $\leq h$ obeys

$$\Phi_{h+1}(z) = \int_0^z \phi\left(\Phi_h(z')\right) dz', \qquad \Phi_0(z) = z$$

with $C_h(n) = n![z^n]\Phi_h(z)$ and $\mathbf{P}\left(\overline{N}_h = h\right) = \frac{C_h(n)}{C_n}$. A bivariate recurrence similar to (22) holds with $(n+1)c_{h+1}(n+1)$ instead of $c_{h+1}(n+1)$ at the left hand-side.

It holds:

- Cayley increasing trees Pittel (1994): $\mathbf{E}(H_n) \sim e\log n$, $H_n/\log n \to e$ (a.s.).

- Ordered increasing trees Pittel (1994): $\mathbf{E}(H_n) \sim \frac{1}{2s}\log n$, $H_n/\log n \to 1/(2s)$ (a.s.), $se^s = 1$ ($s = 0.27846...$).

Increasing recursive trees' height is much smaller than the one of simple trees; concomitantly, their root degrees are much larger than the ones of simple trees.

5 Depth of a size-n tree

The depth D_n of a size-n tree is the height of a randomly chosen node in this tree. So,

$$D_n = h \text{ w.p. } \frac{N_{h,n}}{n}, \qquad h = 0, \ldots, H_n,$$

where $N_{h,n}$ is the number of nodes at height h in the size-n tree ($N_{0,n}=1$), the profile of trees. It requires the joint law of $(N_{h,n};h=0,\ldots,H_n)$ obeying $\sum_{h=0}^{H_n} N_{h,n}=n$. With $N_{h,\nu}^{(r)} \overset{d}{=} N_{h,\nu}$ i.i.d. copies,

$$N_{h+1,n}=\sum_{r=1}^{R_n} N_{h,\nu_{r,n}}^{(r)}, \qquad h=0,\ldots,H_n.$$

- For Cayley increasing trees, Devroye (1998) and Mahmoud et al. (1993) have shown that $\mathbf{E}(D_n)\sim \log n$, $\sigma^2(D_n)\sim \log n$. A CLT holds.

- For ordered increasing trees, Mahmoud et al. (1993) have shown that $\mathbf{E}(D_n)\sim \frac{1}{2}\log n$, $\sigma^2(D_n)\sim \frac{1}{2}\log n$. A CLT holds.

Equivalently, the mean depth can be obtained as the mean path-length/n, (see Bergeron et al. 1992, pp. 36–37), where the path-length is $\sum_{h=0}^{H_n} N_{h,n}$.

Define some local additive parameters g.f. generated by a sequence a_n's as

$$a(z,u)=\sum_{n\geq 1} C_n u^{a_n}\frac{z^n}{n!}.$$

The global additive parameters generated by the a_n's will be

$$A_{\tau_n}=a_n+\sum_{\tau\prec\tau_n} A_\tau,$$

where the sum is over all root sub-trees τ of the full size-n tree τ_n.

For tree size: $a_n=1$, $a(z,u)=u\Phi(z)$; For path length: $a_n=n$, $a(z,u)=\Phi(zu)$; For leaves: $a_n=\delta_{n,1}$, $a(z,u)=zu$.

With $A(z,u)=\sum_{n\geq 1} C_n u^{A_n}\frac{z^n}{n!}$ the additive parameters global g.f., we get

$$\begin{aligned}\text{Simple} &: \quad A(z,u)=a(z,u)-a(z,1)+z\phi(A(z,u)),\\ \text{Recursive} &: \quad A(z,u)=a(z,u)-a(z,1)+\int_0^z \phi\left(A\left(z',u\right)\right)dz'.\end{aligned}$$

Concerning the mean, with $A_u(z,1)$ denoting the derivative with respect to u evaluated at $u=1$,

$$\text{Simple trees: } A_u(z,1)=a_u(z,1)+z\phi'(A(z,1))A_u(z,1),$$

so (owing to: $A(z,1)=\Phi(z)$) $\Phi'(z)=\phi(\Phi(z))/(1-z\phi'(\Phi(z)))$,

$$A_u(z,1)=\frac{a_u(z,1)}{1-z\phi'(\Phi(z))},$$

where $a_u(z,1)=\sum_{n\geq 1} C_n a_n \frac{z^n}{n!}$ is the e.g.f. of the Hadamard product sequence $C_n a_n$.

$$\text{Recursive trees: } A_u(z,1)=a_u(z,1)+\int_0^z \phi'\left(A\left(z',1\right)\right)A_u\left(z',1\right)dz',$$

whose integrated form is ($a_u(z,1)=\sum_{n\geq 1} C_n a_n \frac{z^n}{n!}$, $a_{u,z}(z,1)=\sum_{n\geq 1} C_n a_n \frac{z^{n-1}}{(n-1)!}$)

$$A_u(z,1)=\Phi'(z)\int_0^z \frac{a_{z,u}(z',1)}{\Phi'(z')}dz'.$$

As a result, we get for the mean path lengths: (simple trees) $a_u(z,1) = z\Phi'(z)$ and (recursive trees) $a_{u,z}(z,1) = \Phi'(z) + z\Phi''(z)$.

- For simple Cayley trees: $\phi(z) = e^z$, $\Phi(z) = ze^{\Phi(z)}$, $z\Phi'(z) = \Phi(z)/(1-\Phi(z))$,

$$A_u(z,1) = \frac{a_u(z,1)}{1 - z\phi'(\Phi(z))} = \frac{\Phi(z)}{(1-\Phi(z))^2},$$

$$[z^n]A_u(z,1) = \frac{1}{n}\left[z^{n-1}\right]\frac{1+z}{(1-z)^3}e^{zn},$$

taking the explicit convolution form:

$$[z^n]A_u(z,1) = \frac{1}{n}(a_. * b_.)_{n-1},$$

with $a_m = m(m+1)^2$ and $b_m = n^m/m!$. Next, by the Stirling formula,

$$c_n := [z^n]\Phi(z) = \frac{n^{n-1}}{n!} \sim \frac{1}{\sqrt{2\pi}}n^{-3/2}e^n,$$

translating that $\Phi(z)$ shows up an algebraic singularity of order $-1/2$ at $z_c = 1/e$. Equivalently, observing $\Phi(z_c) = 1$,

$$\Phi(z) \sim 1 - \sqrt{2}(1 - z/z_c)^{1/2} \text{ as } z \to z_c,$$

and so

$$A_u(z,1) = \frac{\Phi(z)}{(1-\Phi(z))^2} \sim \frac{1}{2}(1 - z/z_c)^{-1} \text{ as } z \to z_c,$$

so with $[z^n]A_u(z,1) \sim 1/2 \cdot e^n$. We conclude that

$$\frac{n!\,[z^n]A_u(z,1)}{C_n} = \frac{[z^n]A_u(z,1)}{c_n} \sim \sqrt{\frac{\pi}{2}}n^{3/2}.$$

The mean depth therefore is $\sqrt{\frac{\pi}{2}}n^{1/2}$.

- For simple ordered trees: $\phi(z) = (1-z)^{-1}$, $\Phi(z) = \left(1 - \sqrt{1-4z}\right)/2$, $\Phi'(z) = (1-4z)^{-1/2}$,

$$\begin{aligned} A_u(z,1) &= \frac{a_u(z,1)}{1 - z\phi'(\Phi(z))} = \frac{z(1-4z)^{-1/2}}{\left(1 - \frac{4z}{\left(1+\sqrt{1-4z}\right)^2}\right)} \\ &= \frac{z(1-4z)^{-1}}{2}\left(1 + \sqrt{1-4z}\right). \end{aligned}$$

Hence,

$$[z^n]A_u(z,1) \sim \left[z^{n-1}\right]\frac{(1-4z)^{-1}}{2} = \frac{4^{n-1}}{2}.$$

Owing to $c_n = [z^n]\Phi(z) = \frac{(2n-2)!}{n!(n-1)!} \sim \frac{1}{\sqrt{\pi}}n^{-3/2}4^{n-1}$,

$$\frac{[z^n]A_u(z,1)}{c_n} \sim \frac{1}{2}\sqrt{\pi}n^{3/2}.$$

The mean depth therefore is $\frac{1}{2}\sqrt{\pi}n^{3/2}/n = \frac{1}{2}\sqrt{\pi}n^{1/2}$.

- Concerning recursive Cayley trees (with $\Phi(z) = -\log(1-z)$ and $\Phi'(z) = 1/(1-z)$),

$$\begin{aligned} A_u(z,1) &= \Phi'(z)\int_0^z \frac{\Phi'(z') + z'\Phi''(z')}{\Phi'(z')}dz' \\ &= \frac{1}{1-z}\int_0^z \left(1 + z'\left[\log\Phi'(z')\right]'\right) \\ &= \frac{-\log(1-z)}{1-z}. \end{aligned}$$

Therefore, $[z^n]A_u(z,1) \sim \log n \Rightarrow n!\,[z^n]A_u(z,1)/(n-1)! = n\log n$. The mean depth therefore is of order $\log n$.

- For recursive ordered trees,

$$A_u(z,1) = \frac{2z - \log(1-2z)}{4(1-2z)^{1/2}}.$$

Owing to $c_n = [z^n]\Phi(z) = [z^n]\left(1 - \sqrt{1-2z}\right) \sim -2^n n^{-3/2}/\Gamma(-1/2)$,

$$[z^n]A_u(z,1) \sim 2^n/\Gamma(1/2)\,n^{-1/2}\log n \Rightarrow [z^n]A_u(z,1)/c_n = \frac{n}{2}\log n.$$

The mean depth therefore is of order $\frac{1}{2}\log n$.

6 Forests (nucleation/aggregation process)

Besides aggregation of atoms to a preexisting tree, growth of trees can also result from the creation of new trees forming a forest. Whenever an incoming atom launches on a new tree, we speak of a nucleation event.

With $\Phi(z,1) = \Phi(z)$ solving $\Phi(z) = z\phi(\Phi(z))$ or $\Phi'(z) = \phi(\Phi(z))$, define

$$K(z,u) = e^{u\Phi(z)} \text{ (or } K(z,u) = 1/(1-u\Phi(z))\text{)}, \tag{23}$$

where u marks the number of trees (connected components) in a forest in that

$$\overline{C}_{n,k} := n!\left[u^k z^n\right]K(z,u) = \frac{n!}{k!}[z^n]\Phi(z)^k \text{ or} \tag{24}$$

$$\overline{C}_{n,k} := n!\left[u^k z^n\right]K(z,u) = n!\,[z^n]\Phi(z)^k$$

is the number of unordered (or ordered) forests with n labeled nodes and k trees, $k = 1,\ldots,n$. Furthermore,

$$\begin{aligned} \overline{C}_n &: = \sum_{k=1}^n \overline{C}_{n,k} = [z^n]K(z,1) = [z^n]e^{\Phi(z)} \text{ (unordered)}, \\ \overline{C}_n &: = \sum_{k=1}^n \overline{C}_{n,k} = [z^n]K(z,1) = [z^n]\frac{1}{1-\Phi(z)} \text{ (ordered)}, \end{aligned}$$

is the number of forests with n atoms, regardless of its number of (in) distinguishable trees.

6.1 Nonincreasing (simple) trees

6.1.1 Cayley (unordered) trees Rényi (1959)

- If $\Phi(z) = ze^{\Phi(z)}$, with $K(z,u) = e^{u\Phi(z)}$, the by Lagrange inversion theorem,

$$\begin{aligned}
[z^n] K(z,u) &= \frac{u}{n} \left[z^{n-1}\right] e^{z(u+n)} \\
&= \frac{u}{n!} (u+n)^{n-1}, \\
n! \left[z^n u^k\right] K(z,u) &=: \overline{C}_{n,k} = \binom{n-1}{k-1} n^{n-k}, \qquad k = 1, \ldots, n, \\
\mathbf{P}(K_n = k) &= \frac{n! \left[z^n u^k\right] K(z,u)}{n! [z^n] K(z,1)} = \frac{1}{(1+n)^{n-1}} \binom{n-1}{k-1} n^{n-k}, \\
\binom{n-1}{k-1} n^{n-k} &= \binom{n}{k} k n^{n-k-1}.
\end{aligned}$$

Takács (1990) rather gave $\overline{C}_{n,k} = kn^{n-k-1}$ as the number of unordered forests with k Cayley trees while fixing the k different founders of the distinct trees out of $\binom{n}{k}$ different ways. See also Rényi (1959). Also,

$$\mathbf{E}\left(u^{K_n}\right) = \frac{n! [z^n] K(z,u)}{n! [z^n] K(z,1)} = u \left(1 - \frac{1}{1+n} + \frac{u}{1+n}\right)^{n-1}, \tag{25}$$

a shifted binomial distribution. In particular, $\mathbf{E}(K_n) = \frac{2n}{1+n} \sim 2$. A remarkable fact is that $R_n \overset{d}{=} K_{n-1}$. Note

$$\mathbf{E}\left(u^{K_n}\right) \sim ue^{-(1-u)} \text{ as } n \text{ is large,} \tag{26}$$

the p.g.f. of a shifted mean 1 Poisson r.v..

We finally observe as in Clarke (1958) that the triangular array $\overline{C}_{n,k}$ obeys the backward recursion

$$(n-k)\overline{C}_{n,k} = nk\overline{C}_{n,k+1}.$$

Hence, $\mathbf{P}(K_n = k) = \overline{C}_{n,k}/\overline{C}_n$, $k = 1, \ldots, n$ obeys

$$\mathbf{P}(K_n = k) = \frac{nk}{n-k} \mathbf{P}(K_n = k+1), \qquad k = n-1, \ldots, 1,$$

with terminal condition $\mathbf{P}(K_n = n) = (n+1)^{-(n-1)}$.

6.1.2 Ordered and unordered forests

- (ordered forests of ordered trees) For such models of forests of trees, the trees themselves are assumed distinguishable, resulting in ordered forests. If $\Phi(z) = \left(1 - \sqrt{1-4z}\right)/2$, solving

$\Phi(z) = z/(1-\Phi(z))$, with $K(z,u) = 1/(1-u\Phi(z))$, then

$$\begin{aligned}
\left[z^n\right] K(z,u) &= \frac{u}{n}\left[z^{n-1}\right]\frac{1}{(1-zu)^2}(1-z)^{-n} \\
&= \frac{1}{n}\sum_{k=1}^{n} k\frac{[n]_{n-k}}{(n-k)!}u^k \\
&= \sum_{k=1}^{n}\frac{k}{n}\binom{2n-k-1}{n-k}u^k, \\
\left[z^n\right] K(z,1) &= e\left[z^n\right] e^{-\sqrt{1-2z}} = \frac{1}{n+1}\binom{2n}{n},
\end{aligned}$$

where we used the identity

$$\left[z^n\right] K(z,1) = \sum_{k=1}^{n}\frac{k}{n}\binom{2n-k-1}{n-k} = \frac{1}{n+1}\binom{2n}{n}.$$

Hence

$$\mathbf{P}(K_n = k) = \frac{n!\left[z^n u^k\right] K(z,u)}{n!\left[z^n\right] K(z,1)} = \frac{n+1}{\binom{2n}{n}}\frac{k}{n}\binom{2n-k-1}{n-k}. \tag{27}$$

Observing

$$\partial_u K(z,1) = \Phi(z)/(1-\Phi(z))^2,$$

$$\begin{aligned}
\left[z^n\right]\partial_u K(z,1) &= \frac{1}{n}\left[z^{n-1}\right]\frac{1+z}{(1-z)^{n+3}} \\
&= \frac{1}{n}\left(\left[z^{n-1}\right](1-z)^{-(n+3)} + \left[z^{n-2}\right](1-z)^{-(n+3)}\right) \\
&= \frac{1}{n}\binom{2n}{n-2}\frac{3n+1}{n-1},
\end{aligned}$$

so

$$\begin{aligned}
\mathbf{E}(K_n) &= \frac{\left[z^n\right]\partial_u K(z,1)}{\left[z^n\right] K(z,1)} = \frac{n+1}{n}\frac{3n-1}{n-1}\frac{\binom{2n}{n-2}}{\binom{2n}{n}} \\
&= \frac{3n-1}{n}\frac{n-1}{n+2} \sim 3.
\end{aligned}$$

The variance can be computed using

$$\partial_u^2 K(z,1) = 2\Phi(z)^2/(1-\Phi(z))^3.$$

We get $\mathbf{E}(K_n(K_n-1)) = 10\frac{n-1}{n+2} \sim 10$, so that $\sigma^2(K_n) \sim 4$ and showing that the limit law is not Poisson.

For such tree models, there is a small number of connected components (at least one of which must be very large).

- (Unordered forests of ordered trees): If $K(z,u) = e^{u\Phi(z)}$, for $k = 1, \ldots, n$, then

$$\begin{aligned} \overline{C}_{n,k} : &= n! \left[z^n u^k\right] K(z,u) = \frac{n!}{k!} [z^n] \Phi(z)^k \\ &= \frac{(n-1)!}{(k-1)!} \left[z^{n-1}\right] z^{k-1} (1-z)^{-n} \\ &= \frac{(n-1)!}{(k-1)!} \left[z^{n-k}\right] (1-z)^{-n} \\ &= \frac{(n-1)!}{(k-1)!} \frac{[n]_{n-k}}{(n-k)!} = \frac{(2n-k-1)!}{(k-1)!\,(n-k)!}. \end{aligned}$$

These $\overline{C}_{n,k}$'s are readily checked to obey the three term recurrence

$$\overline{C}_{n+1,k} = (2n-k)\,\overline{C}_{n,k} + \frac{n}{k-1} \overline{C}_{n,k-1}.$$

Furthermore,

$$\begin{aligned} \overline{C}_n &= n! [z^n] e^{\Phi(z)} = (n-1)! \left[z^{n-1}\right] e^z (1-z)^{-n} \\ &= (n-1)! \sum_{m=1}^{n} \frac{1}{(m-1)!} \left[z^{n-m}\right] (1-z)^{-n} = \sum_{k=1}^{n} \frac{(2n-k-1)!}{k!\,(n-k-1)!} \\ &= \sum_{l=1}^{n} \frac{(n+l-1)!}{(l-1)!\,(n-l)!}. \end{aligned} \tag{28}$$

6.2 Increasing trees

6.2.1 Cayley (unordered) trees

- In the case of nonplane increasing trees for which $\Phi(z) = -\log(1-z)$,

$$\begin{aligned} K(z,u) &= e^{u\Phi(z)} = (1-z)^{-u}, \\ n! [z^n] K(z,u) &= [u]_n, \\ n! \left[z^n u^k\right] K(z,u) &= |s_{n,k}|, \\ \mathbf{E}\left(u^{K_n}\right) &= \frac{[u]_n}{n!}. \end{aligned}$$

Here $\overline{C}_{n,k} = |s_{n,k}| \equiv S_{n,k}(-1,0,0)$ are the signless Stirling numbers of the first kind, obeying

$$\overline{C}_{n+1,k} = n\overline{C}_{n,k} + \overline{C}_{n,k-1}, \tag{29}$$

$$\mathbf{E}(K_n) = \frac{[z^n]\,\partial_u K(z,1)}{C_n} = H_n \sim \log n,$$

$$\sigma^2(K_n) \sim \log n.$$

A CLT holds.

6.2.2 Ordered (plane) trees

- In the case of ordered increasing trees for which $\Phi(z) = 1 - \sqrt{1-2z}$ solving $\Phi(z) = 1/(1-\Phi(z))$,

$$K(z,u) = e^{u\left(1-\sqrt{1-2z}\right)}. \tag{30}$$

Here, with $S_{n,k}(-\alpha_2, -\alpha_1; w_2)$ the generalized Stirling triangle defined in Hsu and Shiue (1998),

$$\begin{aligned}
n!\left[z^n u^k\right] K(z,u) &= \frac{n!}{k!}\left[z^n\right]\Phi(z)^k = \overline{C}_{n,k} := S_{n,k}(-2,-1;0),\\
\overline{C}_{n,k} &\equiv S_{n,k}(-2,-1,0) = (2(n-k)-1)!!\binom{2n-k-1}{2(n-k)}\\
&= \frac{(2(n-k)-1)!}{2^{n-k-1}(n-k-1)!}\binom{2n-k-1}{2(n-k)} = \frac{1}{2^{n-k}}\frac{(2n-k-1)!}{(n-k)!(k-1)!},\\
\overline{C}_{n+1,k} &= (2n-k)\overline{C}_{n,k} + \overline{C}_{n,k-1}.
\end{aligned}$$

Hence,

$$\mathbf{P}(K_{n+1} = k) = \frac{\overline{C}_n}{\overline{C}_{n+1}}\left((2n-k)\mathbf{P}(K_n = k) + \mathbf{P}(K_n = k-1)\right). \tag{31}$$

With $\overline{C}_n := \sum_{k=1}^{n} \overline{C}_{n,k} = \sum_{l=0}^{n-1} 2^{-l}\frac{(n+l-1)!}{l!(n-l-1)!}$,

$$\mathbf{P}(K_n = k) = \frac{n!\left[z^n u^k\right] K(z,u)}{n!\left[z^n\right] K(z,1)} = \frac{\overline{C}_{n,k}}{\overline{C}_n},$$

$$\overline{C}_{n+1} = (2n+1)\overline{C}_n - \langle \overline{C}_n \rangle,$$

$$\frac{\overline{C}_{n+1}}{\overline{C}_n} \sim 2n-1 \Rightarrow \mathbf{E}(K_n) = \frac{n!\left[z^n\right]\partial_u K(z,1)}{n!\left[z^n\right] K(z,1)} = \frac{\langle \overline{C}_n \rangle}{\overline{C}_n} \sim 2,$$

K_n has a shifted Poisson(1) limit law. We have

$$K(z,u) = e^u\left(1 + \sum_{k\geq 1}\frac{(-u)^k}{k!}(1-2z)^{k/2}\right),$$

with singularity at $z = 1/2$. Singularity analysis of the singular expansion yields

$$\begin{aligned}
n!\left[z^n\right] K(z,u) &= n! e^u\left(\sum_{k\geq 1}\frac{(-u)^k}{k!}\frac{n^{-k/2}2^n}{n}\frac{1}{\Gamma(-k/2)}\right)\\
&\sim (n-1)! 2^n e^u\left(\frac{un^{-1/2}}{2\sqrt{\pi}} + O\left(n^{-3/2}\right)\right).
\end{aligned}$$

Therefore ,

$$\begin{aligned}
\overline{C}_n &= n!\left[z^n\right] K(z,1) \sim (n-1)! 2^n e\left(\frac{n^{-1/2}}{2\sqrt{\pi}} + O\left(n^{-3/2}\right)\right),\\
\frac{\overline{C}_{n+1}}{\overline{C}_n} &\sim 2n\left(1 - \frac{1}{2n} + O\left(n^{-2}\right)\right) \sim 2n-1.
\end{aligned}$$

Therefore,

$$\mathbf{E}\left(u^{K_n}\right) = \frac{n!\,[z^n]\,K(z,u)}{n!\,[z^n]\,K(z,1)} \sim ue^{-(1-u)}, \tag{32}$$

the p.g.f. of a shifted Poisson(1) r.v..

This three-term recurrence, giving $\mathbf{P}(K_n = k)$ in all cases, is the one of triangular Markov probability sequences. The latter recursion (31), for example, may be written as

$$\mathbf{P}(K_{n+1} = k) = q_{k,k}^{(n)}\mathbf{P}(K_n = k) + q_{k-1,k}^{(n)}\mathbf{P}(K_n = k-1),$$

defining the transition coefficients $q_{k,k}^{(n)}$ and $q_{k-1,k}^{(n)}$ (not probabilities). Introduce the superdiagonal transition matrix

$$Q_{n,n} := \begin{pmatrix} q_{1,1}^{(n)} & q_{1,2}^{(n)} & 0 & \cdots \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & q_{n-1,n-1}^{(n)} & q_{n-1,n}^{(n)} \\ \vdots & \vdots & 0 & q_{n,n}^{(n)} \end{pmatrix},$$

and let $Q_{n-1,n}$ be its $(n-1)\times n$ truncated version. With $\pi_1^K := 1$, taking into account the boundary conditions, the distributions $\pi_n^K := (\mathbf{P}(K_n = k),\ k \in \{1,\ldots,n\})$, $n \geq 1$, satisfy the recursion $\pi_n^K = \left(\pi_{n-1}^K, 0\right) Q_{n,n} = \pi_{n-1}^K Q_{n-1,n}$, $n \geq 2$. Thus, $\pi_n^K = \overleftarrow{\prod}_{m=2,\ldots,n} Q_{m-1,m}$, $n \geq 2$ is an integrated form solution of the recursion, as a left product of nested rectangular matrices. Because for each n, π_n^L is a probability vector, we get that π_n^L is orthogonal to the 1-shifted column sum vector $\mathbf{q}_n - \mathbf{1}$ of $Q_{n,n+1}$ with kth entry $q_{k,k}^{(n+1)} + q_{k,k+1}^{(n+1)} - 1$, $k = 1,\ldots,n$.

7 Weighted trees

Simply generated weighted trees are weighted versions of rooted trees and have been introduced by Meir and Moon (1978). They are obtained while assigning to each node x of a size-n tree τ_n a weight $w_{b_n(x)}$, where $b_n(x)$ is the outdegree of x. The weight of a particular tree τ_n is then the product $\prod_{x\in\tau_n} w_{b_n(x)}$, and while summing over all τ_n, we get the weight of all size-n trees.

7.1 Weighted simple trees

Let $w(\tau_n) = \prod_{b=0}^{n-1} w_b^{n_b(\tau_n)}$ be the (multiplicative) weight of an unlabeled rooted tree τ_n with n nodes ($|\tau_n| = n$) having $n_b(\tau_n)$ nodes with outdegree (branching number) b. The weight $w(\tau_n)$ is the product over the n nodes x of τ_n of the $w_{b_n(x)}$'s, where $b_n(x)$ is the outdegree of x. Then $W_n = \sum_{\tau_n} w(\tau_n)$ is the weight of all size-n such trees associated to the weight sequence $\mathbf{w} := (w_b \geq 0,\ b \geq 0)$. The number of these trees is $c_n := C_n/n! = [z^n]\,\Phi(z)$. Let $\Phi_{\mathbf{w}}(z) = \sum_{n\geq 1} z^n W_n$. Then $\Phi_{\mathbf{w}}(z)$ solves $\Phi_{\mathbf{w}}(z) = zg(\Phi_{\mathbf{w}}(z))$, where $g(z) = \sum_{b\geq 0} w_b z^b$.

By Lagrange inversion formula, we get the identity

$$W_n = [z^n]\,\Phi_{\mathbf{w}}(z) = \frac{1}{n}\left[z^{n-1}\right] g(z)^n = \sum_{\tau_n}\prod_{x\in\tau_n} w_{b_n(x)} = \sum_{\tau_n}\prod_{b=0}^{n-1} w_b^{n_b(\tau_n)}.$$

Examples of $\mathbf{w}$ are as follows:

- $\mathbf{w} = (\varepsilon_b;\ b \geq 0)$ with $\varepsilon_b \in \{0,1\}$. Only the branches of the tree with $\varepsilon_b = 1$ contribute to its weight.

- $w_b = a_1 a_2^b$, $a_1, a_2 > 0$ with $g(z) = a_1 \sum_{b\geq 0} (a_2 z)^b = a_1 / (1 - a_2 z)$. Note $w(\tau_n) = a_1^n a_2^{n-1}$ in view of $\sum_{b=0}^{n-1} n_b(\tau_n) = n$ and $\sum_{b=1}^{n-1} b n_b(\tau_n) = n-1$ (the total tree length). As a result, $W_n = a_1^n a_2^{n-1} c_n$, where $c_n = C_n / n! = \frac{(2n-2)!}{n!(n-1)!}$. In that separable case, each tree τ_n has equal weight $a_1^n a_2^{n-1}$ and $W_n = a_1^n a_2^{n-1} c_n$. Note $\Phi_{\mathbf{w}}(z) = \left(1 - \sqrt{1 - 4a_1 a_2 z}\right) / (2a_2) = \Phi(a_1 a_2 z) / a_2$.

- If the sequence of weights w_b is summable, then $g(z)$'s can be normalized to yield the p.g.f.'s $g(z)/g(1)$; the corresponding integrated solutions change from $\Phi_{\mathbf{w}}(z)$ to $\Phi_{\mathbf{w}}(z/g(1))$.

- If $\mathbf{w} := (w_b \geq 0,\ b \geq 0)$ is directly a probability sequence with $\sum_{b\geq 0} w_b = 1$, then $\Phi_{\mathbf{w}}(z)$ is the g.f. of the total progeny of a Galton–Watson tree with branching mechanism g, solving $\Phi_{\mathbf{w}}(z) = z g(\Phi_{\mathbf{w}}(z))$, $\Phi_{\mathbf{w}}(0) = 0$. The numbers $[z^n] \Phi_{\mathbf{w}}(z) = W_n$ are the (sub-)probabilities of a progeny with size n.

Examples:

• (Poisson offspring) If $w_b = e^{-\mu} \mu^b / b!$, $b \geq 0$ with $\mu > 0$, then, with $g(z) = e^{\mu(z-1)}$ and $\Phi(z)$ solving $\Phi(z) = z e^{\Phi(z)}$, the Cayley e.g.f., we have

$$\Phi_{\mathbf{w}}(z) = \frac{1}{\mu} \Phi\left(\mu e^{-\mu} z\right) = \frac{1}{\mu} \sum_{n\geq 1} \frac{n^{n-1}}{n!} \left(\mu e^{-\mu} z\right)^n.$$

The Borel e.g.f. $\Phi_{\mathbf{w}}(z)$ has thus now a displaced algebraic singularity of order $-1/2$ at $z_c = \frac{1}{\mu} e^{\mu - 1} > 1$ with $\Phi_{\mathbf{w}}(z_c) = 1/\mu$ and $\Phi'_{\mathbf{w}}(z_c) = \infty$. Note

$$\begin{aligned} \mu &\leq 1 : \ \Phi_{\mathbf{w}}(1) = 1, \\ \mu &> 1 : \ \Phi_{\mathbf{w}}(1) = \rho < 1/\mu, \end{aligned}$$

where ρ is the extinction probability of the supercritical Poisson GW process solving $g(\rho) = e^{\mu(\rho - 1)} = \rho$.

• (geometric offspring) If $w_b = \overline{a} a^b$, $b \geq 0$ with $a \in (0,1)$ and $\overline{a} = 1 - a$, then, with $g(z) = \overline{a} / (1 - az)$ and $\Phi(z)$ solving $\Phi(z) = z / (1 - \Phi(z))$, the e.g.f. of ordered trees, we have

$$\Phi_{\mathbf{w}}(z) = \frac{1}{a} \Phi(\overline{a} a z) = \frac{1 - \sqrt{1 - 4\overline{a} a z}}{2a}.$$

Also, $\Phi_{\mathbf{w}}(z)$ has thus now a displaced algebraic singularity of order $-1/2$ at $z_c = 1/(4\overline{a} a) > 1$ with $\Phi_{\mathbf{w}}(z_c) = 1/(2a)$ and $\Phi'_{\mathbf{w}}(z_c) = \infty$. Note, with $\mu = a/\overline{a}$, the mean offspring number is

$$\begin{aligned} \mu &\leq 1 : \ \Phi_{\mathbf{w}}(1) = 1, \\ \mu &> 1 : \ \Phi_{\mathbf{w}}(1) = \rho < 1/(2a), \end{aligned}$$

where $\rho = \overline{a}/a$ is the extinction probability of the supercritical geometric GW process solving $g(\rho) = \overline{a} / (1 - a\rho) = \rho$.

In both cases, there is a possibility for such weighted trees that they are finite with some nonzero probability.

7.2 Weighted increasing trees

Consider the increasing ordered tree g.f. $\Phi(z)$ solving $\Phi'(z) = 1/(1-(\Phi(z)))$, $\Phi(0) = 0$, with $\Phi(z) = 1-\sqrt{1-2z}$ and

$$C_n = n!\,[z^n]\,\Phi(z) = (2n-3)!! = 2^{-(n-1)}\,(2n-2)!/(n-1)!.$$

Let $W_n = \sum_{\tau_n} w(\tau_n)$ (where $w(\tau_n) = \prod_{b=0}^{n-1} w_b^{n_b(\tau_n)}$) be the weight of all size n increasing labeled rooted trees ($|\tau_n| = n$) associated to the weight sequence $\mathbf{w} := (w_b \geq 0,\ b \geq 0)$, with $w_b = [z^b]\,g(z)$ the weight of an atom with out-degree b. Let then $\Phi_{\mathbf{w}}(z) = \sum_{n\geq 1} z^n W_n/n!$, with $W_n = n!\,[z^n]\,\Phi_{\mathbf{w}}(z)$. Then $\Phi_{\mathbf{w}}(z)$ solves $\Phi'_{\mathbf{w}}(z) = g(\Phi_{\mathbf{w}}(z))$, where $g(z) = \sum_{b\geq 0} w_b z^b$. Note $\Phi(z)$ is obtained when $w_b = 1$, $b \geq 0$: the number of such n-trees is $C_n = (2n-3)!!$.

The probability to observe a particular size-n tree τ_n among all size-n trees is $w(\tau_n)/W_n$. The tilted probability to observe a tree of size n among all possible trees is

$$\frac{z^n W_n/n!}{\Phi_{\mathbf{w}}(z)},$$

for those $z \in (0, z_c)$, where $z_c = \inf(z > 0 : \Phi_{\mathbf{w}}(z) = \infty) = \int_0^{z_*} \frac{dz'}{g(z')}$ with $z_* = \inf(z > 0 : g(z) = \infty)$.

Examples of $\mathbf{w}$ are as follows:

- $w_b = a_1 a_2^b [c]_b/b!$, $a_1, a_2, c > 0$ with $g(z) = a_1 \sum_{b\geq 0} \frac{[c]_b}{b!}(a_2 z)^b = a_1(1-a_2 z)^{-c}$.

- $w_b = a_1 a_2^b$, $a_1, a_2 > 0$ with $g(z) = a_1/(1-a_2 z)$. In that separable case, each tree τ_n has equal weight $a_1^n a_2^{n-1}$ and $W_n = a_1^n a_2^{n-1} C_n$ with $W_{n+1}/W_n = a_1 a_2 (2n-1)$.

For the three special g.f.'s,

$$\begin{aligned} g(z) &= (1-\alpha_1 z)^{-(\alpha_2/\alpha_1 - 1)}\ (\alpha_2 > \alpha_1 > 0, 0 < \alpha := \alpha_1/\alpha_2 < 1), \\ g(z) &= e^{\alpha_2 z} \text{ (obtained from the latter } g(z) \text{ when } \alpha_1 \to 0\text{), or} \\ g(z) &= (1+\alpha_1 z)^d, \qquad \alpha_1 > 0, d \in \{2, 3, \ldots\}, \end{aligned}$$

the weighted e.g.f.'s $\Phi_{\mathbf{w}}(z)$ solving $\Phi'_{\mathbf{w}}(z) = g(\Phi_{\mathbf{w}}(z))$, $\Phi_{\mathbf{w}}(0) = 0$ are, respectively,

$$\begin{aligned} \Phi_{\mathbf{w}}(z) &= \frac{1}{\alpha_1}\left(1-(1-\alpha_2 z)^{\alpha_1/\alpha_2}\right) \\ &= -\log(1-\alpha_2 z) \text{ as } \alpha_1 \to 0 \\ &= \frac{1}{\alpha_1}\left(-1+[1-\alpha_1(d-1)z]^{-1/(d-1)}\right). \end{aligned}$$

Note that these special $g(z)$'s can be normalized to the p.g.f.'s $g(z)/g(1)$ with the corresponding integrated solutions changing from $\Phi_{\mathbf{w}}(z)$ to $\Phi_{\mathbf{w}}(z/g(1))$. The normalized p.g.f.'s $g(z)/g(1)$ are then, respectively, negative-binomial, Poisson, and binomial p.g.f.'s.

The formation of such increasing trees admits the following recursive tree evolution scheme (label 1 is assigned to the root):

With probability $p_b(n) := K_n^{-1}(b+1)w_{b+1}/w_b$, attach uniformly node $n+1$ to any of the $n_b(\tau_n)$ nodes with out-degree $b \in \{0, \ldots, b^*\}$ of a previous size-n increasing tree τ_n ($b^* \leq n-1$).

The normalization constant is $K_n = \sum_{b=0}^{n-1} n_b(\tau_n)(b+1)w_{b+1}/w_b$, representing the "number" of ways the new atom with label $n+1$ can be inserted in τ_n. This preferential (or uniform) attachment procedure results in a realization of τ_{n+1}; see Panholzer and Prodinger (2007). With $(B_b(n_b(\tau_n)), b \in \{0,\dots,b^*\})$ mutually exclusive Bernoulli r.v.'s (summing to 1), each with success probability $n_b(\tau_n)\, p_b(n)$, for each $n \geq 1$, we have

$$\begin{aligned} n_0(\tau_{n+1}) &= n_0(\tau_n) + 1 - B_0(n_0(\tau_n)), \\ n_b(\tau_{n+1}) &= n_b(\tau_n) + B_{b-1}(n_{b-1}(\tau_n)) - B_b(n_b(\tau_n)), \qquad b \in \{1,\dots,b^*\}, \\ n_{b^*+1}(\tau_{n+1}) &= 0 + B_{b^*}(n_{b^*}(\tau_n)). \end{aligned}$$

Whenever a connection to a node with outdegree b occurs, the number of nodes with out-degree b (respectively, $b+1$) decreases (increases) by one unit. In addition, a new node with outdegree-0 is always created whatever the degree of the node to which the new incoming atom connects to τ_n . Here $n_0(\tau_n)$ is the number L_n of leaves in a size-n tree.

From the first equation above giving the evolution of the number of **leaves**,

$$\begin{cases} L_{n+1} = L_n + 1 \text{ with probability } 1 - L_n p_0(n), \\ L_{n+1} = L_n \text{ with probability } L_n p_0(n). \end{cases} \tag{33}$$

This is one of standard space-time inhomogeneous Markov chains.

For the three particular $\Phi_{\mathbf{w}}$-models generated by the special g's above, using $\sum_{b=0}^{n-1} n_b(\tau_n) = n$ and $\sum_{b=1}^{n-1} b n_b(\tau_n) = n-1$ for any τ_n, we get

$$p_b(n) = \frac{1-\alpha+\alpha b}{n-\alpha},\ \frac{1}{n},\ \frac{d-b}{1+n(d-1)},$$

respectively, depending only on (b,n) and not on the full sequence of weights $(w_b; b = 0,\dots,b^*)$. In the first two examples, $b \in \{0,\dots,b^* = n-1\}$ while $b \in \{0,\dots,b^* = d-1\}$ in the third d-ary labeled plane trees case. The number C_n of such d-ary plane trees is obtained while integrating $\Phi'_{\mathbf{w}}(z) = g_d(\Phi_{\mathbf{w}}(z))$, where $g_d(z) := (1+z)^d$. We get

$$\Phi_{\mathbf{w}}(z) = -1 + [1-(d-1)z]^{-1/(d-1)}, \text{ with}$$

$$C_n = n!\,[z^n]\,\Phi_{\mathbf{w}}(z) = (d-1)^n\,[1/(d-1)]_n = [1:d-1]_n,$$

where $[a:b]_n := a(a+b)\dots(a+(n-1)b)$.

Note that for $g(z) = (1-\alpha_1 z)^{-(\alpha_2/\alpha_1 - 1)}$, $z_* = 1/\alpha_1$, $z_c = 1/\alpha_2$ and $\Phi_{\mathbf{w}}(z_c) = z_*$.

Such increasing trees may serve as models for phylogenetic trees in which nodes represent species and labels encode their order of appearance, and thus the chronology of evolution. The leaves of the tree are the currently living species; the different trees consist of genera.

Remark:(33) is the weighted tree extension of (17) and (14) obtained, respectively, when $\alpha_2 = 1 > \alpha_1 = 0$ and $\alpha_2 = 2 > \alpha_1 = 1$.

The joint e.g.f. of the number of atoms and the number of trees of the **forests** is

$$K_{\mathbf{w}}(z,u) = e^{u\Phi_{\mathbf{w}}(z)},$$

with

$$\overline{C}_{n,k} = n!\left[z^n u^k\right] K_{\mathbf{w}}(z,u).$$

The generalized Stirling numbers

$$\begin{aligned}\overline{C}_{n,k} &\equiv S_{n,k}(-\alpha_2,-\alpha_1;0), \qquad k=0,\ldots,n,\\ \overline{C}_{n,k} &\equiv 0 \text{ if } k>n,\end{aligned}$$

are defined by the identity Hsu and Shiue (1998)

$$[w:\alpha_2]_n = \sum_{k=0}^{n} \overline{C}_{n,k}\,[w:\alpha_1]_k.$$

The $\overline{C}_{n,k}$ obey the (Stirling's triangle) recurrence

$$\overline{C}_{n+1,k} = (n\alpha_2 - k\alpha_1)\,\overline{C}_{n,k} + \overline{C}_{n,k-1} \tag{34}$$

with boundary conditions $\overline{C}_{n,0} = \delta_{n,0}$ and $\overline{C}_{n,n} = 1$ for all $n \geq 0$. With $\mathbf{P}(K_n = k) = \frac{n!\left[z^n u^k\right]K_{\mathbf{w}}(z,u)}{n!\left[z^n\right]K_{\mathbf{w}}(z,1)} = \frac{\overline{C}_{n,k}}{\overline{C}_n}$, therefore

$$\mathbf{P}(K_{n+1}=k) = \frac{\overline{C}_n}{\overline{C}_{n+1}}\left[(n\alpha_2 - k\alpha_1)\,\mathbf{P}(K_n=k) + \mathbf{P}(K_n=k-1)\right], \tag{35}$$

where $\overline{C}_n := \sum_{k=0}^{n}\overline{C}_{n,k} = n!\,[z^n]\,K_{\mathbf{w}}(z,1) = n!\,[z^n]\,e^{\Phi_{\mathbf{w}}(z)}$. The e.g.f. $K_{\mathbf{w}}(z,1) = e^{\Phi_{\mathbf{w}}(z)}$ has an algebraic singularity of order $-\alpha$ at $z_c = 1/\alpha_2$; by singularity analysis therefore

$$\begin{aligned}\overline{C}_n &\sim -(n-1)!\,\frac{1}{\alpha_1\Gamma(-\alpha)}\,\alpha_2^n n^{-\alpha},\\ \frac{\overline{C}_n}{\overline{C}_{n+1}} &\sim \frac{1}{n\alpha_2}\left(1+O\left(n^{-1}\right)\right), \qquad \text{for large } n.\end{aligned}$$

Summing (34) over k,

$$\overline{C}_{n+1} = (n\alpha_2+1)\,\overline{C}_n - \alpha_1\left\langle \overline{C}_n\right\rangle,$$

$$\frac{\overline{C}_n}{\overline{C}_{n+1}} \sim \frac{1}{n\alpha_2} \Rightarrow \mathbf{E}(K_n) = \frac{n!\,[z^n]\,\partial_u K_{\mathbf{w}}(z,1)}{n!\,[z^n]\,K_{\mathbf{w}}(z,1)} = \frac{\left\langle\overline{C}_n\right\rangle}{\overline{C}_n} \sim 1/\alpha_1.$$

Multiplying (35) by k and summing over k yield $\mathbf{E}\left(K_n^2\right) \sim (1+1/\alpha_1)/\alpha_1$; hence $\sigma^2(K_n) \sim 1/\alpha_1 \sim \mathbf{E}(K_n)$.

By the Lagrange inversion formula, we have

$$\begin{aligned}[z^n]\,K_{\mathbf{w}}(z,1) &= \frac{1}{n}\left[z^{n-1}\right]e^z g(z)^n\\ &= \frac{1}{n}\left(a_{\cdot} * b_{\cdot}\right)_{n-1},\end{aligned}$$

where $a_m = 1/m!$ and
$b_m = [z^m]\,g(z)^n = [z^m]\,(1-\alpha_1 z)^{-n(1/\alpha-1)} = \alpha_2^m\,[n(1-\alpha)]_m/m!$.

Proposition 1 (condensed phase). *If* $g(z) = (1-\alpha_1 z)^{-(\alpha_2/\alpha_1 - 1)}$ *with* $\alpha_2 > \alpha_1 > 0$, $0 < \alpha := \alpha_1/\alpha_2 < 1$, *then* K_n *has a Poisson limit-law with mean* $1/\alpha_1$ *as* $n \to \infty$. *It corresponds to the distribution*

$$\mathbf{P}(K_n = k) = \frac{\overline{C}_{n,k}}{\overline{C}_n}, \qquad k = 0, \ldots, n,$$

of the normalized weight of unordered forests with n labeled atoms and k increasing ordered trees, resulting from a random uniform choice of a configuration proportional to its weight.

Remark: (35) is the weighted tree extension of (29) and (31) obtained, respectively, when $\alpha_2 = 1 > \alpha_1 = 0$ and $\alpha_2 = 2 > \alpha_1 = 1$. When $\alpha_1 \to 0$, K_n diverges logarithmically, with both $\mathbf{E}(K_n)$ and $\sigma^2(K_n) \sim \alpha_2 \log n$.

- Finally, we mention a random version of increasing trees. With $m > 0$, $\mathbf{w} := (w_b = m\pi_b \geq 0,\ b \geq 0)$, where (π_b) is a probability sequence $(\sum_{b\geq 0} \pi_b = 1)$, $\Phi_{\mathbf{w}}(z)$ is the e.g.f. of the total progeny of a Galton–Watson increasing tree with branching mechanism $g(z) = \sum_{b\geq 0} \pi_b z^b$ and parameter m. If the total progeny is finite with probability (w.p.) 1, $\Phi_{\mathbf{w}}(1) = 1$ and $\Phi'_{\mathbf{w}}(1) = m$ is the mean number of its nodes, fixing $m = \int_0^1 \frac{dz'}{g(z')}$.

The numbers $W_n/n! = [z^n]\Phi_{\mathbf{w}}(z) = m^n \sum_{\tau_n} \prod_{b=0}^{n-1} \pi_b^{n_b(\tau_n)}/n!$ are then the probabilities of a progeny with size n. The e.g.f. $\Phi_{\mathbf{w}}(z)$ solves

$$z = \frac{1}{m}\int_0^{\Phi_{\mathbf{w}}(z)} \frac{dz'}{g(z')} =: \frac{1}{m} P(\Phi_{\mathbf{w}}(z)).$$

By Lagrange inversion formula,

$$n!\,[z^n]\,\Phi_{\mathbf{w}}(z) = m^n (n-1)!\,\left[z^{n-1}\right]\left(\frac{z}{P(z)}\right)^n = W_n.$$

With $z_* = \sup(z > 0 : g(z) < \infty) \in [1, \infty]$, the convergence radius of $\Phi_{\mathbf{w}}(z)$ is

$$z_c = \sup(z > 0 : \Phi_{\mathbf{w}}(z) < \infty) = \frac{1}{m}\int_0^{z_*} \frac{dz'}{g(z')},$$

with $\Phi_{\mathbf{w}}(z_c) = z_*$ owing to

$$z_c - z = \frac{1}{m}\int_{\Phi_{\mathbf{w}}(z)}^{z_*} \frac{dz'}{g(z')}.$$

Assuming $m < \int_0^{z_*} \frac{dz'}{g(z')}$ then $z_c > 1$. If so, $\Phi_{\mathbf{w}}(1) < \infty$ and $\Phi_{\mathbf{w}}(z)$ is a g.f. candidate. It is a defective p.g.f. only if $m < \int_0^1 \frac{dz'}{g(z')}$ because then $\rho := \Phi_{\mathbf{w}}(1) < 1$ (in view of $1 = \frac{1}{m}\int_0^{\Phi_{\mathbf{w}}(1)} \frac{dz'}{g(z')}$): the model is supercritical, having ρ as its extinction probability. The total progeny is finite only w.p. ρ and $\Phi'_{\mathbf{w}}(1) = mg(\rho)$ is the mean number of its nodes on the extinction set. Fixing the parameter m to its critical upper value $m_c = \int_0^1 \frac{dz'}{g(z')}$ entails $\Phi_{\mathbf{w}}(1) = 1$, $\Phi'_{\mathbf{w}}(1) = m_c > 1$. The size of the corresponding tree is then finite w.p. 1, with mean value m_c. Note that if $m = 1 < m_c = \int_0^1 \frac{dz'}{g(z')}$, then $\Phi_{\mathbf{w}}(1) < 1$ for all p.g.f.'s g.

When $m = m_c$, the e.g.f. $\Phi_{\mathbf{w}}(z)$ is proper. It admits a closed form integrable expression in the following cases (respectively, binomial, Poisson, and negative-binomial):

$$
\begin{aligned}
g(z) &= (\pi_0+\pi_1 z)^d \;;\; \Phi_{\mathbf{w}}(z)=\frac{\pi_0}{\pi_1}\left((1-z/z_c)^{-1/(d-1)}-1\right),\; z_c=1/\left(1-\pi_0^{d-1}\right),\\
g(z) &= e^{-\mu(1-z)} \;;\; \Phi_{\mathbf{w}}(z)=-\frac{1}{\mu}\log(1-z/z_c)\,,\; z_c=1/\left(1-e^{-\mu}\right),\\
g(z) &= \left(\frac{q}{1-pz}\right)^{\theta} \;;\; \Phi_{\mathbf{w}}(z)=\frac{1}{p}\left(1-(1-z/z_c)^{1/(\theta+1)}\right), \qquad z_c=1/\left(1-q^{\theta+1}\right).
\end{aligned}
$$

With $\alpha,\lambda\in(0,1)$, suppose $g(z)=1-\lambda(1-z)^{\alpha}$ with $w_b=\alpha\lambda\,[1-\alpha]_{b-1}/b!$, $b\geq 1$ and $z_*=1$ (a Sibuya branching mechanism with infinite mean μ). In that case, $z_c=\frac{1}{m}\int_0^1\frac{dz'}{g(z')}=\frac{1}{m}\sum_{k\geq 0}\frac{\lambda^k}{1+k\alpha}<\infty$, with $z_c>1$ if and only if $m<\int_0^1\frac{dz'}{g(z')}=m_c$. When $m=m_c>1$, $z_c=1$ and $\Phi_{\mathbf{w}}(1)=1$. There is no major impact of the mean value μ of the branching mechanism on the extinction possibility of the increasing tree.

Acknowledgements

The author is indebted to his collaborators Francois Dunlop and Arif Mardin for many fruitful discussions during the elaboration of this manuscript.

References

Bergeron, F., Flajolet, P. and Salvy, B. (1992). Varieties of increasing trees. In: CAAP'92: 17th Colloquium on Trees in Algebra and Programming Rennes, **17** (pp. 24–48). Springer Berlin Heidelberg.

Clarke, L. E. (1958). On Cayley's formula for counting trees. *Journal of the London Mathematical Society*, **33**, 471–474.

De Bruijn, N. G., Knuth, D. E., and Rice, S. O. (1972). The average height of planted plane trees. In: *Graph Theory and Computing*, R. C. Read, Ed., Academic Press, pp. 15–22.

Devroye, L. (1998). Universal limit laws for depth in random trees. *SIAM Journal on Computing*, **28**, 409–432.

Drmota, M. (2009) *Random Trees. An Interplay Between Combinatorics and Probability.* Springer, New York.

Flajolet, P. and Sedgewick, R. (2009). *Analytic Combinatorics.* Illustrated Edition. Cambridge University Press.

Hsu, L. C. and Shiue, P. J.-S. (1998). A unified approach to generalized Stirling numbers. *Advances in Applied Mathematics*, **20**, 366–384.

Mahmoud, H., Smythe, R. T. and Szymanski, J. (1993). On the structure of random plane-oriented recursive trees and their branches. *Random Structures and Algorithms*, **4**, 151–176.

Meir, A. and Moon, J. W. (1978). On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, **30**, 997–1015.

Neveu, J. (1986) Arbres et processus de Galton-Watson. *Annales Institue Henri Poincaré Probabilités et Statistiques* **22**, 199–207.

Pitman, J. (1998). Enumerations of trees and forests related to branching processes and random walks. In: *Microsurveys in Discrete Probability* Edited by: D Aldous, J Propp. 163–180 Providence RI: American Mathematical Society.

Pittel, B. (1994). Note on the heights of random recursive trees and random m-array search trees. *Random Structures and Algorithms*, **5**, 337–347.

Rényi, A. (1959). Some remarks on the theory of trees. *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, **4**, 73–85.

Rényi, A. and Szekeres, G. (1967). On the height of trees. *Australian Journal of Mathematics*, **7**, 497–507.

Steele, J. M. (1987). Gibbs'measures on combinatorial objects and the central limit theorem for an exponential family. *Probability in the Engineering and Informational Sciences*, **1**, 47–59.

Surya, E. and Warnke, L. (2023). Lagrange inversion formula by induction. *The American Mathematical Monthly*, **130**, 944–948.

Takács, L. (1990). On Cayley's formula for counting forests. *Journal of Combinatorial Theory*, Series A, **53**, 321–323.

Panholzer, A. and Prodinger, H. (2007). Level of nodes in increasing trees revisited. *Random Structures and Algorithms*, **31**, 203–226.

Wolfram: https://mathworld.wolfram.com/Tree.html