Strict monotonic trees arising from evolutionary processes: combinatorial and probabilistic study

Olivier Bodini, Antoine Genitrini, Cécile Mailler, Mehdi Naima

STRICT MONOTONIC TREES ARISING FROM EVOLUTIONARY PROCESSES: COMBINATORIAL AND PROBABILISTIC STUDY

OLIVIER BODINI, ANTOINE GENITRINI, CÉCILE MAILLER, AND MEHDI NAIMA

Abstract. In this paper we introduce three new models of labelled random trees that generalise the original unlabelled Schröder tree. Our new models can be seen as models for phylogenetic trees in which nodes represent species and labels encode the order of appearance of these species, and thus the chronology of evolution. One important feature of our trees is that they can be generated efficiently thanks to a dynamical, recursive construction.

Our first model is an increasing tree in the classical sense (labels increase along each branch of the tree and each label appears only once). To better model phylogenetic trees, we relax the rules of labelling by, e.g., allowing repetitions in the two other models.

For each of the three models, we provide asymptotic theorems for different characteristics of the tree (e.g. degree of the root, degree distribution, height, etc), thus giving extensive information about the typical shapes of these trees. We also provide efficient algorithms to generate large trees efficiently in the three models. The proofs are based on a combination of analytic combinatorics, probabilistic methods, and bijective methods (we exhibit bijections between our models and well-known models of the literature such as permutations and Stirling numbers of both kinds).

Keywords: Evolution process; Increasing trees; Monotonic trees; Analytic Combinatorics; Uniform sampling.

Date: June 11, 2020.

This research is partially supported by the ANR project *MetACOnc*, ANR-15-CE40-0014. C. Mailler aknowledges EPSRC for support through the fellowship EP/R022186/1.

Contents

1. Introduction

The aim of this paper is to introduce new combinatorial models for phylogenetic trees: the main idea is to add node labels in order to encode chronology in the classical model introduced by Schröder trees in 1870 in the seminal paper *Vier Combinatorische Probleme* [Sch70]. In this paper (see the fourth problem), Schröder introduces a simple model of *phylogenetic tree model*, and enumerate this class of trees by their number of leaves.

In biology, a phylogenetic tree is a classical tool to represent the evolutionary relationship among species. At each branching node of the tree, the descendant species from distinct branches have distinguished themselves in some manner and are no more dependent: the past is shared but the futures are independent.

The main limitation of Schröder's model of phylogenetic trees is that it does not take into account the chronology between the different branching nodes. Since then, probabilistic approaches have been developed to consider this chronology: in particular in the context of binary trees, one can mention, e.g., the stochastic model of Yule [Yul25] and its generalization by Aldous [Ald96].

However, to the best of our knowledge, there seems to have been no attempt to combinatorially enrich Schröder's original model in order to encode the chronology of evolution: this is the aim of this paper. To do so, we consider labelled versions of Schröder's tree, where the labels represent the order at which branchings occur. In Figure 1 we have represented on the left hand-side a classical Schröder trees of size 50 (i.e. with 50 leaves), and, on the right hand-side, a labelled version of the same tree: time is on the vertical axis, from top to bottom, and a node of label x is placed at time x (the horizontal placement is irrelevant).

Discussion of related models: Increasing trees are classical in the literature: for example, Bergeron, Flajolet and Salvy [BFS92] study several families of increasingly-labelled trees, and, to do so, they develop some tools that are now classical in *analytic combinatorics*. As an example, one of these classical tools is the integration of the Greene operators. We refer the reader to [Drm09] where more recent results on various families of increasing trees and the analytic combinatorics methods to quantitatively study them are surveyed.

Beyond their combinatorial description, increasing trees can often be described as the result of a dynamic construction: the nodes are added one by one at successive integer-times in the tree (their labels being the time at which they are added). This description sometimes allow to apply probabilistic method to prove theorems about some characteristics such as the height of the tree, and it also often gives a very efficient way to generate large trees from the considered class using simple, iterative and local rules.

We illustrate this dynamical evolution on the simple case of *recursive trees*. This simple model was originally designed as a simple model for the spread of epidemics [Moo74]. Combinatorially, a recursive tree is a rooted non-plane (i.e. the order of siblings is irrelevant) tree whose nodes are labelled from 1 to the number of nodes in such a way that each label appears exactly once, and the labels increase along all branches. We denote by $\mathcal{R}_n$ the class of all n-node recursive trees. Now, consider a sequence of random trees $(t_n)_{n\geq 1}$ built recursively as follows: t_0 has only one node, labelled by 1. Given t_{n-1}, attach a new child labelled by n to a node picked uniformly at random among the $n-1$ nodes of t_{n-1}. Then, it is

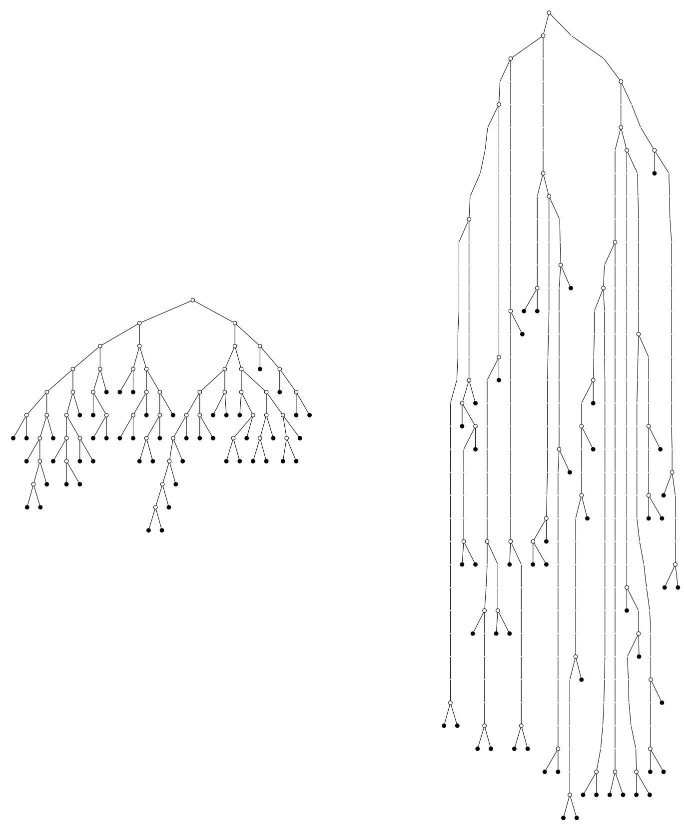


FIGURE 1. A Schröder tree: without chronological evolution (on the left-hand side), and with chronological evolution (on the right-hand side): the label of a node is represented as the distance from this node to the root.

known that for all $n \geq 1$, t_n is a tree taken uniformly at random in $\mathcal{R}_n$, the set of all n-node recursive trees. Both analytic combinatorics and probabilistic methods, as well as a bijection with permutations, have been used to understand the typical shape of a large recursive tree: it is known that the degree of the root grows as $\ln n$ (see [Drm09, Sec. 6.1]), the height as $c \ln n$ (for an explicit constant c – see [Pit94]), the proportion of nodes of arity $k \geq 0$ converges to 2^{-k} (see [Drm09, Th. 6.8]). Although our three models of increasing Schröder trees are more involved, our proofs rely on the same three methods used in the literature for the recursive trees: analytic combinatorics, a dynamical evolution and probabilistic methods, and bijections with classes of permutations.

Our main contributions: Although, as mentioned above, many variations of the recursive tree have been studied, this paper (together with its short version [BGN19]) contains the first study of increasing versions of the classical model of Schröder. We aim at defining an evolution process associating to a given Schröder tree structure an evolution represented by an increasing labelling of its internal nodes. Furthermore we also focus on relaxing the labelling constraints by allowing repetitions of labels. In the dynamical construction of the trees, allowing repetition of labels mean allowing adding several nodes at once in the tree. Our generalisations can be seen as natural discrete-time versions of the classical probabilistic model of Yule trees (see, e.g., [SM01]) where the time between two branchings are exponentially distributed.

This work is a part of a long-term over-arching project, in which we aim at relaxing the classical rules of increasing labelling (described in, e.g., [BFS92]), by, for example, allowing labels to appear more than once in the tree. The following papers are part of this strand of research: [BGGW20, BGNS20] introduce and study models of label trees with less-constrained increasing labelling rules, but also other graphs structures like [BDF$^+$16] focuses on increasingly-labelled "diamonds" and [GGKW20] at a compacted structure that specifies classes of directed acyclic graphs.

In this paper, we introduce three new different models of Schröder trees with chronological evolution: the *increasing Schröder trees*, the *strict monotonic Schröder trees* and the *strict monotonic general trees*. One important feature of these models is that they can all be simulated efficiently. The first two models are based on some *increasingly labelling* of Schröder trees, repetition of labels is allowed in the second model. In the last model increases we increase the expressivity by allowing a new type of internal nodes. For all of the three models, we prove asymptotic results about important characteristics of a typical large tree of this class (e.g. root distribution, number of nodes of arity 2, 3, etc, height of the tree, etc – see Table 1 where our main results are summarised), and design an algorithm that generates a large tree taken uniformly at random among all trees of a given size in the class. The quantitative analysis of the three models and the design of the random samplers rely on a combination of analytic combinatorics methods (see [FS09] for a survey), probabilistic methods (in particular methods developed by Devroye [Dev90] to study the height of *split trees*), and bijective methods (we exhibit bijections between our classes of trees and classes of permutations, these are then useful for the analysis of different characteristics and for the design of the generation algorithms). In particular, we exhibit interesting relations between Striling numbers and parameters on trees such that the labelling of nodes, the number of internal nodes, and the depth of a leaf.

Generic approach highlighted in the paper: Similarly to the recursive tree, all of our three models have a *generic constrained evolution process.* The specificity of each model is induced by small changes of the evolution process: we give here a generic, non precise description of the evolution process, details specific to each family of trees will be detailed in each section:

- *Start with a single (unlabelled) leaf;*
- Iterate the following process: *at step ℓ (for $\ell \geq 1$), select a subset of leaves and replace each selected leaf by an internal node with label ℓ attached to an arbitrary sequence of leaves.*

Note that the increasing labelling corresponds to the chronology of the construction of the tree: internal nodes labelled by an integer ℓ were added at time ℓ. Our three models differ from each other by different constraints on the set of selected leaves: in our first model, this subset is always of size 1, while it can be bigger in the other two models. The difference between our second model and third model is that internal nodes have arity at least 2 on the second model, while they can have arity 1 in the third model. Importantly, in all three models our Schröder trees can be seen as phylogenetic trees of n species (n being the number of leaves): the labels of internal nodes stand for the times at which different branches of the phylogenetic trees split.

	Number of trees	Distinct labels	Internal nodes	Depth LM leaf	Height
Increasing Schröder trees	$n!/2$	$n - \ln n$	$n - \ln n$	$\ln n$	$\Theta(\ln n)$
Strict monotonic Schröder trees	$(n-1)!/(2\,(\ln 2)^n)$	$0.72\ n$	$n - 2\ \ln n$	$\ln n$	
Strict monotonic general trees	$c\ (n-1)!\ 2^{(n-1)(n-2)/2}$	$\Theta(n)$	$\Theta(n^2)$	$\Theta(n)$	$\Theta(n)$

Table 1. Summary of the main analytic results of the paper: behaviour of the characteristics of a large typical tree of each of the three classes of labelled Schröder trees. The parameter n stands for the size of the trees (i.e. their number of leaves) and the results are asymptotic when $n \to +\infty$. (LM stands for "leftmost" and c is a positive constant.)

Plan of the paper: Each of the three main sections (Sections 2, 3 and 4) is dedicated to one of our three new models of labelled Schröder trees. The organisation inside each section is similar: after defining the model we show theorems about different characteristics of the trees using analytic combinatorics and bijective methods. We then exhibit the associated dynamical evolution that generates the considered class of trees, and use this evolution process to (a) design an efficient random sampler for this class of trees and (b), in some cases, to prove some probabilistic results about the height of a typical large tree from this class.

2. Increasing Schröder trees

The first model we are interested in is a generalisation of the Schröder tree, a classical combinatorial structure that was originally introduced in the context of phylogenetics [Sch70]. Our generalisation consists in labelling the internal nodes of a Schröder tree – denote by ℓ their number – with the integers $\{1, \dots, \ell\}$ with the constraints that each label appears exactly once and a node's label is larger than the label of its parent; such a labelling of a tree is called "increasing", we call such a constrained-labelled Schröder tree an *increasing Schröder tree.* In the tree seen as an evolutionary process, the labels can be interpreted as the order of appearance of the different nodes (which, for example, stand for different species). Several classes of increasingly-labelled trees have already been studied in the literature using *analytic combinatorics* [FS09] methods, but these methods applied to the Schröder tree would raise important technical problems. The novelty of our approach is to use a dynamical description of the increasing Schröder tree inspired by its evolutionary interpretation; this allows us to give the first analytical results about this combinatorial structure.

2.1. The model and its context.

In this paper, we define rooted trees as genealogical structures: the root is the unique common ancestor of all nodes of the tree, each node except the root has exactly one parent (the root has no parent), nodes that have no children are called *leaves*, nodes that have at least one child are called *internal nodes*. The *arity* of a node is it's number of children. We say that a tree is *plane* if siblings (nodes that have the same parent) are ordered.

Definition 2.1.1 (see [FS09, p. 69])**.** *A Schröder tree is a rooted plane tree whose internal nodes all have arity at least* 2*. The size of a Schröder tree is its number of leaves.*

Note that a Schröder tree is an unlabelled combinatorial structure (neither the leaves nor the internal nodes are labelled). In the context of analytic combinatorics the combinatorial class $\mathcal{S}$ of Schröder trees is thus specified as

$$\mathcal{S} = \mathcal{Z} \cup \textsc{Seq}_{\geq 2}\,\mathcal{S}. \tag{1}$$

Its combinatorial meaning is direct in the context of decomposable objects (see Flajolet and Sedgewick [FS09] for a detailed introduction to the combinatorial specification): An object from $\mathcal{S}$ is either a leaf (represented by the single atom $\mathcal{Z}$, of size 1), or it is composed of an internal node, parent of a sequence of at least two elements from $\mathcal{S}$. Not that, in the specification, the internal nodes are omitted (because they are of size 0): the expression $\textsc{Seq}_{\geq 2}\,\mathcal{S}$ is a abbreviation of $\mathcal{E} \times \textsc{Seq}_{\geq 2}\,\mathcal{S}$ where $\mathcal{E}$ stands for an atom of size 0 and $\textsc{Seq}_{\geq 2}\,\mathcal{S}$ is a sequence of at least two elements from $\mathcal{S}$.

Once the combinatorial specification is given, the classical *symbolic method* [FS09], translates automatically the equation specifying the objects into a functional equation satisfied by the (ordinary) generating functions associated to the structures. The generating function of $\mathcal{S}$ is defined as the formal series $S(z) = \sum_{n\geq 1} s_n z^n$ where s_n is the number of Schröder trees of size n (i.e. with n leaves). Using the symbolic method on Equation (1), we get that

$$S(z) = z + \frac{S(z)^2}{1 - S(z)}. \tag{2}$$

An elementary iteration allows us to extract the first coefficients of the sequence $(s_n)_{n\in\mathbb{N}}$:

$$(0, 1, 1, 3, 11, 45, 197, 903, 4279, 20793, 103049, 518859, 2646723, 13648869, 71039373, \dots).$$

Equation (2) implies that the generating function S is algebraic and in fact

$$S(z) = \frac{1 + z - \sqrt{1 - 6z + z^2}}{4}.$$

This is sufficient to get the following asymptotic equivalent of s_n when n tends to infinity:

$$s_n = \frac{\sqrt{3\sqrt{2} - 4}}{4\sqrt{\pi}}\, n^{-3/2} \left(3 - 2\sqrt{2}\right)^{-n} \left(1 + \mathcal{O}(1/n)\right).$$

We refer the reader to [FS09, page 69] for a more detailed analysis of this generating function S.

In the rest of the section we are interested in an increasingly-labelled variation of Schröder trees.

Definition 2.1.2. *An* increasing Schröder tree *has a Schröder tree structure and its internal nodes are labelled with the integers between* 1 *and* ℓ *(where* ℓ *is the number of internal nodes) in such a way that each label appears exactly once and each sequence of labels in the paths from the root to any leaf is (strictly) increasing.*

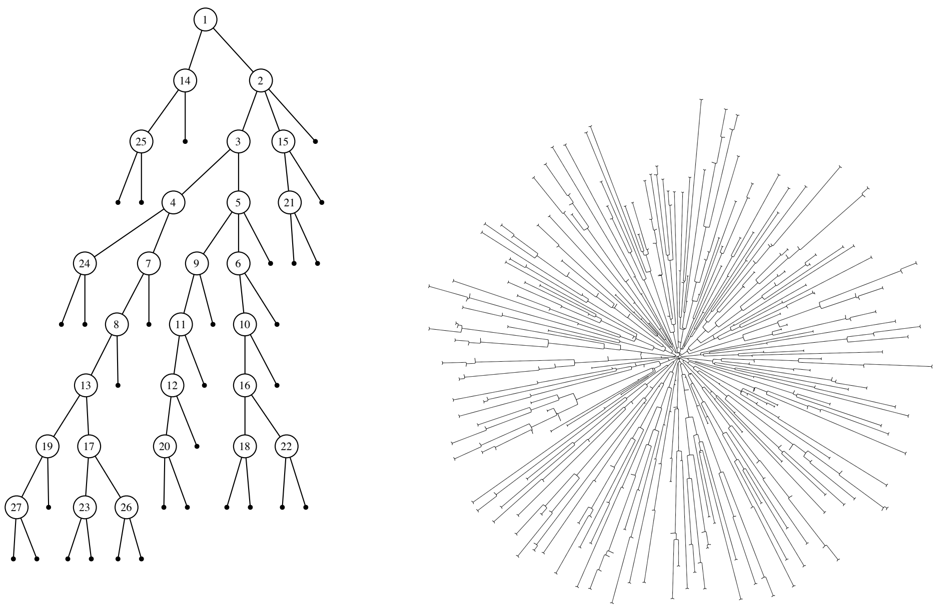


FIGURE 2. Two increasing Schröder trees

Increasing trees can, to a certain extent, be specified using the Greene operator $^\square\star$ (see, for example, [FS09, page 139]), and the specification can then be translated into an equation satisfied by the exponential generating function of the increasing tree class. Since in our context the size of a tree is the number of its leaves while only internal nodes are labelled, we need to introduce a second variable u to mark the internal nodes. Let us denote by $s_{n,\ell}$ the number of increasing Schröder trees with n leaves and ℓ internal nodes. Following standard methods in analytic combinatorics we define a generating function that is ordinary for the leaf marks and exponential for the internal node marks: we set $S^*(z,u) = \sum_{n,\ell} s_{n,\ell}\, z^n\, u^\ell/\ell!$. The specification of this combinatorial class is

$$\mathcal{S}^* = \mathcal{Z} \cup \mathcal{U}^\square \star \operatorname{SEQ}_{\geq 2} \mathcal{S}^*.$$

Using the symbolic method [FS09], we obtain the following equation satisfied by $S^*(z,u)$:

$$S^*(z,u) = z + \int_{v=0}^{u} \frac{S^*(z,v)^2}{1 - S^*(z,v)} \mathrm{d}v.$$

Although this integral equation could be analysed further in order to get information about increasing Schröder trees, this analysis would be very cumbersome; a better approach is to see the Schröder tree as the result of an evolutionary process. Another advantage of this new approach is that it extends to other families of labelled Schröder trees for which there seems to be no (classical) specification, even using the Greene operator: one such example is the family of *strict monotonic Schröder trees* studied in Section 3.

In Figure 2 we have represented two increasing Schröder trees: both are generated uniformly at random among all increasing Schröder trees of the same size:

size 30 on the left, size 500 on the right. The left-hand-side tree has 27 internal nodes (and 30 leaves). It is the same tree as the one represented in Figure 1, where its chronological evolution is represented on the right-hand side: the internal node labelled by ℓ is displayed on level $\ell - 1$ (i.e. at distance $\ell - 1$ from the root on the vertical axis), for all $\ell \in \{1, \dots, 27\}$. The right-hand-side one is drawn using a circular representation, which is often used for phylogenetic trees: the labels are omitted but as in Figure 1, the length of an edge is proportional to the difference of the labels of the two nodes it connects. This right-hand-side tree has 492 internal nodes (and 500 leaves).

Let us introduce an evolution process generating increasing Schröder trees:

- *Start with a single (unlabelled) leaf;*
- Iterate the following process: *at step ℓ (for $\ell \geq 1$), select one leaf and replace it by an internal node with label ℓ attached to an arbitrary sequence of new leaves.*

To make sure that this algorithm generates all increasing Schröder trees and generates each tree exactly "once"; we define this evolution process more rigorously as follows: The process takes as an input two sequences of integers $(d_\ell)_{\ell\geq 1}$ and $(u_\ell)_{\ell\geq 1}$ such that $u_1 = 1$ and for all $\ell \geq 1$, $d_\ell \geq 2$ and $1 \leq u_{\ell+1} \leq \sum_{i=1}^{\ell} d_i - (\ell - 1)$ and gives as an output a sequence $(\tau_\ell)_{\ell\geq 0}$ of ℓ-internal-node increasing Schröder trees. The process is defined inductively as follows:

- Tree τ_0 is the 1-node tree, without any internal node.
- Given τ_ℓ, we define $\tau_{\ell+1}$ as follows: we number the leaves of τ_ℓ in the depth-first order (the choice of the ordering does not matter) from 1 to $\sum_{i=1}^{\ell} d_i - (\ell - 1)$, and replace leaf number $u_{\ell+1}$ by an internal node labelled by $\ell + 1$ to which $d_{\ell+1}$ leaves are attached.

Note that, by construction, τ_ℓ is an increasing Schröder tree with ℓ internal nodes for all $\ell \geq 0$, and the label of a node corresponds to the time in the evolution process when this node became an internal node. In other words, the increasing labelling corresponds to the chronology of the evolution process. Finally, note that the evolution process indeed defines a bijection between the set of increasing Schröder trees having p internal nodes and the set of all sequences $(d_\ell, u_\ell)_{1\leq \ell\leq p}$ such that $u_1 = 1$, and for all $1 \leq \ell \leq p$, $d_\ell \geq 2$ and $1 \leq u_{\ell+1} \leq \sum_{i=1}^{\ell} d_i - (\ell - 1)$.

Recall that we define the size of a Schröder tree to be its number of leaves. It is important to note that, because a Schröder tree with $n - 1$ internal nodes has at least n leaves, the evolution process defines a bijection between the set of all n-leaf Schröder trees and the set of all sequences $(d_\ell^{(n)}, u_\ell^{(n)})_{1\leq \ell<n}$ such that for all $1 \leq \ell < n$, $u_1^{(n)} = 1$, $d_\ell^{(n)} \geq 2$, $1 \leq u_{\ell+1}^{(n)} \leq \sum_{i=1}^{\ell} d_i^{(n)} - (\ell - 1)$, and $\sum_{i=1}^{\ell} d_i^{(n)} = n$.

By taking all trees of the same size together, we obtain the following induction equation, enumerating increasing Schröder trees by size: if, for all $n \geq 0$, t_n is the number of n-leaf Schröder trees, then $t_1 = 1$ and, for all $n \geq 2$,

$$t_n = \sum_{\ell=1}^{n-1} \ell \, t_\ell. \tag{3}$$

2.2. **Exact enumeration and relationship with permutations.** Let $\mathcal{T}$ denote the class of increasing Schröder trees. Using the evolution process, we get the following specification for $\mathcal{T}$:

$$\mathcal{T} = \mathcal{Z} \cup \left(\Theta\mathcal{T} \times \textsc{Seq}_{\geq 1}\,\mathcal{Z}\right). \tag{4}$$

In this specification, $\mathcal{Z}$ stands for the leaves, and the operator Θ is the classical pointing operator (see [FS09, page 86] for details). The specification is a direct rewriting of the evolution process: a tree is either of size 1 ($\mathcal{Z}$), or it has been built by pointing a leaf in a smaller tree ($\Theta\mathcal{T}$) and replacing it by a sequence of at least two leaves. Although the latter sequence is of length at least 2, we use the operator $\textsc{Seq}_{\geq 1}(\mathcal{Z})$ instead of $\textsc{Seq}_{\geq 2}(\mathcal{Z})$ because the leaf that was pointed is reused as the leftmost child of the new internal node.

The symbolic method translates this specification into a functional equation satisfied by the generating function associated to the combinatorial class of increasing Schröder trees. Note that although the increasing Schröder trees are labelled, this labelling is transparent, i.e. it is possible to work with ordinary generating functions (as opposed to exponential generating functions). This is because the size of an increasing Schröder tree is its number of leaves, and the leaves are not labelled. We define the *ordinary generating function* associated to $\mathcal{T}$ by $T(z) = \sum_{n\geq 1} t_n\, z^n$, where t_n is the number of increasing Schröder trees of size n. Using the symbolic method (in particular, pointing at a leaf translates into a differential operator), we get

$$T(z) = z + \frac{z^2}{1-z}\, T'(z). \tag{5}$$

Writing $(1-z)T(z) = z(1-z) + z^2T'(z)$ and extracting the n-th coefficient on both sides of this equation, we get that for all $n \geq 3$, $t_n = n\ t_{n-1}$: we get back the recurrence exhibited earlier in Equation (3). Using the fact that $t_1 = t_2 = 1$, we get that $t_n = n!/2$ for all $n \geq 2$ (this sequence $(t_n)_n$ appears under the reference OEIS A001710[1]). Note that the radius of convergence of the ordinary generating series $T(z)$ is 0; this series is thus purely formal.

2.3. **Analysis of typical parameters.** In this section, our aim is to describe the *shape* of a *typical* increasing Schröder tree, i.e. a tree taken uniformly at random among all increasing Schröder tree of a fixed size. To get information about this *shape*, we focus on four characteristics of the tree: the number of internal nodes, the arity of the root, the number of leaves that are children of the root, and the number of binary nodes (node of arity 2). We show asymptotic theorems for these characteristics in a typical increasing Schröder tree when the size goes to infinity.

2.3.1. *Quantitative analysis of the number of iteration steps.* In this section, we show that although an increasing Schröder tree of size n can have between 1 and $n-1$ internal nodes, it typically has of order $n - \ln n$ internal nodes. This result is particularly interesting to analyse the complexity of the evolutionary process: this means that, on average, this evolutionary process takes of order $n - \ln n$ iteration steps to generate a typical increasing Schröder tree of size n. In fact, our result is stronger than just finding an equivalent for the average number of iterations since we prove a central limit theorem for this quantity. To complete the picture we also quantify the average number of nodes of a fixed degree. We will show that the

[1]Throughout this paper, a reference OEIS A$\cdots$ points to Sloane's Online Encyclopedia of Integer Sequences www.oeis.org.

1						
0 ,	1					
0 ,	1,	2				
0 ,	1,	5,	6			
0 ,	1,	9,	26,	24		
0 ,	1,	14,	71,	154,	120	
0 ,	1,	20,	155,	580,	1044,	720

TABLE 2. Values of $t_{n,k}$ (the number of increasing Schröder trees with n leaves and k internal nodes) for $n \in \{1, 2, \dots, 7\}$, and $k \in \{0, 1, \dots, n-1\}$.

average number of binary nodes in a typical tree is $n - 2\ln n$, the number of ternary nodes is $\ln n$ and higher arity nodes have a constant mean.

Theorem 2.3.1. *For all $n \geq 1$, we denote by X_n the number of internal nodes in a tree taken uniformly at random among all increasing Schröder trees of size n. Then, asymptotically when n tends to infinity, $\mathbb{E}_{\mathcal{T}_n}[X_n] \sim n - \ln n$, $\mathbb{V}_{\mathcal{T}_n}[X_n] \sim \ln n$, and*

$$\frac{X_n - (n - \ln n)}{\sqrt{\ln n}} \xrightarrow{d} \mathcal{N}(0,1) \qquad \textit{in distribution.}$$

To prove this theorem, we enrich the specification (4) with an additional parameter $\mathcal{U}$ marking the internal nodes:

$$\mathcal{T} = \mathcal{Z} \cup \left(\mathcal{U} \times \Theta_{\mathcal{Z}}\mathcal{T} \times \mathrm{SEQ}_{\geq 1}\,\mathcal{Z}\right),$$

where the operator $\Theta_{\mathcal{Z}}$ consists in pointing an element marked by $\mathcal{Z}$. Remark here we do not use the Greene operator: the increasing labelling is a consequence of our point of view, we do not need to care about it. Using the symbolic method, this implies that, if $t_{n,k}$ is the number of increasing Schröder trees with n leaves and k internal nodes, $t_n(u) = \sum_{k=0}^{n-1} t_{n,k}\, u^k$, and $T(z,u) = \sum_{n\geq 1} t_n(u)\, z^n$, then

$$T(z,u) = z + \frac{uz^2}{1-z}\,\partial_z T(z,u), \tag{6}$$

where ∂_z denotes the partial differentiation according to z. Once again, we write $(1-z)T(z,u) = z(1-z) + uz^2$, and extract the coefficient of z^n on both sides; let us denoted by $t_n(u) = \sum_{k=0}^{n-1} t_{n,k}\, u^k$, then this gives $t_1(u) = 1$, $t_2(u) = u$ and, for all $n > 2$,

$$t_n(u) = (1 + (n-1)u)\; t_{n-1}(u). \tag{7}$$

Extracting the coefficient of u^k on both sides of this last equation gives: $t_{1,0} = 1$, $t_{n,1} = 1$ for all $n > 1$,

$$t_{n,k} = t_{n-1,k} + (n-1)\; t_{n-1,k-1} \qquad \text{for all } 0 < k < n,$$

and $t_{n,k} = 0$ otherwise. The first values of $t_{n,k}$ are listed in Table 2. Note that, for all $n \geq 1$, $t_{n,n-1}$ is the number of increasing binary trees (see [FS09, page 143] for details).

From Equation (7), we easily deduce a closed form for $t_n(u)$: for all $n \geq 2$, we have

$$t_n(u) = u \prod_{\ell=2}^{n-1} (1 + \ell u). \tag{8}$$

This is a shifted version of the sequence OEIS A145324, which is related to Stirling cycle numbers. Our proof of Theorem 2.3.1 relies on the following lemma, which is a straightforward consequence of Equation (8).

Lemma 2.3.2. *Let $SC_n(u) = \prod_{i=0}^{n-1}(u+i)$ be the generating functions of the respective rows of the Stirling Cycle numbers (see* [FS09, page 735]*), which enumerate all permutations of a set of size n that decompose into k cycles (i.e. Stirling numbers of the first kind). If we set $\hat{t}_n(u) = \sum_{k=1}^{n} t_{n,k}\, u^{n-k}$, which is the row-reversed generating function, then*

$$\hat{t}_n(u) = \frac{SC_n(u)}{1+u} = u \prod_{\ell=2}^{n-1} (u+\ell).$$

Proof of Theorem 2.3.1. One could apply Hwang's quasi-powers theorem [Hwa98], but since we have an explicit formula for $t_n(u)$, we decide instead to apply Lévy's continuity theorem directly. By Lemma 2.3.2, we have that, if $\bar{X}_n = n - X_n$, for all $\xi \in \mathbb{R}$,

$$\mathbb{E}\left[\mathrm{e}^{i\xi \cdot \frac{\bar{X}_n - \ln n}{\sqrt{\ln n}}}\right] = \frac{1}{t_n}\, \mathrm{e}^{-i\xi\sqrt{\ln n}} \hat{t}_n\left(\mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}\right) = \frac{2}{n!}\, \mathrm{e}^{-i\xi\sqrt{\ln n} + \frac{i\xi}{\sqrt{\ln n}}} \cdot \frac{\Gamma\left(n + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}\right)}{\Gamma\left(2 + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}\right)}$$

$$= \frac{2 + o(1)}{\Gamma(3 + o(1))}\, \frac{\left(n - 1 + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}\right)^{n + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}} - \frac{1}{2}} \mathrm{e}^{n} \mathrm{e}^{-i\xi\sqrt{\ln n}}}{\mathrm{e}^{n - 1 + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}}\, n^{n + \frac{1}{2}}},$$

where we have used Stirling's formula. Note that

$$\lim_{n\to\infty} \mathrm{e}^{1 - \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}} = 1,$$

and $\Gamma(3) = 2$, which implies that

$$\mathbb{E}\left[\mathrm{e}^{i\xi \cdot \frac{\bar{X}_n - \ln n}{\sqrt{\ln n}}}\right] = (1 + o(1)) \frac{\left(n - 1 + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}\right)^{n + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}} - \frac{1}{2}} \mathrm{e}^{-i\xi\sqrt{\ln n}}}{n^{n + 1/2}}$$

$$= (1 + o(1))\, \frac{n^{\mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}}}{n} \left(1 + O\left(\frac{1}{n\sqrt{\ln n}}\right)\right)^{n - \frac{1}{2} + \mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}} \mathrm{e}^{-i\xi\sqrt{\ln n}}$$

$$= (1 + o(1))\, \frac{n^{\mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}} \mathrm{e}^{-i\xi\sqrt{\ln n}}}{n}.$$

Since

$$\begin{aligned} n^{\mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}} &= \exp\left((\ln n)\mathrm{e}^{\frac{i\xi}{\sqrt{\ln n}}}\right) \\ &= \exp\left((\ln n)\left(1+\frac{i\xi}{\sqrt{\ln n}}-\frac{\xi^2}{2\ln n}+O\left((\ln n)^{-3/2}\right)\right)\right) \\ &\underset{n\to\infty}{=} n\mathrm{e}^{i\xi\sqrt{\ln n}-\xi^2/2}, \end{aligned}$$

we get

$$\mathbb{E}\left[\mathrm{e}^{i\xi\cdot\frac{\bar{X}_n-\ln n}{\sqrt{\ln n}}}\right] = (1+o(1))\,\mathrm{e}^{-\xi^2/2},$$

which, by Lévy's continuity theorem concludes the proof; recall that $\bar{X}_n = n - X_n$. □

2.3.2. *Quantitative characteristics of the root node.* In this section, we study two parameters of the root of a typical increasing Schröder tree: the total number of its children (i.e. its arity), and the number of its children that are leaves.

Theorem 2.3.3. *Let A_n be the arity of the root of a tree taken uniformly at random among all increasing Schröder trees of size n. For all $n \geq 2$, $k \geq 2$, we have*

$$\mathbb{P}(A_n = k) = \frac{2k}{(k+1)!}.$$

And the second result is:

Theorem 2.3.4. *Let L_n be the number of children of the root that are leaves in a tree taken uniformly at random among all increasing Schröder trees of size n. Asymptotically when n tends to infinity,*

$$\mathbb{E}[L_n] = \frac{2\mathrm{e}}{n} + \Theta\left(1/(n\cdot n!)\right) \quad \textit{and} \quad \mathbb{V}[L_n] = \frac{2\mathrm{e}}{n} + \Theta\left(1/n^2\right).$$

Theorem 2.3.3 is a direct consequence of the following lemma.

Lemma 2.3.5. *If $t_{n,k}$ is the number of increasing Schröder trees whose root has arity k, then $t_{1,0} = 1$, for all $n \geq 0, t_{n,1} = 0$ and for all $n \geq 2, 2 \leq k \leq n-1$,*

$$t_{n,k} = \frac{k\ n!}{(k+1)!}, \qquad \textit{and } t_{n,n} = 1.$$

Indeed, this lemma together with the fact that $t_n = n!/2$, imply, for all $2 \leq k < n$,

$$\mathbb{P}(A_n = k) = \frac{2k}{(k+1)!},$$

which concludes the proof of Theorem 2.3.3.

We refer the reader to Table 3 where the first values of $t_{n,k}$ are listed. The sequences $(t_n(u))_{n\geq 1}$ and $(t_{n,k})_{2\leq k<n}$ are related to the sequences OEIS A094112 and OEIS A092582, which enumerate some families of permutations (the former enumerates a family of permutations avoiding some pattern, the second permutations with initial cycle of a given size). Since the number of increasing Schröder trees of size $n \geq 2$ is equal to $n!/2$, it is natural to expect some links between this family of trees and permutations: in Section 2.4 we exhibit a bijection between the two families.

1,	0						
0,	0,	1					
0,	0,	2,	1				
0,	0,	8,	3,	1			
0,	0,	40,	15,	4,	1		
0,	0,	240,	90,	24,	5,	1	
0,	0,	1680,	630,	168,	35,	6,	1

TABLE 3. Values of $t_{n,k}$, the number of size-n increasing Schröder trees of root-arity k, and $0 \leq k \leq n \in \{1, \ldots, 7\}$.

Proof of Lemma 2.3.5. In this proof, the variable $\mathcal{U}$ marks the arity of the root (we re-use the same notation as in the previous section, but with a different meaning; this is done to avoid having too many different notations). Using the evolution process, we get that

$$\mathcal{T} = \mathcal{Z} \cup \big(\mathcal{U} \times \mathcal{Z} \times \mathrm{SEQ}_{\geq 1}(\mathcal{U} \times \mathcal{Z})\big) \cup \big(\Theta_{\mathcal{Z}}(\mathcal{T} \setminus \mathcal{Z}) \times \mathrm{SEQ}_{\geq 1} \mathcal{Z}\big).$$

Indeed, the root is either a leaf ($\mathcal{Z}$), or it is an internal node to which is attached a sequence of at least 2 leaves ($\mathcal{U} \times \mathcal{Z} \times \Theta_{\mathcal{Z}}(\mathcal{Z}) \times \mathrm{SEQ}_{\geq 1}(\mathcal{U} \times \mathcal{Z})$), or the tree is larger, i.e. the last step in the evolution process was replacing another leaf by an internal node to which is attached a sequence of non-marked leaves ($\Theta_{\mathcal{Z}}(\mathcal{T} \setminus \mathcal{Z}) \times \mathrm{SEQ}_{\geq 1} \mathcal{Z}$). Using the symbolic method, we thus get that

$$T(z, u) = z + \frac{u^2 z^2}{1 - uz} + \frac{z^2}{1 - z}\, \partial_z \left(T(z, u) - z\right).$$

In the same way as before, through a direct extraction $[z^n](1 - zu)(1 - z)T(z, u)$, we prove that $t_1(u) = 1$, $t_2(u) = u^2$, and for all $n > 2$,

$$t_n(u) = (u - 1)\, u^{n-1} + n\, t_{n-1}(u).$$

This implies $t_{1,0} = 1$, $t_{n,n} = 1$ for all $n \geq 1$, $t_{n,k} = n\, t_{n-1,k}$ for all $1 \leq k \leq n - 1$, and $t_{n,k} = 0$ for all $k > n$, which concludes the proof. □

The proof of Theorem 2.3.4 is a little more involved.

Proof of Theorem 2.3.4. The operators needed for the specification are not so classical so we prefer to directly write the differential equation satisfied by $T(z, u) = \sum_{n,k} t_{n,k}\, u^k z^n$, where $t_{n,k}$ is the number of size-n increasing Schröder trees with k leaves attached to the root. Like in the proof of Theorem 2.3.5 at each step we must remove the tree reduced to the leaf, i.e. $\mathcal{T} \setminus \mathcal{Z}$. So let us introduce the function $V(z, y) = T(z, y) - z$. Thus we get

$$T(z, u) = z + \frac{u^2 z^2}{1 - uz} + \frac{z^2}{1 - z}\, \frac{\partial_u V(z, u)}{z} + \frac{z^2}{1 - z} \left(\partial_z V(z, u) - \frac{u}{z}\, \partial_u V(z, u)\right). \quad (9)$$

Indeed, by looking at the last step in the evolution process, four possibilities occur:
- the tree is reduced to a leaf z, i.e. the evolution process did not already start
- the tree contains a single internal node to which a sequence of at least 2 leaves is attached ($\frac{u^2\, z^2}{1 - uz}$), i.e. the evolution process has gone through one step only,
- in the last step of the evolution process, a leaf of the root has been replaced by an internal node to which a sequence of leaves is attached ($\frac{z^2}{1-z}\, \frac{\partial_u V(z,u)}{z}$), in fact, leaves attached to the root are marked as zu, the differentiation by u followed by

the division by z gives the result,
- in the last step of the evolution process, a leaf that is not attached to the root has been selected and replaced by an internal node attached to at least two leaves:

$$\frac{z^2}{1-z}\left(\partial_z V(z,u) - \frac{u}{z}\,\partial_u V(z,u)\right).$$

The second term removes the trees built in the first one where a leaf attached to the root has been selected. As an example, take a tree counted by $z^{\ell+r}u^r$, thus containing ℓ leaves such that r of them are attached to the root. The operation gives $(\ell+r)z^{\ell+r-1}u^r - \frac{u}{z}\, r\, z^{\ell+r}u^{r-1}$ and thus gives exactly $\ell z^{\ell+r-1}u^r$.

After some simplifications and multiplications by $(1-uz)(1-z)$ we get

$$(1-uz)(1-z)\,V(z,u) = u^2z^2(1-z) + z^2(1-uz)\left(\partial_z V(z,u) - \frac{u}{z}\,\partial_u V(z,u)\right).$$

By extracting the coefficient of z^n from the latter equation, we directly get that, for all $n \geq 4$,

$$v_n(u) = (n+u)\,v_{n-1}(u) - u(n-1)\,v_{n-2}(u) - (u-1)\,v'_{n-1}(u) + u(u-1)\,t'_{n-2}(u),$$

and $v_1(u) = 0$, $v_2(u) = u^2$ and $v_3(u) = 2u^2 + u^3$.

To evaluate the average number of leaves attached to the root we must compute the limit of the ratio $v'_n(u)/v_n(u)$ when n tends to infinity and evaluate it for $u = 1$. Differentiating the last equation we get

$$\begin{aligned} v'_n(u) \quad &= \quad v_{n-1}(u) + (n+u-1)\,v'_{n-1}(u) - (u-1)\,v''_{n-1}(u) \\ &\quad -(n-1)\,v_{n-2}(u) + (2u-(n-1)u-1)\,v'_{n-2}(u) + u(u-1)\,v''_{n-2}(u) \end{aligned} \tag{10}$$

We thus define the sequence of the average values $m_n = v'_n(1)/v_n(1)$ and get for $n \geq 4$

$$m_n = m_{n-1} - \frac{n-2}{n(n-1)}m_{n-2},$$

with $m_1 = 0, m_2 = 2$ and $m_3 = 5/3$. In order to analyse the sequence of real values m_n we introduce an alternative sequence ℓ_n such that $\ell_n = n\ m_n$ and thus we obtain for all $n \geq 1$, we get, for all $n \geq 4$,

$$\ell_n = \left(1 + \frac{1}{n-1}\right)\ell_{n-1} - \frac{1}{n-1}\ell_{n-2}, \tag{11}$$

and $\ell_1 = 0, \ell_2 = 4$ and $\ell_3 = 5$. Finally, we set $e_n = 2\sum_{i=0}^{n-1} 1/i!$ for all $n \geq 1$, and prove by induction that, for all $n \geq 1$, $\ell_n = e_n$. First note that $\ell_n = e_n$ for $n = \{1,2,3\}$. Now take $n \geq 4$ and assume that for all $i < n$ we have $\ell_i = e_i$. Using the fact that $e_n = e_{n-1} + 2/(n-1)!$, and Equation (11), we have

$$\begin{aligned} \ell_n - e_n &= \ell_{n-1} + \frac{1}{n-1}\,(\ell_{n-1} - \ell_{n-2}) - e_n \\ &= \ell_{n-1} - e_{n-1} + \frac{1}{n-1}\left(\ell_{n-1} - \ell_{n-2} - \frac{2}{(n-2)!}\right) \\ &= \ell_{n-1} - e_{n-1} + \frac{1}{n-1}\,(\ell_{n-1} - e_{n-1} - (\ell_{n-2} - e_{n-2})) = 0, \end{aligned}$$

and thus $\ell_n = e_n$, which concludes the induction argument. Since, by definition of e_n, $e_n = 2\mathrm{e} + \Theta(1/n!)$, and since $e_n = \ell_n = nm_n = n\mathbb{E}[L_n]$, we get

$$m_n = \mathbb{E}[L_n] = \frac{2}{n}\sum_{i=0}^{n-1}\frac{1}{i!} \underset{n\to\infty}{=} \frac{2\mathrm{e}}{n} + \Theta\left(1/(n\cdot n!)\right) \tag{12}$$

We now estimate the variance of L_n; to do so, we use the following identity (see, e.g., [FS09, p. 159]):

$$\mathbb{V}[L_n] = \mathbb{E}[L_n(L_n-1)] + \mathbb{E}[L_n] - (\mathbb{E}[L_n])^2 \tag{13}$$

Since we already have estimated $\mathbb{E}L_n$, we only need to estimate $\mathbb{E}[L_n(L_n-1)] = v_n''(1)/v_n(1)$, which we denote by k_n. Differentiating Equation (10) we get that, for all $n \geq 4$,

$$k_n = k_{n-1} - \frac{1}{n}\left(k_{n-1} - \frac{n-3}{n-1}\,k_{n-2}\right) + \frac{2}{n}\left(m_{n-1} - \frac{n-2}{n-1}\,m_{n-2}\right),$$

where we recall that $m_n = v_n'(1)/v_n(1) = \mathbb{E}[L_n]$. The first terms of $(k_n)_{n\geq 1}$ are $k_1 = 0, k_2 = 2$ and $k_3 = 2$. For all $n \geq 1$, set $r_n = n(n-1)k_n$. Using Equation (12), we get that, for all $n \geq 4$,

$$r_n = r_{n-1} + \frac{1}{n-2}\left(r_{n-1} - r_{n-2}\right) + \frac{4}{(n-2)!},$$

with the initial values $r_1 = 0, r_2 = 4$ and $r_3 = 12$. Finally, for all $n \geq 3$, we set

$$\tilde{e}_n = 4\sum_{i=2}^{n-1}\frac{i-1}{(i-2)!} = 4\sum_{i=3}^{n-1}\frac{1}{(i-3)!} + 4\sum_{i=2}^{n-1}\frac{1}{(i-2)!},$$

and $\tilde{e}_1 = 0$, $\tilde{e}_2 = 4$. By induction, one can prove that, for all $n \geq 1$, $r_n = \tilde{e}_n$, which implies

$$\begin{aligned} k_n = \mathbb{E}[L_n(L_n-1)] = \frac{r_n}{n(n-1)} &= \frac{4}{n(n-1)}\sum_{i=3}^{n-1}\frac{1}{(i-3)!} + \frac{4}{n(n-1)}\sum_{i=2}^{n-1}\frac{1}{(i-2)!} \\ &= \frac{8\mathrm{e}}{n^2} + \Theta\left(1/n^3\right). \end{aligned}$$

Using this last estimate together with equations (13) and (12), we get

$$\mathbb{V}L_n \underset{n\to\infty}{=} \frac{2\mathrm{e}}{n} + \Theta\left(1/n^2\right). \qquad \square$$

2.3.3. *Quantitative analysis of the number of nodes of a given arity.* In this section, we prove asymptotic results for the number of nodes of a given arity in a typical increasing Schröder tree, starting with binary nodes:

Theorem 2.3.6. *Let B_n be the number of binary nodes (nodes of arity 2) in a tree taken uniformly at random among all increasing Schröder trees of size n. Asymptotically when n tends to infinity, we have*

$$\mathbb{E}[B_n] = n - 2\ln n + 2\gamma - \frac{7}{3} + \mathcal{O}(1/n), \qquad \textit{and } \mathbb{V}[B_n] = 4\ln n + 4\gamma - \frac{2}{3}\pi^2 - \frac{17}{6} + \mathcal{O}(1/n),$$

where γ is the Euler-Mascheroni constant. Moreover, in distribution when $n \to +\infty$,

$$\frac{B_n - (n - 2\ln n)}{2\sqrt{\ln n}} \to \mathcal{N}(0,1).$$

In other words, almost all internal nodes are binary, only a proportion of order $2\ln n/n$ of internal nodes are at least ternary. In the following theorem, we show that, on average, half of all non-binary nodes are ternary.

Theorem 2.3.7. *Let $C_n^{(\ell)}$ be the number of nodes of arity $\ell \geq 3$ in a tree taken uniformly at random among all increasing Schröder trees of size n. Asymptotically when n tends to infinity, we have*

$$\mathbb{E}C_n^{(3)} = \ln n + \mathcal{O}(1), \qquad \text{and } \mathbb{E}C_n^{(4)} \sim c_\ell,$$

for some positive constants $(c_\ell)_{\ell\geq 4}$; and, for $\ell = 4$, we have $c_4 = {}^{23}/_{90}$.

Proof of Theorem 2.3.6. Here the specification is easy to exhibit, and its translation via the symbolic method is direct (in this proof, $\mathcal{U}$ marks the binary nodes):

$$\mathcal{T} = \mathcal{Z} \cup \left(\Theta_{\mathcal{Z}}\mathcal{T} \times \left(\mathcal{U} \times \mathcal{Z} \cup \textsc{Seq}_{\geq 2}\,\mathcal{Z}\right)\right);$$

$$T(z,u) = z + \left(u\,z^2 + \frac{z^3}{1-z}\right)\partial_z T(z,u).$$

The method we use to analyse this differential equation is similar to [CHY00]. For all $n \geq 3$,

$$t_n(u) = (1 + u(n-1))t_{n-1}(u) + (1-u)(n-2)t_{n-2}(u), \tag{14}$$

with $t_1(u) = 1$, $t_2(u) = u$, and $t_3(u) = 1 + u^2$. Once again (see also Lemma 2.3.2) we take the row-reversed generating function $\hat{t}_n(u) = \sum_{k=1}^n t_{n,k}\, u^{n-k} = t^n t_n({}^1/_u)$. From Equation (14), we get that, for all $n \geq 4$,

$$\hat{t}_n(u) = \frac{n+u-1}{n}\,\hat{t}_{n-1}(u) + \frac{u(u-1)(n-2)}{n(n-1)}\,\hat{t}_{n-2}(u), \tag{15}$$

with $\hat{t}_2(u) = u$ and $\hat{t}_3(u) = (2u+u^3)/3$. Let us now define $F(z,u) = \sum_{n\geq 2} n\hat{t}_n(u)z^n$; this generating function satisfies the following differential equation:

$$z\,(1-z)\,\partial_z F(z,u) = \left(1 + uz - u(1-u)z^2\right)\,F(z,u) + 2uz^2\,(1 - u(1-u)z)\,,$$

with initial condition $\partial_z^2 F(z,u)|_{z=0} = 4\,u$. This last equation gives

$$F(z,u) = 2uz\exp\left(u(1-u)z\right)(1-z)^{-1-u^2}\int_0^z (1-u(1-u)t)\,(1-t)^{u^2}\exp\left(-u(1-u)t\right)\,\mathrm{d}t.$$

Let $\phi(z,u) = (1-u(1-u)z)\,(1-z)^{u^2}\mathrm{e}^{-u(1-u)z}$; with this definition, we get

$$F(z,u) = (1-z)^{-1-u^2}\left(g(u) + E(z,u)\right),$$

where,

$$g(u) = 2u\mathrm{e}^{u(1-u)}\,(1-z)^{-1-u^2}\int_0^1 \phi(t,u)\,\mathrm{d}t$$

and,

$$E(z,u) = \left(z\,\mathrm{e}^{u(1-u)z}\right) - \mathrm{e}^{u(1-u)}\int_0^1 \phi(t,u)\,dt - z\mathrm{e}^{u(1-u)z}\int_z^1 \phi(t,u)\,\mathrm{d}t.$$

Therefore, asymptotically when $n \to +\infty$,

$$n\hat{t}_n(u) = \frac{g(u)}{\Gamma\left(1+u^2\right)}\,n^{u^2}\left(1 + \mathcal{O}({}^1/_n)\right)$$

uniformly for all u such that $|u-1| \leq \delta$ for some $\delta > 0$. This thus falls into the scope of the quasi-powers framework and Theorem IX.8 [FS09, p. 645] is applicable with

$B(u) = \exp(2u)$ and $\beta_n = \ln n$, which concludes the proof (the mean and variance expansions can be calculated automatically using, e.g., a computer software such as Maple). $\square$

Proof of Theorem 2.3.7. We first look at ternary nodes: the specification (with $\mathcal{U}$ marking ternary nodes) is given by

$$\mathcal{T} = \mathcal{Z} \cup \left(\Theta_{\mathcal{Z}}\mathcal{T} \times \left(\mathcal{Z} \cup \mathcal{U} \times \mathcal{Z}^2 \cup \mathrm{SEQ}_{\geq 3}\,\mathcal{Z}\right)\right)$$

which implies

$$T(z,u) = z + \left(z^2 + u\, z^3 + \frac{z^4}{1-z}\right)\partial_z T(z,u),$$

and thus, for all $n \geq 4$:

$$t_n(u) = nt_{n-1}(u) + (u-1)(n-2)t_{n-2}(u) + (n-3)(1-u)t_{n-3}(u),$$

with $t_1(u) = 1$ and $t_2(u) = 1$. Differentiating this equation, we get that, for all $n \geq 5$,

$$t_n'(u)|_{u=1} = nt_{n-1}'(u)|_{u=1} + \frac{(n-2)(n-2)!}{2} - \frac{(n-3)(n-3)!}{2}.$$

This thus implies that, for all $n \geq 5$,

$$\mathbb{E}[C_n^{(3)}] = \mathbb{E}[C_{n-1}^{(3)}] + \frac{(n-2)}{n(n-1)} - \frac{(n-3)}{n(n-1)(n-2)} = \frac{10}{24} + \sum_{\ell=5}^{n}\left(\frac{(k-2)}{k(k-1)} - \frac{(k-3)}{k(k-1)(k-2)}\right).$$

since $\mathbb{E}[C_4^{(3)}] = 10/24$. Using again the fact that $\sum_{k=1}^{n}\frac{1}{k} = \ln n + \mathcal{O}(1)$ and $\sum_{k=1}^{n}\frac{1}{k^2} = \mathcal{O}(1)$ when n tends to infinity, we get

$$\mathbb{E}[C_n^{(3)}] = \ln n + \mathcal{O}(1),$$

as claimed.

We reason similarly for $\ell = 4$ ($\mathcal{U}$ now marks nodes of arity 4):

$$\mathcal{T} = \mathcal{Z} \cup \left(\Theta_{\mathcal{Z}}\mathcal{T} \times \left(\mathcal{Z} \cup \mathcal{Z}^2 \cup \mathcal{U} \times \mathcal{Z}^3 \cup \mathrm{SEQ}_{\geq 4}\,\mathcal{Z}\right)\right);$$

$$T(z,u) = z + \left(z^2 + z^3 + u\, z^4 + \frac{z^5}{1-z}\right)\partial_z T(z,u).$$

Thus, for all $n \geq 4$, we have

$$t_n(u) = nt_{n-1}(u) + (u-1)(n-3)t_{n-3}(u) + (n-4)(1-u)t_{n-4}(u),$$

with $t_1(u) = 1$ and $t_2(u) = 1$, which, after differentiating at $u = 1$ and dividing by t_n gives

$$\mathbb{E}[C_n^{(4)}] = \mathbb{E}[C_{n-1}^{(4)}] + \frac{(n-3)}{n(n-1)(n-2)} - \frac{(n-4)}{n(n-1)(n-2)(n-3)},$$

with $\mathbb{E}[C_5^{(4)}] = 12/120$. A simple look to this recurrence shows that it converges to a constant since it is a modified geometric sum. Solving the recurrence we obtain,

$$\mathbb{E}[C_n^{(4)}] = \frac{23}{90} - \frac{13}{6n} - \frac{1}{6\,(n-2)} + \frac{4}{3\,(n-1)},$$

which proves the statement for $\ell = 4$.

Let us now treat the general $\ell \geq 5$ case ($\mathcal{U}$ now marks nodes of arity ℓ):

$$\mathcal{T} = \mathcal{Z} \cup \left(\Theta_{\mathcal{Z}}\mathcal{T} \times \left(\left(\cup_{i=1}^{\ell-2}\mathcal{Z}^{\ell}\right) \cup \mathcal{U} \times \mathcal{Z}^{\ell-1} \cup \mathrm{SEQ}_{\geq 4}\,\mathcal{Z}\right)\right);$$

$$T(z,u) = z + \left(\left(\sum_{i=2}^{\ell-1} z^i \right) + u\, z^\ell + \frac{z^{\ell+1}}{1-z} \right) \partial_z T(z,u).$$

This implies that, for all $n \geq \ell$:

$$t_n(u) = n t_{n-1}(u) + (u-1)(n-\ell+1) t_{n-\ell+1}(u) + (n-\ell)(1-u) t_{n-\ell}(u),$$

with $t_n(u) = 1$ for all $n < \ell$. Therefore, we get

$$\begin{aligned}\mathbb{E}[C_n^{(\ell)}] &= \mathbb{E}[C_{n-1}^{(\ell)}] + \frac{(n-\ell+1)}{n(n-1)\cdots(n-\ell+2)} - \frac{(n-\ell)}{n(n-1)\cdots(n-\ell+1)} \\ &= \mathbb{E}[C_\ell^{(\ell)}] + \sum_{k=\ell+1}^{n} \left(\frac{(k-\ell+1)}{k(k-1)\cdots(k-\ell+2)} - \frac{(k-\ell)}{k(k-1)\cdots(k-\ell+1)} \right).\end{aligned}$$

Since

$$\left(\frac{(k-\ell+1)}{k(k-1)\cdots(k-\ell+2)} - \frac{(k-\ell)}{k(k-1)\cdots(k-\ell+1)} \right) \underset{k\to\infty}{\sim} \frac{1}{k^{\ell-2}},$$

which implies that, for all $\ell \geq 4$,

$$\lim_{n\to\infty} \mathbb{E}[C_n^{(\ell)}] = \mathbb{E}[C_\ell^{(\ell)}] + \sum_{k=\ell+1}^{\infty} \left(\frac{(k-\ell+1)}{k(k-1)\cdots(k-\ell+2)} - \frac{(k-\ell)}{k(k-1)\cdots(k-\ell+1)} \right) < +\infty.$$

All these recurrences converges to constants that get smaller and smaller when ℓ increases. $\square$

Note that the constants c_ℓ are computable by solving the simple recurrences for each case; Table 3 gives a summary of the typical number of nodes for the smallest arities.

	2-ary	3-ary	4-ary	5-ary	6-ary	7-ary	8-ary	9-ary	10-ary
$\mathbb{E}C_n^{(\ell)}$	$n - 2\ln n$	$\ln n$	$\frac{23}{90}$	$\frac{1}{32}$	$\frac{107}{25200}$	$\frac{47}{86400}$	$\frac{101}{1587600}$	$\frac{229}{33868800}$	$\frac{659}{1005903360}$

FIGURE 3. The asymptotic number of ℓ-ary nodes

2.3.4. *Typical depth of the leftmost leaf.* In this section, we prove a central limit theorem for the depth of the leftmost leaf in a typical increasing Schröder tree; this gives us a lower bound for the height of a typical increasing Schröder tree:

Lemma 2.3.8. *Let Y_n be the depth of the leftmost leaf in a tree taken uniformly at random among all increasing Schröder trees of size n. For all $n \geq 1$, $Y_n = n - X_n$, where X_n is the number of internal nodes in a typical increasing Schröder tree of size n (see Theorem 2.3.1), and thus, we have convergence in distribution when n tends to infinity:*

$$\frac{Y_n - \ln n}{\sqrt{\ln n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Note that the choice of the leftmost leaf is arbitrary, although it has the advantage that the specification is straightforward. Table 4 exhibits the smallest values of $(t_{n,k})$.

Proof. We directly look at the differential equation satisfied by $T(z,u)$, where u marks the internal nodes that belong to the leftmost path (between the root and the leftmost leaf).

$$T(z,u) = z + \partial_z\left(\frac{T(z,u)}{z}\right)\frac{z^3}{1-z} + T(z,u)\frac{uz}{1-z}.$$

Indeed, the tree is either a unique leaf (which is thus also the leftmost leaf) at height zero (z), or at the last step of the evolution process, we have selected a leaf that is not the leftmost one and replaced it by a sequence of at least two leaves ($\partial_z(T(z,u)/z)\frac{z^3}{1-z}$), or we have replaced the leftmost leaf by an internal node and a sequence of at least two leaves ($T(z,u)\frac{uz}{1-z}$). We rewrite this equation as

$$(1-uz)T(z,u) = z(1-z) + z^2\partial_z T(z,u),$$

and thus, identifying the coefficient of z^n on both sides gives that

$$t_n(u) = (u+n-1)t_{n-1}(u) \quad (\forall n \geq 3),$$

$t_1(u) = 1$, and $t_2(u) = u$. This implies that, for all $n \geq 3$,

$$t_n(u) = u\prod_{i=2}^{n-1}(u+i) = \frac{1}{1+u}SC_n(u),$$

where $SC_n(u)$, defined in Lemma 2.3.2, is the generating function of all size n permutations with k cycles. Therefore, using Lemma 2.3.2, we get that $Y_n = n - X_n$ in distribution, where X_n is the number of internal nodes in a typical increasing Schröder tree, which concludes the proof, by Theorem 2.3.1. □

1						
0 ,	1					
0 ,	2,	1				
0 ,	6,	5,	1			
0 ,	24,	26,	9,	1		
0 ,	120,	154,	71,	14,	1	
0 ,	720,	1044,	580,	155,	20,	1

TABLE 4. The values of $t_{n,k}$, the number of increasing Schröder trees of size n trees whose leftmost leaf has depth k, for all $0 \leq k < n \in \{1, 2, \ldots, 7\}$.

2.4. **Bijection with permutations.** The fact that the number of increasing Schröder trees of size n is equal to $t_n = n!/2$ hints at the existence of a relationship between our model of increasing trees and a subclass of permutations. In this section, we aim at exhibiting this relationship.

We denote by $\sigma = (\sigma_1, \ldots, \sigma_n)$ the size-n permutation that sends i to $\sigma_i \in \{1, \ldots, n\}$ for all $i \in \{1, \ldots, n\}$. For all $k \in \{1, \ldots, n\}$, we denote by $\sigma^{-1}(k)$ the pre-image of k by σ, and sometimes call $\sigma^{-1}(k)$ the "position" of k in the permutation σ.

We now define recursively a map $\mathcal{M}$ between $\mathcal{HP}$, the class of permutations such that 1 appears before 2, and the class $\mathcal{T}$ of increasing Schröder trees.

The only element of $\mathcal{HP}$ of size 2 is the permutation $(1,2)$; we set its image to be the tree whose root is labelled by 1 and has two (unlabelled) leaf-children. Now assume that we have defined $\mathcal{M}(\sigma)$ for all permutations $\sigma \in \mathcal{HP}$ of size at most $n-1$ for some $n \geq 2$ and let σ be a size-n permutation in $\mathcal{HP}$. We distinguish two cases according to the pre-image of n by σ; we denote by $\hat{\sigma}_i = \sigma_i$ if $\sigma_i < \sigma_n$ and $\hat{\sigma}_i = \sigma_i - 1$ otherwise. For example, if $\sigma = (4,1,5,2,3)$, then $\hat{\sigma} = (3,1,4,2)$; $\hat{\sigma}$ can be seen as the permutation induced by σ on $\{1,\dots,n-1\}$.

- If $\sigma_n = n$ then, we set $\mathcal{M}(\sigma)$ to be the tree $\mathcal{M}(\hat{\sigma})$ in which we add a new rightmost leaf to the internal node with the largest label.
- If $\sigma_n = k < n$, then, we build $\mathcal{M}(\sigma)$ as follows: create a new binary node ν labelled with the smallest integer that does not appear as a label in $\mathcal{M}(\hat{\sigma})$ and attach two new leaves to this internal node. Insert this tree in $\mathcal{M}(\hat{\sigma})$ by placing ν in the k-th leaf (we assume, for example, that the leaves are ordered in the depth-first order) of $\mathcal{M}(\hat{\sigma})$.

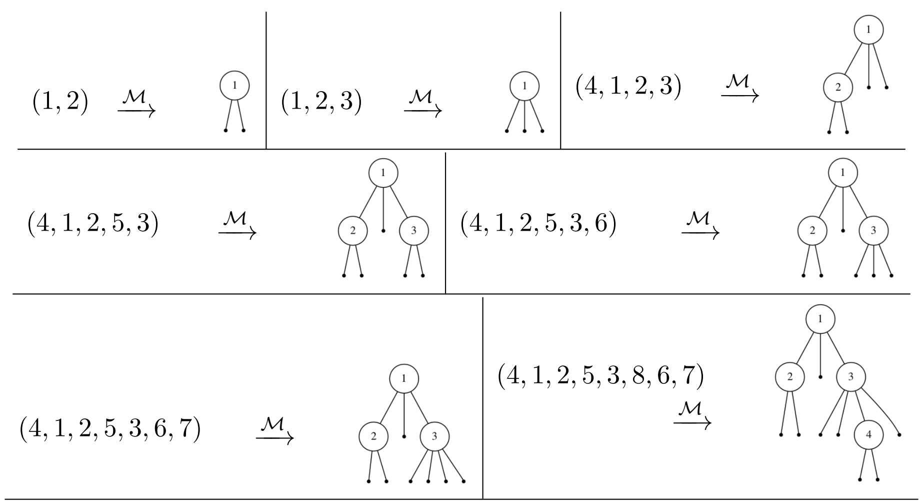


FIGURE 4. A size-8 example of the mapping $\mathcal{M}$

In Figure 4 we present the mapping on an example. Remark that we have ordered the steps reversely to understand the process in a constructive way.

Theorem 2.4.1. *The map $\mathcal{M}$ is a one-to-one correspondence between $\mathcal{HP}$ and $\mathcal{T}$.*

Proof. First note that the image by $\mathcal{M}$ of a permutation of size n is a Schröder tree of size n: indeed, at each iteration we remove exactly one element from the permutation and add exactly one leaf to the tree by either adding a leaf to the node with largest label or by removing one leaf and adding two new ones. Since the number of permutations of size n in $\mathcal{HP}$ is equal to the number of Schröder trees of size n, it is enough to prove that $\mathcal{M}$ is injective to conclude the proof. The mapping is injective since by induction at each iteration we remove the greatest element of the permutation and the following actions are performed on the resulting tree in a non-ambiguous manner. This concludes the proof. $\square$

2.5. **Uniform random sampling.** In this section, we present an algorithm that samples a Schröder tree uniformly at random among all Schröder trees of a given size. Our aim is to use this algorithm to generate trees of large size (typically several thousands of leaves): we thus provide a detailed analysis of the complexity of our sampler.

Note that the uniform sampling of structures with increasing labelling constraints is not so classical in the context of analytic combinatorics. Martínez and Molinero [MM03, Mol05] focus on the recursive method: using and generalising recursive and unranking generation methods, they give a method that, given a combinatorial specification, automatically outputs a uniform generation algorithm and its complexity analysis. Using a different approach based on Boltzmann generation, Bodini, Roussel and Soria [BRS12] give an algorithmic framework to develop Boltzmann samplers in the context of specifications that lead to differential equation of the first order. The paper [BDF$^+$16] show that this framework can be extended to the context of differential equations of higher order; in particular, they apply this method to the generation of diamonds satisfying differential equations of order 2.

The bijection presented in Section 2.4 immediately gives an algorithm that samples a tree uniformly among all Schröder trees of size n: first sample a permutation uniformly at random among all permutations of size n in $\mathcal{HP}$, and then build its image by $\mathcal{M}$. While there exists fast algorithms to sample permutations (see for example [BBHT17]), it is not clear how to make the application of $\mathcal{M}$ efficient.

Instead, we use the bijection $\mathcal{M}$ as a basis for a direct *probabilistic construction.* Indeed, one can sample a uniform bijection uniformly at random in $\mathcal{H}P$ by doing the following recursive procedure: if $n = 2$, then return $\sigma^{(2)} = (2, 1)$. If $n \geq 3$, assume we have sampled $\sigma^{(n-1)}$ uniformly among all permutations of size $n - 1$. Draw an integer k_n uniformly at random in $\{1, \dots, n\}$, and set $\sigma_n^{(n)} = k_n$, and

$$\sigma_i^{(n)} = \begin{cases} \sigma_i^{(n-1)} & \text{if } \sigma_i^{(n-1)} < k_n \\ \sigma_i^{(n-1)} + 1 & \text{otherwise.} \end{cases}$$

One can indeed check that $\sigma^{(n)}$ is uniformly distributed among all permutations of size n in $\mathcal{HP}$. Executing this random sampling of $\sigma^{(n)}$ simultaneously with $\mathcal{M}$ (note that, for all $n \geq 3$, $\sigma^{(n-1)} = \hat{\sigma}^{(n)}$, where the notation $\hat{\sigma}$ is defined in the definition of $\mathcal{M}$) is the idea of our sampler:

Algorithm 1 Increasing Schröder Tree Builder

```
function TreeBuilder(n)
    if n = 1 then
        return the single leaf
    T = the root labelled by 1 and attached to two leaves
    ℓ = 2
    for i from 3 to n do
        k = rand_int(1, i)
        if k = i then
            Add a new leaf to the last added internal node in T
        else
            Create a new binary node at position k in T
                    with label ℓ and attached to two leaves
            ℓ = ℓ + 1
    return T
```

The function $rand\_int(a, b)$ returns uniformly at random an integer in $\{a, a+1, \dots, b\}$.

Using the adequate data structures, as for example by keeping an array of pointers to all leaves and another one to the last inserted internal node, each insertion in the tree under construction is done in constant time. We thus get

Theorem 2.5.1. *The function* TreeBuilder(n) *in Algorithm 1 is a uniform sampling algorithm for size-n trees. Asymptotically, it operates in $\mathcal{O}(n)$ operations on trees and necessitates $\mathcal{O}(n \ln n)$ random bits.*

2.6. **Analysis of the height of a typical increasing Schröder tree.** The probabilistic construction used in our uniform sampler allows us to prove the following result.

Theorem 2.6.1. *For all $n \geq 2$, let H_n be the height of a tree taken uniformly at random among all increasing Schröder trees of size n. Asymptotically when n tends to infinity,*

$$\mathbb{P}\left(\frac{H_n}{\ln n} \in [1-\varepsilon, \gamma+\varepsilon]\right) \to 1,$$

where $\gamma = \inf\{c > 0 \colon c - 1 + c\ln(2/c) < 0\} \approx 4.311$. This implies, in particular that $\mathbb{E}[H_n] = \Theta(\ln n)$ when n tends to infinity.

Definition 2.6.2. *Given a sequence of integers $\boldsymbol{d} = (d_i)_{i\geq 1}$, we define the random $\boldsymbol{d}$-ary tree $(\tau_n^{(\boldsymbol{d})})_{n\geq 0}$ recursively as follows: $\tau_0^{(\boldsymbol{d})}$ is reduced to its root, given $\tau_{\ell-1}^{(\boldsymbol{d})}$, we build $\tau_\ell^{(\boldsymbol{d})}$ as the tree obtained by picking a leaf uniformly at random in $\tau_{\ell-1}^{(\boldsymbol{d})}$ and replacing it by a node to which d_ℓ leaves are attached.*

Lemma 2.6.3. *Let $\boldsymbol{D} = (D_\ell)_{\ell\geq 1}$ be the sequence of integer-valued random variables defined by:*

- *$\mathbb{P}(D_1 = k) = 2k/(k+1)!$ for all $k \geq 2$, and*
- *if, for all $\ell \geq 1$, we denote by $\bar{D}_\ell = \sum_{i=1}^{\ell} D_i$, then,*

$$\mathbb{P}(D_{\ell+1} = k | D_1, \ldots, D_\ell) = \frac{(\bar{D}_\ell + 1)!(k - 1 + \bar{D}_\ell)}{(k + \bar{D}_\ell)!}.$$

Then, for all $\ell \geq 1$, the tree $\tau_\ell^{(\boldsymbol{D})}$ given its size is equal in distribution to an increasing Schröder tree taken uniformly at random among all trees of that size.

Proof. This follows from Theorem 2.5.1. Indeed, note that the degree of the last inserted internal node increases as long as the random integer $k = k_i$ (see line 7 of Algorithm 1) drawn in the i-th loop is not equal to i. Note that this happens with probability $1/i$. For example, the degree of the root starts at 2, we draw the first integer $k_3 \in \{1, 2, 3\}$ and if $k_3 \neq 3$, then we can conclude that $D_1 = 2$, otherwise, we know that $D_1 \geq 3$ and we need to look at k_4. Therefore, $\mathbb{P}(D_1 = 2) = 2/3$, as claimed, and $\mathbb{P}(D_1 \geq 3) = 1/3$. Iterating this argument, we get that

$$\mathbb{P}(D_1 \geq k) = \prod_{i=3}^{k} \mathbb{P}(k_i = i) = \prod_{i=3}^{k} \frac{1}{i} = \frac{2}{k!},$$

and thus

$$\mathbb{P}(D_1 = k) = \mathbb{P}(D_1 \geq k) - \mathbb{P}(D_1 \geq k+1) = \frac{2}{k!} - \frac{2}{(k+1)!} = \frac{2k}{(k+1)!},$$

as claimed.

By definition of our sampling algorithm, we know that the $(\ell+1)$-th internal node is inserted into the tree during the loop number $i = D_1+\dots+D_\ell+1 = \bar{D}_\ell+1$. Therefore, we get

$$\mathbb{P}(D_{\ell+1} = 2|D_1,\dots,D_\ell) = \mathbb{P}(k_{i+1} \neq i+1) = 1 - \frac{1}{\bar{D}_\ell+2}, \text{ as claimed,}$$

and

$$\mathbb{P}(D_{\ell+1} \geq 3|D_1,\dots,D_\ell) = \frac{1}{\bar{D}_\ell+2}.$$

Iterating this argument, we get that, for all $k \geq 3$,

$$\mathbb{P}(D_{\ell+1} \geq k|D_1,\dots,D_\ell) = \prod_{j=\bar{D}_\ell+2}^{\bar{D}_\ell+k-1} \mathbb{P}(k_j = j) = \prod_{j=\bar{D}_\ell+2}^{\bar{D}_\ell+k-1} \frac{1}{j} = \frac{(\bar{D}_\ell+1)!}{(\bar{D}_\ell+k-1)!}.$$

This concludes the proof because

$$\begin{aligned}\mathbb{P}(D_{\ell+1} = k|D_1,\dots,D_\ell) &= \mathbb{P}(D_{\ell+1} \geq k|D_1,\dots,D_\ell) - \mathbb{P}(D_{\ell+1} \geq k+1|D_1,\dots,D_\ell)\\ &= \frac{(\bar{D}_\ell+1)!}{(\bar{D}_\ell+k-1)!} - \frac{(\bar{D}_\ell+1)!}{(\bar{D}_\ell+k)!} = \frac{(\bar{D}_\ell+1)!(\bar{D}_\ell+k-1)}{(k+\bar{D}_\ell)!},\end{aligned}$$

as claimed. □

Proof of Theorem 2.6.1. For this proof, we use the fact that the increasing Schröder tree is equal in distribution to the random $\boldsymbol{D}$-ary tree (see Lemma 2.6.3). For the lower bound, we use Lemma 2.3.8 and the fact that, almost surely for all $\ell \geq 1$, $H_\ell \geq Y_{\bar{D}_\ell+1}$, where we recall that Y_n is the depth of the leftmost leaf in an n-leaf uniform increasing Schröder tree and $\bar{D}_\ell = \sum_{i=1}^{\ell} D_i$. By Lemma 2.3.8, we have that, for all $\varepsilon > 0$,

$$\mathbb{P}(H_n \leq (1-\varepsilon)\ln n) \leq \mathbb{P}(Y_n \leq (1-\varepsilon)\ln n) \leq \mathbb{P}\left(\frac{Y_n - \ln n}{\sqrt{\ln n}} \leq -\varepsilon\sqrt{\ln n}\right) \to 0,$$

when n tends to infinity, which concludes the proof for the lower bound.

The proof for the upper bound is an adaptation of Devroye [Dev90] in

which the case of regular trees is treated (in regular trees, nodes have all the same degree they are also known as random k-ary trees). We denote by $N_1(n),\dots,N_{D_1}(n)$ the sizes of the subtrees of the root of $\tau_n^{(\boldsymbol{D})}$; a straightforward adaptation of [Dev90, Lemma 2] gives that, conditionally on D_1,

$$\mathbb{P}((n-m+2)S_1 \geq x) \leq \mathbb{P}(N_1(n) \geq x) \leq \mathbb{P}(nS_1 \geq x), \tag{16}$$

where S_1 is the minimum of $D_1 - 1$ i.i.d. random variables uniform on $[0,1]$. We reason conditionally on the sequence $\boldsymbol{D}$ of random degrees, and denote by $\mathbb{P}_{\boldsymbol{D}}$ the law under this conditioning. We denote by $S_1,\dots,S_{D_1}$ the spacings induced on $[0,1]$ by a sample of $D_1 - 1$ i.i.d. random variables uniform on $[0,1]$. Using the fact that the sizes of the subtrees of the root, $N_1(n),\dots,N_{D_1}(n)$ all have the same distribution, we get

$$\begin{aligned}\mathbb{P}_{\boldsymbol{D}}(H_n \geq k) &\leq \sum_{i=1}^{D_1} \mathbb{P}_{\boldsymbol{D}}(H_{N_i(n)} \geq k-1) = D_1\mathbb{P}_{\boldsymbol{D}}(H_{N_1(n)} \geq k-1)\\ &\leq D_1\mathbb{P}_{\boldsymbol{D}}(H_{nS_1} \geq k-1),\end{aligned}$$

where we have used Equation (16) in the last inequality. We now iterate this identity: we denote by $I(n) = n\prod_{i=1}^{k} S(D_i)$, where, for all $d \geq 2$, $S(d)$ is the minimum of $d-1$ i.i.d. random variables uniform on $[0,1]$. We get

$$\mathbb{P}_{\boldsymbol{D}}(H_n \geq k) \leq \left(\prod_{i=1}^{k} D_i\right) \mathbb{P}_{\boldsymbol{D}}\left(H_{I(n)} \geq 0\right) = \left(\prod_{i=1}^{k} D_i\right) \mathbb{P}_{\boldsymbol{D}}\left(n\prod_{i=1}^{k} S(D_i) \geq 1\right),$$

because a tree has height at least 1 as soon as it has at least one internal node. We now use Chebychev's inequality, which implies that, for all $\alpha \geq 1$,

$$\mathbb{P}_{\boldsymbol{D}}(H_n \geq k) \leq \left(\prod_{i=1}^{k} D_i\right) n^{\alpha} \mathbb{E}_{\boldsymbol{D}}\left[\prod_{i=1}^{k} S(D_i)^{\alpha}\right] = n^{\alpha} \prod_{i=1}^{k} \left(\frac{\Gamma(D_i+1)}{\prod_{i=1}^{D_i-1}(\alpha+i)}\right).$$

See [Dev90, Equation (1)] for the last equality. For all $\alpha \geq 1$, and for all $d \geq 2$, we have

$$\begin{aligned}\ln\Gamma(d+2) - \sum_{i=1}^{d} \ln(\alpha+i) &= \ln\Gamma(d+1) - \sum_{i=1}^{d-1} \ln(\alpha+i) + \ln(d+1) - \ln(\alpha+d)\\ &\leq \ln\Gamma(d+1) - \sum_{i=1}^{d-1} \ln(\alpha+i).\end{aligned}$$

Therefore, since $D_i \geq 2$ almost surely for all $i \geq 1$, we get

$$\mathbb{P}_{\boldsymbol{D}}(H_n \geq k) \leq n^{\alpha} \prod_{i=1}^{k} \left(\frac{\Gamma(3)}{\alpha+1}\right) = n^{\alpha}\left(\frac{2}{\alpha+1}\right)^{k}.$$

This expression is minimized for $\alpha = k/\ln n - 1$; taking $k = c\ln n$ and $\alpha = c-1$, we get that, for all $c > 0$,

$$\mathbb{P}_{\boldsymbol{D}}(H_n \geq c\ln n) \leq n^{c-1+c\ln(2/c)}.$$

If we take $c > \gamma$ where $\gamma = \inf\{c > 0 \colon c - 1 + c\ln(2/c) < 0\}$, then

$$\mathbb{P}_{\boldsymbol{D}}(H_n \geq c\ln n) \underset{n\to\infty}{\to} 0,$$

which concludes the proof for the upper bound. □

3. Strict monotonic Schröder trees

3.1. **The model and its context.** In this section we introduce and study a generalisation of the increasing Schröder trees, which we call *strict monotonic Schröder trees*. The main difference between the two models is that in strict monotonic Schröder trees, several internal nodes can be labelled by the same integer as long as they are not on the same ancestral line:

Definition 3.1.1. *A* strict monotonic Schröder tree *is a classical Schröder tree structure whose internal nodes are labelled by the integers between* 1 *and* ℓ *(for some* $\ell \geq 1$*), in such a way that each integer in* $\{1, \ldots, \ell\}$ *appears at least once in the tree and the sequence of labels in the path from the root to any leaf is (strictly) increasing.*

Remark that the trees are qualified by "strict" in the sense that the sequence of labels along the paths from the root to any leaf is strictly increasing.

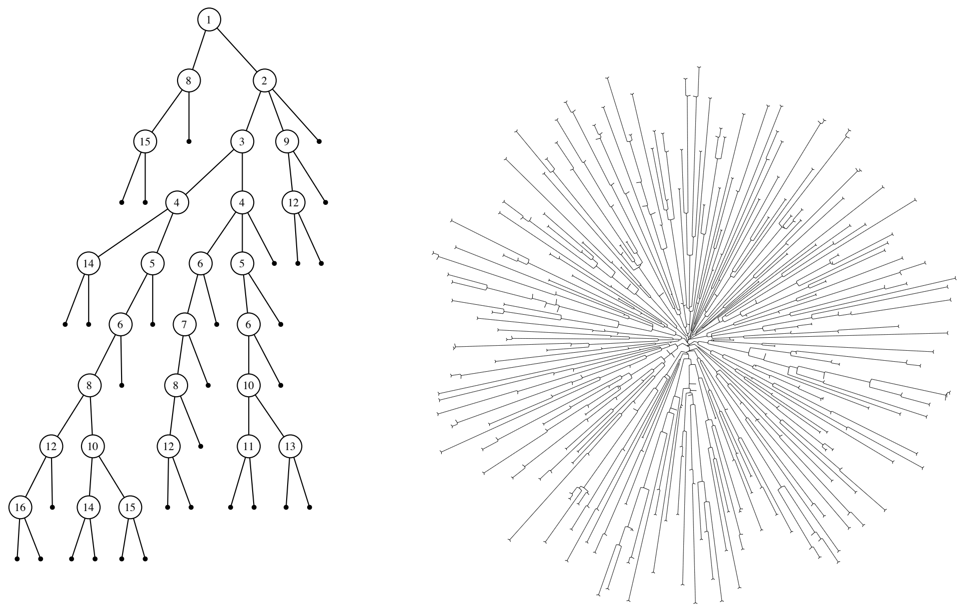


FIGURE 5. Two strict monotonic Schröder trees

In Figure 5 we show two strict monotonic trees: the left-hand-side one is of size 30 with 16 distinct labels, the right-hand-side one is of size 500 (sampled uniformly at random among all trees of size 500), with 495 internal nodes labelled with 372 distinct labels.

Because of the possible repetition of labels, this class of labelled trees cannot be directly specified using the classical analytic combinatorics operators for labelled structures. However, the following recursive construction allows us to specify the class of strict monotonic Schröder trees using operators for unlabelled structures. Every strict monotonic Schröder tree can be built as follows:

- *Start with a single (unlabelled) leaf.*
- *At step each step ℓ (for $\ell \geq 1$), select a non-empty subset of leaves and replace each of them by an internal node with label ℓ attached to a sequence of at least two leaves.*

3.2. **Enumeration and relationship with ordered Bell numbers.** Using the iterative construction described above, we deduce the following specification for the class $\mathcal{G}$ of all strict monotonic Schröder trees:

$$\mathcal{G} = \mathcal{Z} \cup \big(\mathcal{G}[\mathcal{Z} \to (\mathcal{Z} \cup \textsc{Seq}_{\geq 2}\, \mathcal{Z})]\big) \setminus \mathcal{G}.$$

Note that again the labelling is transparent and does not appear directly in the specification. The combinatorial meaning of this specification is the following: A tree of $\mathcal{G}$ is either a single leaf, or it is obtained by taking an already constructed tree in $\mathcal{G}$, and replace each leaf by either a leaf (i.e. no change) or an internal node attached to a sequence of at least two leaves. Furthermore we omit the case where no leaf is changing (this is why we subtract the set $\mathcal{G}$). Note that subtracting $\mathcal{G}$

is important, otherwise some integer values could be absent in the final tree. For example, if there is no change at step 2 but then the evolution continues, then 2 would not appear in the final tree but larger integers would appear as labels.

Using the symbolic method, we can translate this specification into a functional equation (with substitution) for the ordinary generating series:

$$G(z) = z + G\left(z + \frac{z^2}{1-z}\right) - G(z) = z + G\left(\frac{z}{1-z}\right) - G(z). \tag{17}$$

From this equation we extract the recurrence for the number g_n of strict monotonic Schröder trees with n leaves: we get

$$\begin{aligned} g_n = [z^n]G(z) &= [z^n]\left(z + G\left(z + \frac{z^2}{1-z}\right) - G(z)\right) \\ &= \delta_{n,1} + [z^n]\sum_{\ell \geq 1} g_\ell \left(\frac{z}{1-z}\right)^\ell - g_n \\ &= \delta_{n,1} - g_n + \sum_{\ell \geq 1} g_\ell [z^{n-\ell}] \left(\frac{1}{1-z}\right)^\ell . \end{aligned}$$

We use Kronecker's notation: $\delta_{n,1} = 1$ if $n = 1$ and 0 otherwise. The last coefficient extraction is similar to the integer composition (see [FS09, Example I.3, p. 44]). This implies

$$g_n = \begin{cases} 1 & \text{if } n = 1, \\ \sum\limits_{\ell=1}^{n-1} \binom{n-1}{\ell-1} g_\ell & \text{otherwise.} \end{cases} \tag{18}$$

The first coefficients are equal to a shift of the sequence of ordered Bell numbers (also called Fubini numbers or surjection numbers) referenced as OEIS A000670:

$$(g_n)_{n\in\mathbb{N}} = (0, 1, 1, 3, 13, 75, 541, 4683, 47293, 545835, 7087261, 102247563, 1622632573, \dots).$$

We recall that the n-th ordered Bell number counts the number of ordered partitions of a set of size n, where an ordered partition of a set $\mathcal{S}$ is an ordered sequence of disjoint subsets of $\mathcal{S}$ whose union is equal to $\mathcal{S}$. Ordered Bell numbers are specified by

$$B = \mathrm{SEQ}\left(\mathrm{SET}_{\geq 1}\, \mathcal{Z}\right). \tag{19}$$

Motivated by this remark, we define in Section 3.3 a bijection between the set of strict monotonic Schröder trees and the set of ordered partitions.

Following the approach developed by Pippenger in [Pip10] for ordered Bell numbers, we compute the exponential generating function of $\mathcal{G}$, i.e. we apply the Borel transform on $G(z)$. But first let us recall some basic properties of the latter transform. The *Borel transform*, which $\mathcal{B}$ denotes, takes as an argument an ordinary generating function and gives as its image the corresponding exponential generating series. More precisely, for all real-valued sequence $(a_n)_{n\geq 0}$, we set

$$\mathcal{B}\left[\sum_{n\geq 0} a_n z^n\right] = \sum_{n\geq 0} a_n \frac{z^n}{n!}.$$

Note that if $t_n \leq \rho^n\, n!$ for n sufficiently large then $\mathcal{B}T(z)$ is analytic around 0. It is easy to check that:

Fact 3.2.1. *For all ordinary generating function $f = f(z)$, we have*

$$(i)\ \mathcal{B}\left[zf(z)\right] = \int_0^z \mathcal{B}(f)(t)\mathrm{d}t \qquad \text{and} \qquad (ii)\ \mathcal{B}\left[f'(z)\right] = (\mathcal{B}[f(z)])' + z(\mathcal{B}[f(z)])''.$$

Proposition 3.2.2. *The exponential generating function enumerating strict monotonic Schröder trees is*

$$\mathcal{B}G(z) = \frac{1}{2}\left(z - \ln\left(2 - \mathrm{e}^z\right)\right).$$

Proof. Using Equation (18) and the fact that $g_0 = 1$, we obtain

$$g_n = \delta_{n,1} + \sum_{\ell=1}^{n-1} \binom{n-1}{\ell-1} g_\ell.$$

Adding g_n to both sides (multiplied by $\binom{n-1}{n-1} = 1$ on the right-hand side) gives

$$2g_n = \delta_{n,1} + \sum_{\ell=1}^{n} \binom{n-1}{\ell-1} g_\ell.$$

This recurrence can be directly used to derive an equation for the exponential generating function of $\mathcal{G}$:

$$2\mathcal{B}G(z) = z + \sum_{n\geq 1}\sum_{\ell=1}^{n} \binom{n-1}{\ell-1} g_\ell \frac{z^n}{n!},$$

which is the classical equation satisfied by the ordered Bell numbers. Following the approach of [Pip10], we differentiate the equation with respect to z and get

$$2(\mathcal{B}G(z))' = 1 + \sum_{n\geq 1}\sum_{\ell=1}^{n} \binom{n-1}{\ell-1} g_\ell \frac{z^{n-1}}{(n-1)!}.$$

Since the sum is the convolution of $\mathcal{B}G'(z)$ with $\exp(z)$, we get

$$(\mathcal{B}G(z))' = \frac{1}{2-\mathrm{e}^z},$$

which implies $\mathcal{B}G(z) = \left(z - \ln\left(2 - \mathrm{e}^z\right)\right)/2$ as claimed. □

Recall that ordered Bell numbers are specified by $\mathcal{B} = \textsc{Seq}(\textsc{Set}_{\geq 1}\,\mathcal{Z})$ and thus have exponential generating function $B(z) = 1/(2 - \mathrm{e}^z)$. This directly implies that our sequence $(g_n)_{n\geq 0}$ is equal to the sequence of ordered Bell numbers shifted by one, since $B(z)$ is the derivative of $\mathcal{B}G(z)$. This link between strict monotonic trees and ordered Bell numbers has the interesting following consequence: we have shown that the (shifted) ordinary generating function of the ordered Bell numbers satisfies Equation (17). As far as we can tell, this was not known before.

The asymptotic behaviour of ordered Bell numbers is known (see, e.g., [FS09, p. 109]): if we denote by b_n the n-th ordered Bell number, then

$$b_n = \sum_{\ell=0}^{n} \ell! \left\{ {n \atop \ell} \right\} \underset{n\to\infty}{\sim} \frac{n!}{2\left(\ln 2\right)^{n+1}},$$

where the $\left\{ {n \atop \ell} \right\}$'s are the Stirling partition numbers (also called Stirling numbers of the second kind, see [FS09, Appendix A.8]). They count the number of ways to partition a set of n objects into k non-empty subsets.

The number b_n is equal to the number g_{n+1} of strict monotonic Schröder trees of size $n+1$, which implies that, for all $n \geq 1$,

$$g_n = \sum_{\ell=0}^{n-1} \ell! \left\{ {n-1 \atop \ell} \right\} \underset{n\to\infty}{\sim} \frac{(n-1)!}{2\,(\ln 2)^n}.$$

3.3. **Bijection with ordered Bell numbers.** Since the number of strict monotonic Schröder trees of size $n+1$ is equal to the number of ordered partitions of a set of size n, it is natural to try to find an explicit bijection between the two classes. In this section, we exhibit such a bijection.

To describe precisely the bijection we need the following definitions and notations. Recall that the the subsets of an ordered partitions are ordered but the elements inside each subset are not. In the following, we denote by $p = (p_1, p_2, \dots, p_\ell)$ the ordered partition of ordered subsets $p_1, \dots, p_\ell$; for example, $(\{3,4\},\{1,5,7\},\{2,6\}) \neq (\{2,6\},\{3,4\},\{1,5,7\})$. We denote by $|p_i|$ the size of the i-th subset of p, and by $|p| = \sum_{i=1}^{\ell} p_i$ its total size (i.e. the number of elements of $\cup_{i=1}^{\ell} p_i$). Let $a = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ (with $r \geq 1$) be a subset of $\mathbb{N}$; without loss of generality, we can assume that $\alpha_1 < \alpha_2 < \ldots < \alpha_r$. A `run` of a is a maximal sequence $(\alpha_i, \alpha_{i+1}, \dots, \alpha_j)$ $(1 \leq i \leq j \leq r)$ of consecutive integers, i.e. $(\alpha_i, \alpha_{i+1}, \dots, \alpha_j) = (\alpha_i, \alpha_i + 1, \dots, \alpha_i + j - i)$, $\alpha_{i-1} < \alpha_i - 1$ and $\alpha_{j+1} > \alpha_j + 1$. We define the function `runs` as the function that lists all the runs of a subset: for example, $\texttt{runs}(\{3,4\}) = (\{3,4\})$ and $\texttt{runs}(\{1,3,6,7\}) = (\{1\},\{3\},\{6,7\})$.

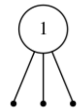
An ordered partition $p = (p_1, \dots, p_\ell)$ is called *incomplete* if and only if $\cup_{i=1}^{\ell} p_i \neq \{1, 2, \dots, |p|\}$: e.g. the partition $(\{3,4\},\{1,5,7\})$ is incomplete due to the fact that $\cup_{i=1}^{\ell} p_i = \{1,3,4,5,7\} \neq \{1,2,3,4,5\}$. We define the normalization of a partition p (either incomplete or not), denoted by $\texttt{norm}(p)$, as the ordered partition of $\{1, \dots, |p|\}$ that keeps the relative order between the elements. For example, if $p = (\{3,4\},\{1,5,7\})$, then $\texttt{norm}(p) = [\{2,3\},\{1,4,5\}]$.

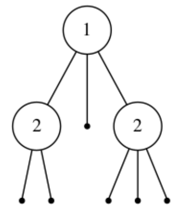
We are now ready to describe our bijection: we first define the mapping `M'`, which associates a strict monotonic Schröder tree to each (possibly incomplete) ordered partition $p = (p_1, \dots, p_\ell)$. Before starting we fix an arbitrary order for the leaves in the tree once and for all (for example, the one given by the postorder traversal of the tree). Then The tree $\texttt{M'}(p)$ is the result of the following recursive procedure:

- At time zero, consider a tree with one internal node labelled by 1 to which are attached $|p_1| + 1$ leaves.
- At each time $2 \leq i \leq \ell$, we denote by $p'_1, \dots, p'_i$ the ordered subsets of the renormalization of $(p_1, \dots, p_i)$, i.e. $\texttt{norm}((p_1, \dots, p_i)) = (p'_1, \dots, p'_i)$. We denote by $r_1, \dots, r_j$ the runs of p'_i, i.e. $\texttt{runs}(p'_i) = (r_1, \dots, r_j)$; recall that each of $r_1, \dots, r_j$ is a set of successive integers, possibly reduced to a singleton and iterate the following process: for k from 1 to j, take the leaf whose index is the first element of r_k and replace it with an internal node with label k attached to $|r_k| + 1$ leaves.

In Figure 6 we show how to construct $\texttt{M}'(p)$ when $p = (\{3,4\},\{1,5,7\},\{2,6\})$. The resulting strict monotonic Schröder tree is of size 8. It is straightforward to check that `M'` is indeed a bijection.

3.4. **Analysis of typical parameters.** In this section, we give information about the shape of a *typical* strict monotonic Schröder tree: more precisely, we prove limit

$\mathtt{norm}((\{3,4\})) = (\{1,2\}) \xrightarrow{\mathtt{M}'}$ 

$\mathtt{norm}((\{3,4\},\{1,5,7\})) = (\{2,3\},\{1,4,5\}) \xrightarrow{\mathtt{M}'}$ 

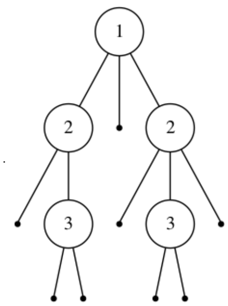
$\mathtt{norm}((\{3,4\},\{1,5,7\},\{2,6\})) = (\{3,4\},\{1,5,7\},\{2,6\})$ 

FIGURE 6. The constructive bijection between an ordered partition and a strict monotonic Schröder tree

theorems for the number of distinct labels, the number of internal nodes and the arity of the root in a tree picked uniformly at random among all strict monotonic Schröder trees of size n (i.e. with n leaves).

3.4.1. *Quantitative analysis of the number of iteration steps.* The main novelty of strict monotonic Schröder trees compared to increasing Schröder trees is that repetitions of labels are allowed: it is thus natural to ask how many repetitions there are in a *typical* strict monotonic Schröder tree. To answer this question, on can mark iterations by adding a new variable u in Equation (17):

$$G(z,u) = z + u\, G\left(\frac{z}{1-z}, u\right) - u\, G(z,u),$$

which implies

$$g_{n,k} = \begin{cases} 1 & \text{if } n = 1 \text{ and } k = 0, \\ \sum\limits_{\ell=1}^{n-1} \binom{n-1}{\ell-1}\, g_{\ell,k-1} & \text{otherwise,} \end{cases} \tag{20}$$

with n being the size and k the number of iteration steps (i.e. the number of distinct labels). In Figure 7, we show the first values of $(g_{n,k})$ that are stored in OEIS A019538.

This recurrence is analogous to the one relating ordered Bell numbers and Stirling partition numbers (see Equation (18)).

1,						
0,	1,					
0,	1,	2,				
0,	1,	6,	6,			
0,	1,	14,	36,	24,		
0,	1,	30,	150,	240,	120,	
0,	1,	62,	540,	1560,	1800,	720

FIGURE 7. Distribution of $(g_{n,k})_k$ for $n \in \{1, \dots, 7\}$

Theorem 3.4.1. *The number of strict monotonic Schröder trees of size n with exactly k distinct labels is given by*

$$g_{n,k} = k! \left\{ {n+1 \atop k} \right\}.$$

We denote by $X_n^{\mathcal{G}}$ the number of distinct labels in a tree picked uniformly at random among all strict monotonic Schröder trees of size n: for all $n \geq 1$, $X_n^{\mathcal{G}}$ is a random variable such that $\mathbb{P}(X_n^{\mathcal{G}} = k) = g_{n,k} / \sum_{k=1}^{n} g_{n,k}$. Then, asymptotically when n tends to infinity,

$$\frac{X_n^{\mathcal{G}} - \frac{n}{2\ln 2}}{\sqrt{\frac{(1-\ln 2)n}{(2\ln 2)^2}}} \xrightarrow{d} \mathcal{N}(0,1).$$

The analysis of the limiting distribution is classical in the quasi-powers framework established by Hwang [Hwa98]; see [FS09, p. 645, 653] for details and applications.

Proof. Recall that $g_{n,k} = k! \left\{ {n+1 \atop k} \right\}$ is the number of ordered partitions of a set of size n having k non-empty parts. It is known (see, e.g. [Ben73, Example 3.4]) that, if K_n is the number of parts in an ordered set partition of size n, then

$$\frac{K_n - \frac{n}{2\ln 2}}{\sqrt{\frac{(1-\ln 2)n}{(2\ln 2)^2}}} \xrightarrow{d} \mathcal{N}(0,1),$$

in distribution. This concludes the proof since K_n has the same distribution as $X_n^{\mathcal{G}}$ for all $n \geq 1$. $\square$

3.4.2. *Quantitative analysis of the number of internal nodes.* In this model the number of internal nodes is different from the number of distinct labels that appear in the tree: this is because one integer can label several internal nodes. It is thus natural to ask how many internal nodes a typical strict monotonic Schröder trees of size n (i.e. with n leaves) has. The specification marking both leaves (with variable z) and internal nodes (with variable u) is

$$G(z,u) = z + G\left(z + \frac{uz^2}{1-z}, u\right) - G(z,u). \tag{21}$$

We recall that the substitution $z \to z + \frac{uz^2}{1-z}$ means that at each iteration each leaf can be left as it is ($z \to z$) or expanded into an internal node attached to an arbitrary number of leaves ($z \to \frac{z^2}{1-z}$). A new internal, marked with the variable u, is created only in the second case.

$$
\begin{array}{lllllll}
1, & & & & & & \\
0, & 1, & & & & & \\
0, & 1, & 2, & & & & \\
0, & 1, & 5, & 7, & & & \\
0, & 1, & 9, & 31, & 34, & & \\
0, & 1, & 14, & 86, & 226, & 214 & \\
0, & 1, & 20, & 190, & 874, & 1946, & 1652
\end{array}
$$

FIGURE 8. Distribution of $(g_{n,k})_k$ for $n \in \{1, \dots, 7\}$

For all $1 \leq n$ and $1 \leq k \leq n-1$, we denote by $g_{n,k}$ the number strict monotonic Schröder trees with n leaves and k internal nodes: Figure 8 shows the values of $(g_{n,k})_{1\leq k\leq n-1}$ for $n \in \{1, 2, \dots, 7\}$. This triangle of integers is not yet stored in `OEIS`. However, its diagonal is equal to `OEIS A171792`. In fact in the diagonal the numbers corresponds to the number of strict monotonic trees with n leaves and $n-1$ internal nodes, i.e. binary strict monotonic trees: this class of trees is studied in [BGGW20].

Theorem 3.4.2. *If we denote by $I_n^{\mathcal{G}}$ the (random) number of internal nodes in a tree picked uniformly at random among all strict monotonic Schröder trees of size n, then, asymptotically when n tends to infinity,*

$$\mathbb{E}[I_n^{\mathcal{G}}] \underset{n\to\infty}{=} n - (\ln 2)(\ln n) + \frac{\pi^2}{12} - 1 + (\ln 2)\left(-\gamma + \frac{\ln 2}{2} + \ln\ln 2\right) + o(1),$$

where γ is the Euler-Mascheroni constant.

Proof. For all $n \geq 1$, we denote by $h_n = \sum_{k=1}^{n-1} k g_{n,k}$, and let H be the ordinary generating function of $(h_n)_{n\geq 1}$; we have

$$H(z) = \left(\frac{\partial G(z,u)}{\partial u}\right)_{|u=1}.$$

The ratio h_n/g_n is equal to the expected number of internal nodes in a tree taken uniformly at random among all strict monotonic Schröder trees of size n; we are thus interested in the asymptotic behaviour of this ratio. Differentiating according to u and then substituting u by 1 in Equation (21) gives

$$H(z) = \frac{z^2}{1-z} G'\left(\frac{z}{1-z}\right) + H\left(\frac{z}{1-z}\right) - H(z), \tag{22}$$

because

$$\left(\frac{\partial G(z,u)}{\partial z}\right)_{|u=1} = G'(z).$$

Since Equation (22) is similar to Equation (17), we apply the same method as in the proof of Proposition 3.2.2. We first derive

$$(\mathcal{B}H(z))' = \frac{1}{2-e^z}\left(\mathcal{B}\left[\frac{z^2}{1-z} G'\left(\frac{z}{1-z}\right)\right]\right)'.$$

Then using Equation (17) we deduce

$$\left(\mathcal{B}\left[\frac{z^2}{1-z} G'\left(\frac{z}{1-z}\right)\right]\right)' = -z + \frac{z^2}{2} + 2\left(\mathcal{B}\left[z^2(1-z)G'(z)\right]\right)'.$$

Furthermore since for any function F we have $\mathcal{B}zF(z) = \int_0^z \mathcal{B}F(t)\mathrm{d}t$, we can simplify the equation into

$$(\mathcal{B}H(z))' = \frac{1}{2-\mathrm{e}^z}\left(-z + \frac{z^2}{2} + 2\int_0^z \mathcal{B}G'(t)\,\mathrm{d}t - 2\int_0^z\int_0^t \mathcal{B}G'(u)\,\mathrm{d}u\,\mathrm{d}t\right).$$

Then, since $\int_0^z \mathcal{B}G'(t)\mathrm{d}t = z(\mathcal{B}G(z))'$, we obtain

$$\begin{aligned}(\mathcal{B}H(z))' &= \frac{1}{2-\mathrm{e}^z}\left(-z + \frac{z^2}{2} + 2z(\mathcal{B}G(z))' - 2\int_0^z t(\mathcal{B}G(t))'\mathrm{d}t\right)\\ &= \frac{1}{2-\mathrm{e}^z}\left(-z + \frac{z^2}{2} + \frac{2z}{2-\mathrm{e}^z} - 2\int_0^z \frac{t}{2-\mathrm{e}^t}\,\mathrm{d}t\right)\\ &= \frac{1/2}{1-\mathrm{e}^z/2}\left(-\frac{\pi^2}{12} + \frac{(\ln 2)^2}{2} - z\left(1 - \ln(1-\mathrm{e}^z/2) - \frac{1}{1-\mathrm{e}^z/2}\right) + \mathrm{Li}_2(\mathrm{e}^z/2)\right),\end{aligned}$$

where Li_2 is the dilogarithm function, defined in [FS09, section VI.8.]. Using its asymptotic development at 1, we get

$$\begin{aligned}(\mathcal{B}H(z))' \underset{z\to\ln 2}{\sim}\ & \frac{1}{2\ln 2}\frac{1}{(1-z/\ln 2)^2}\\ & - \left(\frac{1}{2\ln 2} - \frac{\pi^2}{24\ln 2} + \frac{\ln 2}{4} - \frac{\ln 2 + \ln\ln 2 + \ln(1-z/\ln 2)}{2}\right)\frac{1}{1-z/\ln 2}\\ & - \frac{1}{2} - \frac{7\ln 2}{24} + \frac{\pi^2}{48} + \frac{(\ln 2)^2}{8} + \frac{\ln 2\ln\ln 2}{4} + O\left(\ln\left(\frac{1}{1-z/\ln 2}\right)\right).\end{aligned}$$

By using classical transfer theorems we obtain the result by extracting the $(n-1)$-th coefficient of $(\mathcal{B}H(z))'$ and dividing it by the n-th coefficient of $\mathcal{B}G(z)$. □

3.4.3. *Quantitative characteristics of the root node.* In this section, we look at the arity of the root in a typical strict monotonic Schröder tree. We denote by $A_n^{\mathcal{G}}$ the arity of the root in a tree picked uniformly at random among all strict monotonic Schröder trees of size n, and by p_n its probability generating function:

$$p_n(u) = \sum_{k\geq 0} \mathbb{P}(A_n^{\mathcal{G}} = k)\ u^k.$$

Theorem 3.4.3. *Asymptotically when n tends to infinity, $A_n^{\mathcal{G}}$ converges in distribution to a (shifted) zero-truncated Poisson law with parameter $\ln 2$, i.e. for all $u \geq 0$,*

$$p_n(u) \underset{n\to\infty}{\to} u\mathrm{e}^{u\ln 2} - u.$$

This implies that $\mathbb{E}[A_n^{\mathcal{G}}] \to 2\ln 2 + 1$ when n tends to infinity.

Proof. Thanks to the bijection of Section 3.3, we know that $A_n^{\mathcal{G}}$ is equal to the size of the first subset in an ordered partition picked uniformly at random among all ordered partitions of $\{1, \dots, n-1\}$. We denote by $\mathcal{P}$ the class of ordered partitions, 1 is the empty partition, $\mathcal{Z}$ is a singleton, and $\mathcal{U}$ marks the elements in the first subset. Here the specification is defined in the context of labelled object, thus the associated generating functions are exponential (see [FS09] for notation details):

$$\mathcal{P} = 1 + \underset{\geq 1}{\mathrm{SET}}(\mathcal{U}\mathcal{Z}) \star \mathrm{SEQ}(\underset{\geq 1}{\mathrm{SET}}\,\mathcal{Z}).$$

Using the symbolic method for exponential generating function, we get

$$P(z,u) = 1 + \frac{e^{uz} - 1}{2 - e^z}.$$

Thus, if we set

$$\tilde{p}_n(u) = \frac{[z^n]P(z,u)}{[z^n]P(z,1)},$$

for all $n \geq 0$, then

$$[z^n]P(z,u) \underset{n\to\infty}{\to} \frac{1}{2}\left(2^u - 1\right)(\ln 2)^{-n-1}.$$

This implies that, for all $u \geq 0$,

$$\tilde{p}_n(u) \underset{n\to\infty}{\to} 2^u - 1.$$

Note that, by definition, $\tilde{p}_n(u)$ is the probability generating function of the size S_n of the first subset in an ordered partition picked uniformly at random among all ordered partitions of $\{1, \ldots, n-1\}$. Because of the bijection of Section 3.3, we know that $A_n^{\mathcal{G}}$ and S_{n-1} have the same distribution, implying that $p_n(u) = u\tilde{p}_n(u)$. This concludes the proof. □

3.4.4. *Typical depth of the leftmost leaf.* In this section, we prove a central limit theorem for the depth of the leftmost leaf in a typical strict monotonic Schröder tree:

Proposition 3.4.4. *Let $Y_n^{\mathcal{G}}$ be the depth of the leftmost leaf in a tree taken uniformly at random among all increasing Schröder trees of size n. In distribution when n tends to infinity, we have*

$$\frac{Y_n^{\mathcal{G}} - \ln n}{\sqrt{\ln n}} \xrightarrow{d} \mathcal{N}(0,1).$$

The depth of the leftmost leaf is a lower bound for the height (since the height of a tree is the maximal depth of its leaves), and it has the advantage of being easier to specify and analyse than the height itself. Note that the choice of the leftmost leaf is arbitrary (i.e. choosing another leaf would lead to the same result), but it has the advantage that the specification is straightforward. Recall that, in Section 2.3.4, we proved a similar central limit theorem for the depth of the leftmost leaf in a typical increasing Schröder tree.

In this section, the variable $\mathcal{U}$ marks the depth of the leftmost leaf. Using the evolution process, we get that

$$G(z,u) = z + \left(\frac{G(y,u)}{y}\right)_{|y=\frac{z}{1-z}} \cdot \left(z + \frac{uz^2}{1-z}\right) - G(z,u).$$

At each iteration step we start by chopping off the leftmost leaf (this corresponds to $G(y,u)/y$). Each of the other leaves either stays unchanged or is replaced by an internal node to which is attached a sequence of at least two leaves (this corresponds to substituting y by $z/(1-z)$). Finally we put back the leaf that has been chopped off and there we have 2 choices, either it stays unchanged (z) or it is replaced by an internal node with at least two leaves attached to it ($z^2/(1-z)$) in which case we multiply by u because the depth of the leftmost leaf has been increased by one.

Iterating this specification, we can calculate the first coefficients (see Table 5): they are equal to the first coefficients of a shifted version of OEIS A129062. From the specification it is possible to derive a recurrence relation on the coefficients $g_{n,k}$. We have $g_{1,0} = 1$ and for all $n \geq 2$ and $1 \leq k \leq n-1$,

$$g_{n,k} = \sum_{\ell=k+1}^{n-1} g_{\ell,k} \binom{n-2}{\ell-2} + \sum_{\ell=k}^{n-1} g_{\ell,k-1} \binom{n-2}{\ell-1}. \tag{23}$$

This equation can be interpreted combinatorially using the evolution process (this reasoning is similar to the one leading to Equation (18).): At each iteration step, the height of the leftmost leaf either stays unchanged or increases by 1. Therefore each tree in $\mathcal{G}_{n,k}$, the set of all strict monotonic Schröder trees whose leftmost leaf is at depth k, was, before the last iteration, either a tree of $\mathcal{G}_{\ell,k}$ of a tree of $\mathcal{G}_{\ell,k-1}$, for some $\ell < n$. There are $\binom{n-2}{\ell-1}$ to expand a tree of $\mathcal{G}_{\ell,k-1}$ into a tree of $\mathcal{G}_{n,k}$ in one iteration step: it is the number of ways to partition the n leaves of the size-n tree into ℓ parts of size at least one (when a part is of size 1, the corresponding leaf in the ℓ-size node stays unchanged, otherwise, it becomes an internal node of out-degree the size of the part) in a way that the first part is of size at least 2 (the left-most leaf becomes an internal node attached to two leaves). Similarly, there are $\binom{n-2}{\ell-2}$ number of ways to expand a tree of $\mathcal{G}_{\ell,k}$ into a tree of $\mathcal{G}_{n,k}$ in one iteration step: it is the number of ways partition the n leaves of the size-n tree into ℓ parts of size at least one such that the first part is of size 1.

Rather than trying to solve the above recurrence we exhibit an interesting relationship between the depth of leftmost leaf and the cycles of ordered Bell numbers. In fact, it is well-known that, in the symbolic method, sequences are in one-to-one correspondence with sets of cycles:

$$\textsc{Seq} \cong \textsc{Set} \circ \textsc{Cyc}\,. \tag{24}$$

In Section 3.3 we exhibited a bijection between strict monotonic Schröder trees and ordered Bell numbers. This bijection thus implies that

$$\begin{aligned} G(\mathcal{Z}) &\cong \textsc{Seq}\,(\textsc{Set}_{\geq 1}\,\mathcal{Z}) \\ &\cong \textsc{Set}\,(\textsc{Cyc}\,(\textsc{Set}_{\geq 1}\,\mathcal{Z}))\,. \end{aligned}$$

To prove Proposition 3.4.4, we first exhibit the bijection $\mathtt{S}$ between the two latter sets. To do so, we first need to define better our notations: for example, since the order in a set is not relevant, i.e. $\{1,3,2\} = \{1,2,3\} = \{3,2,1\}$, we choose to always use the representation $\{i_1, \dots, i_m\}$ such that $i_1 < i_2 < \cdots < i_m$. Similarly, a cycle of sets, e.g. $(\{3,4\},\{1,5,6\},\{2\})$, is invariant by cyclic permutation of its elements, i.e.

$$C = (\{2,4\},\{1,5,6\},\{3\}) = (\{1,5,6\},\{3\},\{2,4\}) = (\{3\},\{2,4\},\{1,5,6\}).$$

In the following we choose to always use the representation such that the first element in the cycle contains 1: for our example, $C = (\{1,5,6\},\{3\},\{2,4\})$. Finally, given a set of cycles, we choose the representation in which the cycles are in decreasing lexicographic order: to each cycle, we associate the string of integers obtained from reading its elements from left to right, for example to $(\{1,5,6\},\{3\},\{2,4\})$, we associate 156324, and then order the cycles of sets according to this order. For example, the set $\{(\{1,2,4\},\{3\}),(\{1,3\},\{2,4,5\})\}$ has 2 cycles, the list of the string of the first one is 1243, the string of the second one is 13245. Since $3 > 2$, the canonical representation of this set of cycles is

$$\{(\{1,3\},\{2,4,5\}),(\{1,2,4\},\{3\})\}.$$

We are now ready to define the mapping $\mathtt{S}$: take S a set of cycles of sets of integers (in its canonical representation), and we denote by $X(S)$ the string of integers read from left to right in this canonical representation. E. g. if $S = \{(\{1,3\},\{2,4,5\}),(\{1,2,4\},\{3\})\}$, then $X(S) = 132451243$. Now define $\hat{X}(S)$ as a string of zeros of the same length as $X(S)$, and $c = 1$, and for all i between 1 and the maximum integer in S, go through the string $X(S)$ from right to left, i.e. for all j from length$(X(S))$ down to 1, if the digit in j-th position is a 1, replace the j-th digit in $\hat{X}(S)$ by c and increase c by 1. In our example, we eventually obtain $\hat{X}(S) = 264891375$. We denote by $s_1, \dots, s_m$ as the sizes of the sets of the cycles of S (in the order of the canonical representation); in our example, there are $m = 4$ sets in total (in the two cycles) and their sizes are $2, 3, 3, 1$. Define $\mathtt{S}(S)$ as the ordered partition having m parts of respective sizes $s_1, \dots, s_m$ and such that the elements of the first part are the first s_1 digits of $\hat{X}(S)$, the elements of the second part are the following s_2 digits of $\hat{X}(S)$ and so on. On our example, we get

$$\mathtt{S}(S) = (\{2,6\},\{4,8,9\},\{1,3,7\},\{5\}).$$

Lemma 3.4.5. *The mapping* $\mathtt{S}$ *is a one-to-one map from* $\textsc{Set}(\textsc{Cyc}(\textsc{Set}_{\geq 1}\,\mathcal{Z}))$ *onto* $\textsc{Seq}(\textsc{Set}_{\geq 1}\,\mathcal{Z})$.

Proof. Given an ordered partition, i.e. a sequence os sets $S_1, \dots, S_m$. Denote by k the first integer in S_1 (since we use the canonical representation, it is also the smallest integer in S_1). And denote by i_j the integer such that S_{i_j} contains j, for all $1 \leq j \leq k-1$. Note that $1 = i_k < i_{k-1} < \ldots < i_1$, and define $C_1 = (S_1, \dots, S_{i_{k-1}-1})$, $C_2 = (S_{i_{k-1}}, \dots, S_{i_{k-2}})$, until $C_k = (S_{i_1}, \dots, S_m)$. And set $\mathtt{S}^{-1}((S_1, \dots, S_m)) = \{C_1, \dots, C_k\}$. One can check that this is indeed the inverse of $\mathtt{S}$, which concludes the proof. □

Recall that, in Section 3.3, we have defined $\mathtt{M}'$, a bijection from the set of ordered partitions onto the set of strict monotonic Schröder trees. Therefore, $\mathtt{M}' \circ \mathtt{S}$ is a bijection from the $\textsc{Set}(\textsc{Cyc}(\textsc{Set}_{\geq 1}\,\mathcal{Z}))$ onto the set of strict monotonic Schröder trees.

Lemma 3.4.6. *If* $X \in \textsc{Set}\,(\textsc{Cyc}\,(\textsc{Set}_{\geq 1}\,\mathcal{Z}))$, *then the number of cycles of* X *is equal to the depth of the leftmost leaf of* $\mathtt{M}' \circ \mathtt{S}(X)$.

Proof. If X contains m cycles, then the integers $1, 2, \dots, m$ appear in reverse order and in different sets $s_m, s_{m-1}, \dots, s_1$ of the ordered partition $\mathrm{S}(X)$: s_i is the set containing the integer i for all $1 \leq i \leq m$. Moreover s_m is the first set in $\mathtt{S}(X)$ because the cycles are ordered in the canonical order. In the mapping $\mathtt{M}'$, s_m will form the root of the tree. Then s_{m-1} will create a node in the leftmost leaf, then s_{m-2} will create a node in the leftmost leaf and so on until s_1 is added to create a last node on the leftmost leaf. Thus the depth of the leftmost leaf is m. □

Proposition 3.4.7. *The exponential generating function of* $g_{n,k}$ *is*

$$G(z,u) = \sum_{n\geq 0}\sum_{k\geq 0} g_{n,k} u^k \frac{z^n}{n!} = \int_0^z \left(\frac{1}{2-\mathrm{e}^x}\right)^u \mathrm{d}x.$$

Proof. From the discussion above, since the depth of the leftmost leaf is the number of cycles we get a direct specification by marking the cycles in the following

$$\mathcal{D} = \textsc{Set}\,(\mathcal{U}\,\textsc{Cyc}\,(\textsc{Set}_{\geq 1}\,\mathcal{Z}))\,.$$

1						
0	1					
0	2	1				
0	6	6	1			
0	26	36	12	1		
0	150	250	120	20	1	
0	1082	2040	1230	300	30	1

TABLE 5. Values of $g_{n,k}$, the number of n strict monotonic Schröder trees of size n with leftmost leaf at depth k, for all $0 \leq k \leq n \in \{1, \ldots, 7\}$.

Therefore $D(z, u) = \exp\left(u \ln\left(\frac{1}{1-(\exp(z-1))}\right)\right)$. The number of trees of size n is the number of ordered Bell numbers of size $n-1$, so we integrate the last expression. □

The discussion above also leads to the following identity

Proposition 3.4.8. *For all $n \geq 2$ and $1 \leq k \leq n-1$,*

$$g_{n,k} = \sum_{m=0}^{n-1} \left\{ {n-1 \atop m} \right\} \left[{m \atop k} \right].$$

Where $\left[{n \atop k} \right]$ are Stirling Cycle numbers (also known as Stirling numbers of the first kind). They count the number of cycles of size k in a permutation of size n.

Proof. The proof is a direct consequence of the previous construction. The number of trees of size n with leftmost leaf at depth k can be constructed by looking at set partitions of size $n-1$ elements into i subsets for all possible sizes of i which is counted by $\left\{ {n-1 \atop i} \right\}$ then for each partition of size i we see how many cycles of size k we can build with it. □

Proof of Proposition 3.4.4. We can make the calculations on the bivariate generating function $D(z, u)$ which enters the scope of quasi-powers framework. Theorem IX.11 in [FS09, p. 669] is applicable. The exponent $\alpha(u) = u$ is analytic and $\alpha(1) = 1$ and it satisfies $\alpha'(1) + \alpha''(1) = 1 \neq 0$. So $D(z, u)$ is asymptotically Gaussian with mean and variance as announced. Finally the shift that we have between the size of trees and the ordered partitions does not affect the first orders since $\ln(n+1) \sim \ln n$. □

3.5. **Uniform random sampling.** To sample uniformly at random a strict monotonic Schröder tree of size n, we could choose a two-step algorithm. First we sample uniformly an ordered partition of the set $\{1, \ldots, n-1\}$ and then with the use the bijection of Section 3.3 we transform it into a strict monotonic Schröder tree. But here, in this section, we prefer to present a direct algorithm that generates uniformly a strict monotonic Schröder tree, i.e. without the intermediate step of generating another combinatorial object like an ordered partition.

The global approach for our algorithmic framework deals with the *recursive generation method* adapted to the analytic combinatorics point of view in [FZVC94]. But in our context we note that we can obtain for free (from a complexity view) an *unranking algorithm.* This fact is sufficiently rare to mention it: usually unranking algorithm are less efficient than recursive generation ones. Unranking algorithmic has been developed in the 70's by Nijenhuis and Wilf [NW75] and then has been introduced to the context of analytic combinatorics by Martínez and Molinero [MM03]. Here the idea is not to draw uniformly an object, but first to define a total order over the objects under consideration (in our context, strict monotonic Schröder trees) and then an integer (named the rank) is chosen to build deterministically the associated object. Obviously if the rank is uniformly chosen among all possible ranks, then the unranking algorithm is nothing else than a uniformly random sampler. But the unranking approach gives also a way for obtaining an exhaustive sampler, just by iterating the sampling over all possible ranks (the reader can refer to the paper [BDGV18] for an example of both methods: recursive generation and unranking).

For both types of algorithms (unranking or recursive generation) some precomputations are done (only once before the sampling of many objects). We compute (and store) the numbers of trees of sizes from 1 to n. This calculation is be done with a quadratic complexity (in the number of arithmetic operations) using the recursive formula for $(g_n)_{n\geq 1}$ (see Equation (20)). This complexity is only achieved if we first compute and memorize all values of $(i!)_{1\leq i\leq n}$.

Then it only remains to build the tree of rank r recursively. If r is sampled uniformly at random in $\{0, 1, \dots, g_n - 1\}$ the algorithm is a uniform sampler and if r is deterministically chosen, then the algorithm is a classical unranking algorithm. To do this, we recall that (see Equation (18)), for all $n \geq 1$,

$$g_n = \binom{n-1}{n-2} g_{n-1} + \binom{n-1}{n-3} g_{n-2} + \cdots + \binom{n-1}{0} g_1, \tag{25}$$

and interpret this equation combinatorially: to build a tree of size n, we take a size $\ell \in \{1, \dots, n-1\}$ tree T_ℓ constructed with exactly one less iteration. To grow it into a size-n tree, we interpret the binomial coefficient $\binom{n-1}{\ell-1}$ as the number of composition of n in ℓ parts: some of the ℓ leaves of T_ℓ are replaced by some internal nodes to which leaves are attached, some leaves remain leaves. To do that we traverse the tree T_ℓ and each time we see a leaf, we do the following action: if the next part (in the composition) is of value 1, we keep the leaf unchanged otherwise for a value $s > 1$, we replace the leaf by an internal node (well labelled with the currently step number) and attached s leaves to it. We then take the next part of the composition into consideration and continue the tree traversal.

Focusing on Equation (25) and the equation above we see that a function allowing the unranking of compositions is necessary. Recall the composition of the integer n into ℓ parts is in bijection with the number of combinations of $(\ell - 1)$ elements chosen in $(n - 1)$ ones. A way to prove it consists in laying $(\ell - 1)$ barriers in the sequence of n bullets in order to define ℓ parts. There are classical algorithms to unrank combinations in the lexicographical order. A first algorithm has been described by Buckles and Lybanon [BL77]. Another, more efficient, has just been settled in the technical report [DGH20]. For both of them we can easily prove that their average complexity (when ℓ rages over all possibilities) is $\Theta(n)$ in the number of arithmetic operations by having first memoized all factorial values of the numbers from 0 to n. In the following we develop a simpler approach based on the classical recursive generation without any lexicographic constraint like in the two mentioned papers. The algorithm is an unranking method for the composition of integers. It is based on the reverse lexicographic order (cf e.g. [Rus03]) so that we get an easier implementation[2]. For simplification, we suppose having memoized all values of $\binom{r}{s}$ for $r \in \{1, n\}$ and $s \in \{1, r\}$. Using the classical Pascal's rule for binomial coefficients, we obtain the following recurrence for the number of composition of n into ℓ:

$$C_{n,\ell} = \binom{n-1}{\ell-1} = C_{n-1,\ell} + C_{n-1,\ell-1}. \tag{26}$$

We thus deduce Algorithm 2 for the unranking method.

Algorithm 2 Reverse Lexicographic Composition Unranking

function $\textsc{UnrankComposition}(n, \ell, r)$
    if $n = \ell$ and $r = 0$ **then**
        return $(1, 1, \ldots, 1)$
    if $r < \binom{n-2}{\ell-1}$ **then**
        $C := \textsc{UnrankComposition}(n-1, \ell, r)$
        $C[0] := C[0] + 1$
        return C
    else
        $s := r - \binom{n-2}{\ell-1}$
        $C := (1) \cup \textsc{UnrankComposition}(n-1, \ell-1, s)$
        return C

Theorem 3.5.1. *The function* UNRANKCOMPOSITION *is an unranking algorithm (based the reverse lexicographic order) and calling it with the parameters $\ell \leq n$ and a uniformly-sampled integer r in $\{0, \ldots, \binom{n-1}{\ell-1} - 1\}$, gives as output a uniform composition of n in ℓ parts.*

Using the memorization of binomial coefficients, the algorithm needs at most $(\ell - 1)$ arithmetic operations on big integers.

Proof. We prove that the algorithm is correct by induction on n. The result is true when $n = \ell = r = 1$ since the algorithm returns (1). Fix an integer n and assume that the algorithm is correct for all $\ell \leq n-1$, and that the total order over compositions is the reverse lexicographic one (see, e.g., [Rus03] for the definition of the reverse lexicographic order). Let ℓ be an integer between 0 and n, and r be an integer chosen uniformly at random in $\{0, \ldots, C_{n,\ell} - 1\}$. Equation (26) implies that a composition of n in ℓ parts is either a composition of $(n-1)$ in ℓ parts whose first part has been increased by one, or it is a composition of $(n-1)$ in $(\ell-1)$ parts, and a new part equal to 1 is added at the beginning of the composition. In both cases, the first elements are all greater than the second elements according to the lexicographic order. The recurrence hypothesis ends the proof since the rank value r (or s in the second case) is adapted to each of the latter cases.

The number of arithmetic operations is direct when all binomial coefficients are first memorised. ☐

In Equation (25) the first term is much bigger than the second one, which is much bigger than the third one and so on. This approach, focusing first on the dominant terms is an adaptation to the idea underlying the *Boustrophedonic order* presented in [FZVC94]. It allows to improve essentially the average complexity of

[2]For the composition unranking, note that it would suffice to look for the rank $\binom{n}{\ell} - 1 - r$ (instead of r) in order to get the lexicographic order.

the random sampling algorithm. In our case of strict monotonic Schröder trees do not follow a standard specification (cf. [FZVC94] for details), the complexity gain is even better. The loop starting in line 6 aims at determining the interesting term

Algorithm 3 Strict monotonic Schröder Tree Unranking

1: **function** UNRANKTREE(n, s)
2: **if** $n = 1$ **then**
3: **return** the tree reduced to a single leaf
4: $\ell := 1$
5: $r := s$
6: **while** $r >= 0$ **do**
7: $r := r - \binom{n-1}{\ell} \cdot g_{n-\ell}$
8: $\ell := \ell + 1$
9: $\ell := \ell - 1$
10: $r := r + \binom{n-1}{\ell} \cdot g_{n-\ell}$
11: $T :=$UNRANKTREE$(n - \ell, r \mod g_{n-\ell})$
12: $C :=$UNRANKCOMPOSITION$(n, n - \ell, r // g_{n-\ell})$
13: Substitute in T, using traversal $\mathcal{T}$, some leaves according to C
14: **return** the tree T

The sequences $(g_\ell)_{\ell \leq n}$ and $(\ell!)_{\ell \in \{1,\dots,n\}}$ have been pre-computed and stored.
Line 13: The operation // is the Euclidean division.

in the sum (25), thus the size of the tree in the evolution process letting to build the tree of rank s and size n.

The traversal $\mathcal{T}$ used to substitute some leaves in line 13 determines partly the total order over the strict monotonic trees. Let α be an strict monotonic tree, and $\mathcal{T}$ a given traversal of all trees. Remark that there is a single evolution process building α (the construction is unambiguous). If α is built at the step ℓ, then we denote by $\tilde{\alpha}$ the single tree (built with $\ell - 1$ steps) and $\underline{\alpha}$ the single composition such that at step ℓ replacing the leaves from $\tilde{\alpha}$ according to the composition $\underline{\alpha}$, using the traversal $\mathcal{T}$, we obtain α.
Here we remark that the whole tree α is strongly dependent from the traversal of the leaves of $\tilde{\alpha}$ (while some leaves are substituted by an internal nodes attached to new leaves according to $\underline{\alpha}$). We define now how to compare strict monotonic trees (we use the analogous notations than the latter for all trees).

Definition 3.5.2. *Let α and β be two trees. We define $\alpha < \beta$ if*

- *the size of α is smaller than the one of β, or*
- *if both sizes are equal to n and if the size of $\tilde{\alpha}$ is strictly greater than the one of $\tilde{\beta}$ or if both sizes of $\tilde{\alpha}$ and $\tilde{\beta}$ are equal and the composition $\underline{\alpha}$ is smaller than $\underline{\beta}$, using the reverse lexicographic order over compositions.*

Proposition 3.5.3. *The order defined over strict monotonic trees is a total order.*

The result is direct since all possible cases according to the trees α and β for comparing them are explored.

Theorem 3.5.4. *The function* UNRANKTREE *is an unranking algorithm and calling it with the parameters n and a uniformly-sampled integer s in $\{0, \dots, g_n - 1\}$ gives as output a uniform strict monotonic Schröder tree of size n .*

The correctness of the algorithm follows directly from the total order over the trees and Equation (25).

Theorem 3.5.5. *Once the pre-computations have been done, the function* UnrankTree *needs in average $\Theta(n)$ arithmetic operations to construct a tree of size n.*

Proof. Let us assume that all binomial coefficients $(C_{n,\ell})_{0\leq\ell\leq n}$ have been memorized and prove that, with this information stored, the complexity in terms of arithmetic operations is of order $\Theta(n)$. Note that if we only memorize the factorial numbers $(i!)_{0\leq i\leq n-1}$ the complexity is at most three times the complexity obtained when memorizing the binomial coefficient and thus still of order $\Theta(n)$.

For all $n \geq 1$, we denote by a_n the number of arithmetic operations that come from the loop in line 7 and the calls in lines 11 and 12, when building all trees of size n (i.e. we sum the number needed for each $r \in \{0, \ldots, C_{n,\ell} - 1\}$). The exact value of arithmetic operations is $a_n + O(ng_n)$, because at each recursive call there is at most a constant number of operations that are not counted in a_n. We first analyze a_n: we have

$$a_n = \sum_{\ell=1}^{n-1} \binom{n-1}{\ell} \big((\min(\ell, n-1-\ell) - 1 + 2\ell)\, g_{n-\ell} + a_{n-\ell} \big).$$

In fact, for the terms with index ℓ, we are interested in the trees α of size n such that their corresponding tree $\tilde{\alpha}$ is of size $n-\ell$. Thus such trees α are counted by $\binom{n-1}{\ell} g_{n-\ell}$. And for each of them the factor $\min(\ell, n-1-\ell)-1$ is the the number of operations needed for the unranking of the composition (we use the symmetry in the binomial coefficients), the factor 2ℓ is the number of multiplication and subtractions in the loop in line 7. Furthermore we have $a_1 = 0$. By taking an upper bound for the min function, we get that if $\bar{a}_1 = 0$, and, for all $n \geq 2$,

$$\bar{a}_n = \sum_{\ell=1}^{n-1} \binom{n-1}{\ell} (3\ell\, g_{n-\ell} + \bar{a}_{n-\ell}),$$

then $a_n \leq \bar{a}_n$ for all $n \geq 1$. Using similar calculations as in the proof of Proposition 3.2.2, we obtain an equation satisfied by the Borel transform of the series associated to $(\bar{a}_n)$:

$$2\left(\mathcal{B}\bar{A}(z)\right)' = \mathrm{e}^z \left(\mathcal{B}\bar{A}(z)\right)' + 3z\mathrm{e}^z\ \left(\mathcal{B}G(z)\right)'.$$

We thus deduce $\bar{a}_n \sim 3ng_n$, which concludes the proof. □

Let us give some final remarks for this algorithm. In order to obtain a better time complexity for the implementation, we must handle an array of pointers to the leaves of the tree under construction so that the tree traversal is efficient. At each step ℓ, a leaf stored in the array is replaced by $n-\ell+1$ leaves that must be stored in the array. An efficient way consists in reusing the cell from the replaced leaf, and the to append all other leaves at the end on the array. Thus, the most efficient traversal $\mathcal{T}$ of the leaves consists to the left right traversal of the array. But obviously this is not really a natural traversal for the tree. Thus in practice we use this efficient traversal $\mathcal{T}$.

4. Strict monotonic general trees

In this section, we introduce a generalisation of the strict monotonic Schröder tree model of Section 3: the difference is that we allow internal nodes to have only one child (we call these nodes "unary" nodes). Since the size of a tree is the number of its leaves, allowing unary nodes without adding any other constraint would mean that there would be infinitely many trees of any given size n. To avoid this, we add the following constraint: at each growth step, at least one leaf is expanded as an internal node of arity greater or equal to 2.

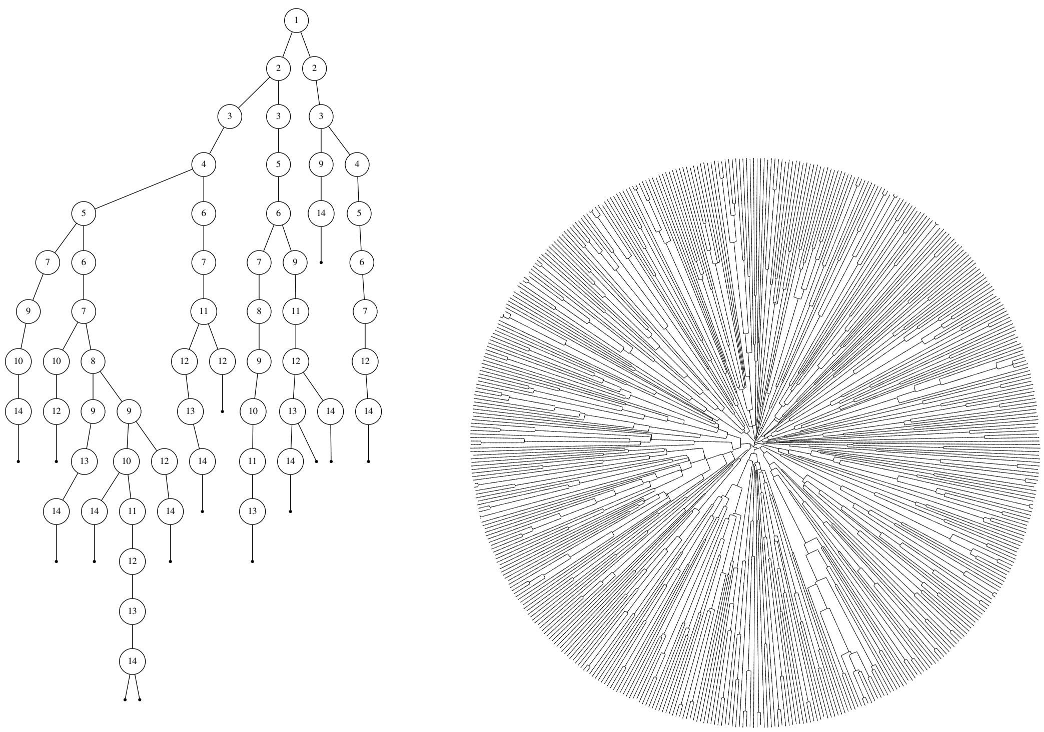


FIGURE 9. Two strict monotonic general trees

4.1. **The model and its enumeration.**

Definition 4.1.1. *A* strict monotonic general tree *is a labelled tree that can be obtained by the following evolution process:*

- *Start with a single (unlabelled) leaf.*
- *At every step $\ell \geq 1$, select a non-empty subset of leaves, replace all of them by internal nodes labelled by ℓ, attach to at least one of them a sequence of at least two leaves, and attach to all others a unique leaf.*

The two trees in Figure 9 are sampled uniformly among all strict monotonic general trees of respective sizes (i.e. number of leaves) 15 and 500. The left-hand-side tree has 14 distinct node-labels, i.e. it can be built in 14 steps using Definition 4.1.1. The right-hand-side tree is represented as a circular tree with stretched edges like in the right-hand-side of Figure 2. Here the tree contains 500 leaves built with 499 iterations of the growth process. But in comparison with the increasing and strict monotonic Schröder trees drawn in the latter sections and containing respectively

492 and 495 internal nodes, this one contains 62494 internal nodes, most of them being unary nodes.

We can specify strict monotonic general trees using the symbolic method; once again the labelling is transparent and does not appear in the specification (i.e. we use ordinary generating functions). In this section, we denote by $F(z)$ the generating function of strict monotonic general trees and by $\mathcal{F}_n$ the set of all strict monotonic general trees of size n; from Definition 4.1.1, we get

$$F(z) = z + F\left(z + \frac{z}{1-z}\right) - F(2z). \tag{27}$$

The combinatorial meaning of this specification is the following: A tree of is either a single leaf, or it is obtained by taking an already constructed tree, and replace each leaf by either a leaf (i.e. no change) or an internal node attached to a sequence of at least one leaf. Furthermore we omit the case where no leaf is replaced by an internal node with at least to children (this is encoded in the subtracting $F(2z)$).

From this equation we extract the recurrence for the number f_n of strict monotonic general trees with n leaves. In fact we get

$$\begin{aligned} f_n = [z^n]F(z) = [z^n]\left(z + F\left(z + \frac{z}{1-z}\right) - F(2z)\right) \\ &= \delta_{n,1} - 2^n f_n + [z^n]\sum_{\ell\geq 1} f_\ell \left(z + \frac{z}{1-z}\right)^\ell \\ &= \delta_{n,1} - 2^n f_n + \sum_{\ell\geq 1} f_\ell\,[z^{n-\ell}]\sum_{i=0}^{\ell}\binom{\ell}{i}\left(\frac{1}{1-z}\right)^i, \end{aligned}$$

which implies that

$$f_n = \begin{cases} 1 & \text{if } n=1, \\ \sum_{\ell=1}^{n-1} \sum\limits_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i}\, 2^{\ell-i}\,\binom{n-\ell-1}{i-1}\, f_\ell & \text{for all } n\geq 2. \end{cases} \tag{28}$$

The combinatorial meaning of the inner sum is the following: starting with a tree of size ℓ we reach a tree of size n in one iteration by adding $n-\ell$ leaves. The index i in the inner sum stands for the number of leaves that are replaced by internal nodes or arity at least 2, by definition of the model (see Definition 4.1.1), we have $1 \leq i \leq \min(n-\ell,\ell)$. There are $\binom{\ell}{i}$ possible choices for the i leaves that are replaced by nodes of arity at least 2. Each of the remaining $\ell - i$ leaves is either kept unchanged or replaced by a unary node, which gives $2^{\ell-i}$ possible choices. And finally, there are $\binom{n-\ell-1}{\ell-1}$ possible ways to distribute the (indistinguishable) $n-\ell$ additional leaves among the i new internal nodes so that each of the i nodes is given at least one additional leaf (it already has one leaf, which is the leaf that was replaced by an internal node). The first terms of the sequence are the following:

$$(f_n)_{n\geq 0} = (0, 1, 1, 5, 66, 2209, 180549, 35024830, 15769748262, 16187601252857, \dots).$$

Theorem 4.1.2. *There exists a constant c such that the number f_n of strict monotonic general trees of size n satisfies, asymptotically when n tends to infinity,*

$$f_n \underset{n\to\infty}{\sim} c\,(n-1)!\,2^{\frac{(n-1)(n-2)}{2}}.$$

In the proof of the latter theorem we exhibit the following bounds $1.4991 < c < 1.8932$. But through several experimentations we see that $c < 3/2$ but it is close to it. For instance when $n = 1000$ we get $c \approx 1.49913911$. We postpone the proof to the next section to make use of the number of iteration steps.

4.2. **Iteration steps and asymptotic enumeration of the trees.** In this section, we look at the number of distinct internal-node labels that occur in a typical strict monotonic general tree, i.e. the number of iterations needed to build it:

Proposition 4.2.1. *Let $f_{n,k}$ denotes the number of strict monotonic general trees of size n with k distinct node-labels, then, for all $n \geq 1$,*

$$f_{n,n-1} = (n-1)!\, 2^{\frac{(n-1)(n-2)}{2}}.$$

Note that the first terms are

$$(f_{n,n-1})_{n\geq 0} = (0, 1, 1, 4, 48, 1536, 122880, 23592960, 10569646080, 10823317585920, \dots).$$

Proof. We use a new variable u to mark the number of iterations (i.e. the number of distinct node-labels) in the iterative Equation (28). We get

$$F(z,u) = z + u\, F\left(z + \frac{z}{1-z}, u\right) - u\, F(2z, u). \tag{29}$$

Using either Equation (29) or a direct combinatorial argument, we get that, for all $k \geq n$, $f_{n,k} = 0$ and

$$f_{n,k} = \begin{cases} 1 & \text{if } n = 1 \text{ and } k = 0, \\ \displaystyle\sum_{\ell=k}^{n-1} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i}\, 2^{\ell-i}\, \binom{n-\ell-1}{i-1}\, f_{\ell,k-1} & \text{if } 1 \leq k < n. \end{cases}$$

In particular, for $k = n-1$, we get

$$\begin{aligned} f_{n,n-1} &= (n-1)\, 2^{n-2}\, f_{n-1,n-2} = f_{1,0} \prod_{j=1}^{n-1} j 2^{j-1} = (n-1)! 2^{\sum_{j=0}^{n-2} j} \\ &= (n-1)! 2^{\frac{(n-1)(n-2)}{2}}, \end{aligned}$$

because $f_{1,0} = 1$. This concludes the proof.

Alternatively the recurrence of $f_{n,n-1}$ can be obtained by extracting the coefficient $[z^n]$ in the following functional equation

$$T(z) = z + z^2\, T'(2z).$$ □

Lemma 4.2.2. *Both sequences (f_n) and $(f_{n,n-1})$ have the same asymptotic behaviour up to a multiplicative constant.*

Proof. Let us start with the definition of a new sequence

$$g_n = \begin{cases} 1 & \text{if } n = 1, \\ f_n / f_{n,n-1} & \text{otherwise.} \end{cases}$$

This sequence g_n satisfies the following recurrence:

$$g_n = \begin{cases} 1 & \text{if } n = 1, \\ \sum_{\ell=1}^{n-1} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i}\, 2^{\ell-i}\, \binom{n-\ell-1}{i-1}\, g_\ell \frac{(\ell-1)!\, 2^{(\ell-1)(\ell-2)/2}}{(n-1)!\, 2^{(n-1)(n-2)/2}} & \text{otherwise.} \end{cases}$$

When $n > 1$, extracting the term g_{n-1} from the sum we get

$$g_n = g_{n-1} + \sum_{\ell=1}^{n-2} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i} 2^{\ell-i} \binom{n-\ell-1}{i-1} g_\ell \frac{(\ell-1)!\, 2^{(\ell-1)(\ell-2)/2}}{(n-1)!\, 2^{(n-1)(n-2)/2}}.$$

Since all summands are non-negative, this implies that $g_n \geq g_{n-1}$, and thus that this sequence is non-decreasing. To prove that this sequence converges, it only remains to prove that it is (upper-)bounded.

Equation (28) implies that, for $n \geq 2$,

$$f_n \leq \sum_{\ell=1}^{n-1} 2^{\ell-1} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i} \binom{n-\ell-1}{i-1} f_\ell.$$

Chu-Vandermonde's identity states that, for all $\ell \leq n$,

$$\sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i} \binom{n-\ell-1}{i-1} = \binom{n-1}{\ell-1}.$$

This implies the following upper-bound for f_n:

$$f_n \leq \sum_{\ell=1}^{n-1} 2^{\ell-1} \binom{n-1}{\ell-1} f_\ell = \sum_{\ell=1}^{n-1} 2^{n-\ell-1} \binom{n-1}{\ell} f_{n-\ell}$$

Using the same argument for g_n we get

$$g_n \leq g_{n-1} + \sum_{\ell=2}^{n-1} \frac{2^{(\ell-1)(\ell-2n+2)/2}}{\ell!} \; g_{n-\ell}.$$

We look at the exponent of 1 in the sum: For all $\ell \geq 2$ (as in the sum), we have $2\ell \geq \ell + 2$, and thus $2n - \ell - 2 \geq 2(n-\ell)$. This implies that for all $\ell \geq 2$, $(\ell-1)(\ell-2n+2)/2 \leq -(n-\ell)$, and thus that

$$g_n \leq g_{n-1} + \sum_{\ell=2}^{n-1} \frac{1}{\ell!\, 2^{n-\ell}} \; g_{n-\ell}.$$

Since the sequence $(g_n)_n$ is non-decreasing, we obtain

$$g_n \leq g_{n-1} + \frac{g_{n-1}}{2^n} \sum_{\ell=2}^{n-1} \frac{2^\ell}{\ell!} \leq g_{n-1} + g_{n-1} \frac{\mathrm{e}^2 - 3}{2^n}.$$

We set $\alpha = \mathrm{e}^2 - 3$. Iterating the last inequality, we get that

$$g_n \leq g_{n-1} \left(1 + \frac{\alpha}{2^n}\right) \leq g_1 \prod_{i=2}^{n} \left(1 + \frac{\alpha}{2^i}\right) = \exp\left(\sum_{i=2}^{n} \ln\left(1 + \frac{\alpha}{2^i}\right)\right),$$

because $g_1 = 1$. Note that, when $i \to +\infty$, we have $\ln(1 + \alpha 2^{-i}) \leq \alpha 2^{-i}$ (because $\ln(1+x) \leq x$ for all $x \geq 0$). This implies that, for all $n \geq 1$,

$$g_n \leq \exp\left(\alpha \sum_{i=2}^{\infty} 2^{-i}\right) = \exp(\alpha/2).$$

In other words, the sequence $(g_n)_n$ is bounded. Since it is also non-decreasing, it converges to a finite limit c, which is also non-zero since $g_n \geq g_1 \neq 0$ for all $n \geq 1$. This is equivalent to $f_n \sim c f_{n,n-1}$ when $n \to +\infty$ as claimed. To get a lower wound on c, note that, for all $n \geq 1$, $c \geq g_n \geq g_9 = f_9/f_{9,8} \approx 1.4956$. $\square$

Proof of Theorem 4.1.2. The latter Lemma 4.2.2 gives a proof of Theorem 4.1.2. But in order to get a better upper bound for the constant c, let us introduce another proof. In the proof of Lemma 4.2.2 we have proved

$$g_n \leq g_{n-1} + g_{n-1}\frac{\mathrm{e}^2 - 3}{2^n}.$$

We set $\alpha = \mathrm{e}^2 - 3$ and define two other sequences as

$$\bar{g}_n = \begin{cases} 1 & \text{if } n = 1 \text{ or } n = 2, \\ \bar{g}_{n-1} + \frac{\alpha}{2^n}\, \bar{g}_{n-2} & \text{otherwise,} \end{cases}$$

and

$$\bar{\bar{g}}_n = \begin{cases} 1 & \text{if } n = 1 \text{ or } n = 2, \\ \bar{\bar{g}}_{n-1} + \frac{1}{n(n+1)}\, \bar{\bar{g}}_{n-2} & \text{otherwise.} \end{cases}$$

Due to the two first terms and the recurrence equation we have for all positive n, $g_n \leq \bar{g}_n \leq \bar{\bar{g}}_n$. By induction we prove a new expression for $\bar{\bar{g}}_n$:

$$\bar{\bar{g}}_n = \begin{cases} \bar{g}_n & \text{if } n \leq 3, \\ \bar{\bar{g}}_{n-1} + \frac{2}{(n+1)!}\, a_{n-1} & \text{otherwise,} \end{cases}$$

with the sequence $(a_n)_n$ such that $a_1 = 0, a_2 = 1$ and for $n \geq 3$, $a_n = na_{n-1} + a_{n-2}$. This sequence (a_n) is a shifted version of `OEIS A058307`. We can either follow the work of Janson [Jan10] to study it, but we need less details than him so we describe here an easier approach. We define a new sequence as $b_n = a_n/n!$. We easily prove that $b_n = b_{n-1} + b_{n-2}/(n(n-1))$ with $b_1 = 0$ and $b_2 = 1/2$. Using the later recurrence, we obtain an equation satisfied by its generating function $B(z) = \sum_{n>0} b_n z^n$:

$$B(z) = \frac{z^2}{2} + zB(z) + \int_{t=0}^{u}\int_{t=0}^{z} B(u)\mathrm{d}u.$$

we thus obtain

$$(z-1)B''(z) + 2B'(z) + B(z) + 1 = 0,$$

with $B(0) = 0$ and $B'(0) = 0$. By dividing the equation by $\mathrm{i}\sqrt{1-z}$ and then by a change of variable: $u := 2\mathrm{i}\sqrt{1-z}$, we recognize the classical differential equation satisfied by Bessel functions [BO99]. We thus derive

$$B(z) = -1 + \frac{1}{\sqrt{1-z}}\left(\alpha\, J_1(2\mathrm{i}\sqrt{1-z}) + \beta\, Y_1(2\mathrm{i}\sqrt{1-z})\right),$$

where $J_.(\cdot)$ and $Y_.(\cdot)$ are the Bessel functions and α and β are two complex constants determined with the initial conditions:

$$\alpha = \frac{Y_1(2\mathrm{i}) - \mathrm{i}Y_0(2\mathrm{i})}{J_1(2)Y_0(2\mathrm{i}) + \mathrm{i}J_0(2)Y_1(2\mathrm{i})}, \qquad \beta = -\frac{J_1(2) - \mathrm{i}J_0(2)}{J_1(2)Y_0(2\mathrm{i}) + \mathrm{i}J_0(2)Y_1(2\mathrm{i})}.$$

We are interested in the asymptotic behaviour of b_n. The dominant singularity of $B(z)$ is at $z = 1$ and there

$$B(z) \underset{z\to 1}{\sim} -\frac{\beta}{\mathrm{i}\pi}\,\frac{1}{1-z}.$$

We thus deduce that b_n tends to $-\beta/(\mathrm{i}\pi) \approx 0.68894$. Since the sequence $\bar{\bar{g}}_n$ satisfies $\bar{\bar{g}}_n = \bar{\bar{g}}_{n-1} + \frac{2}{n(n+1)}\, b_{n-1}$. We deduce that the increasing sequence $(\bar{\bar{g}}_n)$ admits a finite limit. Hence it is also the case for the increasing sequence (g_n). Finally, Proposition 4.2.1 allows to conclude for the existence of the constant c. Furthermore we get

$$c < \bar{\bar{g}}_3 + \sum_{\ell \geq 4} \frac{2}{\ell(\ell+1)} \cdot \lim_{n\to\infty} b_n \approx 1.8932. \qquad \square$$

This result means that asymptotically a constant fraction of the strict monotonic general trees of size n are built in $(n-1)$ steps. For these trees, at each step of construction only one single leaf expands into a binary node. All other leaves either become a unary node or stay unchanged, meaning that on average half of the leaves will expand into unary node with one leaf expanding into a binary node. The number of internal nodes of these trees then grow like $n^2/4$.

4.3. **Analysis of typical parameters.**

4.3.1. *Quantitative analysis of the number of internal nodes.*

Theorem 4.3.1. *Let $I_n^{\mathcal{F}}$ be the number of internal nodes in a tree taken uniformly at random among all strict monotonic general trees of size n. Then for all $n \geq 1$, we have*

$$\frac{(n-1)(n+2)}{6} \leq \mathbb{E}[I_n^{\mathcal{F}}] \leq \frac{(n-1)n}{2}.$$

To prove this theorem, we use the following proposition.

Proposition 4.3.2. *Let us denote by $s_{n,k}$ the number of strict monotonic general trees of size n that have $n-1$ distinct node-labels and k internal nodes. For all $n \geq 1$ and $k \geq 0$,*

$$s_{n,k} = (n-1)! \binom{(n-1)(n-2)/2}{k-(n-1)},$$

and thus, if $I_n^{\mathcal{S}}$ is the number of internal nodes in a tree taken uniformly at random among all strict monotonic general trees of size n that have $n-1$ distinct label nodes, then, for all $n \geq 1$,

$$\mathbb{E}[I_n^{\mathcal{S}}] = \frac{(n-1)(n+2)}{4}.$$

Proof. Let us prove the formula for $s_{n,k}$ by induction. For $n=1$, k can only be 0 thus $s_{1,0} = 1 = 0!\binom{0}{0}$.

We suppose that $s_{m,k} = (m-1)!\binom{(m-1)(m-2)/2}{k-(m-1)}$ holds for $m = n-1$ and $k \in \{n-1, \dots, (n-2)(n-3)/2\}$.

Then, we are interested in the value of $s_{n,k}$:

$$s_{n,k} = \sum_{s=0}^{k-(n-1)} (n-2)! \binom{(n-2)(n-3)/2}{s-(n-2)} \binom{n-1}{k-s-1} (k-s-1).$$

Let $k' = k-(n-1)$ and $s' = s-(n-2)$. Replacing k' and s' in the equation gives,

$$\begin{aligned}
\tilde{s}_{n,k'} &= \sum_{s'=0}^{k'} (n-2)! \binom{(n-2)(n-3)/2}{s'} \binom{n-1}{k'-s'+1} (k'-s'+1) \\
&= (n-1)! \sum_{s'=0}^{k'} \binom{(n-2)(n-3)/2}{s'} \binom{n-2}{k'-s'}.
\end{aligned}$$

Using Chu-Vandermonde identity, we finally obtain

$$s_{n,k} = (n-1)!\binom{(n-1)(n-2)/2}{k-(n-1)}.$$

We now can compute the average number of internal nodes of $\mathcal{S}_n$:

$$\mathbb{E}_n[I_n^{\mathcal{S}}] = \frac{\sum_{k=n-1}^{n(n-1)/2} k(n-1)!\binom{(n-1)(n-2)/2}{k-(n-1)}}{(n-1)!\,2^{(n-1)(n-2)/2}}.$$

Again we reverse the sum: $k' = k-(n-1)$,

$$\begin{aligned}\mathbb{E}[I_n^{\mathcal{S}}] &= \frac{\sum_{k'=0}^{(n-1)(n-2)/2} (k'+(n-1))(n-1)!\binom{(n-1)(n-2)/2}{k'}}{(n-1)!\,2^{(n-1)(n-2)/2}}\\ &= \frac{\sum_{k'=0}^{(n-1)(n-2)/2} k'\binom{(n-1)(n-2)/2}{k'} + (n-1)\sum_{k'=0}^{(n-1)(n-2)/2}\binom{(n-1)(n-2)/2}{k'}}{2^{(n-1)(n-2)/2}}\\ &= \frac{(n-1)(n-2)}{4} + (n-1) = \frac{(n-1)(n+2)}{4}.\end{aligned}$$

□

We are now ready to prove the main theorem of this section.

Proof of Theorem 4.3.1. Note that the number of internal nodes of a strict monotonic general tree of size n belongs to $\{1,\dots,n(n-1)/2\}$. The upper bound follows from the fact that, at the ℓ-th iteration in Definition 4.1.1, a maximum of ℓ internal nodes is added to the tree, and $\sum_{\ell=1}^{n}\ell = n(n-1)/2$. In particular, we thus have that, almost surely for all $n\geq 1$, $I_n^{\mathcal{F}} \leq n(n-1)/2$, and thus $\mathbb{E}[I_n^{\mathcal{F}}] = \mathcal{O}(n^2)$.

For the lower bound, we denote by $\mathcal{S}_n$ the set of strict monotonic general trees of size n that have $n-1$ distinct node-labels. Moreover, we denote by t_n a tree taken uniformly at random in $\mathcal{F}_n$, and by $I_n^{\mathcal{F}}$ its number of internal nodes. We have, for all $n\geq 1$,

$$\begin{aligned}\mathbb{E}[I_n^{\mathcal{F}}] &= \mathbb{E}[I_n^{\mathcal{F}} \mid t_n\in\mathcal{S}_n]\cdot\mathbb{P}(t_n\in\mathcal{S}_n) + \mathbb{E}[I_n^{\mathcal{F}} \mid t_n\notin\mathcal{S}_n]\cdot\mathbb{P}(t_n\notin\mathcal{S}_n)\\ &\geq \mathbb{E}[I_n^{\mathcal{F}} \mid t_n\in\mathcal{S}_n]\cdot\mathbb{P}(t_n\in\mathcal{S}_n) = \mathbb{E}[I_n^{\mathcal{S}}]\ \cdot\frac{f_{n,n-1}}{f_n},\end{aligned}$$

where we have used conditional expectations and the fact that conditionally on being in $\mathcal{S}_n$, t_n is uniformly distributed in this set, and, in particular, $\mathbb{E}[I_n^{\mathcal{F}} \mid t_n \in \mathcal{S}_n] = \mathbb{E}I_n^{\mathcal{S}}$. Using Proposition 4.3.2 and the upper bound of Proposition 4.2.1, we thus get

$$\mathbb{E}[I_n^{\mathcal{F}}] \geq \frac{2}{3}\ \frac{(n-1)(n+2)}{4},$$

which concludes the proof. □

4.3.2. *Quantitative analysis of the number of distinct labels.*

Theorem 4.3.3. *Let $X_n^{\mathcal{F}}$ denotes the number of distinct internal-node labels (or construction steps) is a tree taken uniformly at random among all strict monotonic general trees of size n, then for all $n\geq 1$,*

$$\frac{2}{3}\ (n-1) \leq \mathbb{E}[X_n^{\mathcal{F}}] \leq n-1.$$

Proof. First note that since at every construction step in Definition 4.1.1 we add at least one leaf in the tree, then after ℓ construction steps, there are exactly ℓ distinct labels and at least $\ell+1$ leaves in the tree. Therefore, $n \geq X_n^{\mathcal{F}}+1$ almost surely for all $n \geq 1$, which implies in particular that $\mathbb{E}[X_n] \leq n-1$, as claimed.

For the lower bound, we reason as in the proof of Theorem 4.3.1, and using the same notations:

$$\mathbb{E}[X_n^{\mathcal{F}}] \geq \mathbb{E}[X_n^{\mathcal{F}} \mid t_n \in \mathcal{S}_n] \cdot \mathbb{P}(t_n \in \mathcal{S}_n) = (n-1)\,\frac{f_{n,n-1}}{f_n},$$

because $\mathbb{E}[X_n^{\mathcal{F}} \mid t_n \in \mathcal{S}_n] = n-1$ by definition of $\mathcal{S}_n$ (being the set of all strict monotonic general trees of size n that have $n-1$ distinct node-labels). Using the upper bound of Proposition 4.2.1 gives that $\mathbb{E}[X_n^{\mathcal{F}}] \geq 2(n-1)/3$, which concludes the proof. □

4.3.3. *Quantitative analysis of the height of the trees.*

Theorem 4.3.4. *Let $H_n^{\mathcal{F}}$ denotes the height of a tree taken uniformly at random in $\mathcal{F}_n$, the set of all strict monotonic general trees of size n. Then we have, for all $n \geq 0$,*

$$\frac{n}{3} \leq \mathbb{E}[H_n^{\mathcal{F}}] \leq n-1.$$

To prove this theorem, we first prove the following:

Proposition 4.3.5. *Let us denote by $H_n^{\mathcal{S}}$ the height of a tree taken uniformly at random in $\mathcal{S}_n$, the set of all strict monotonic general trees of size n that have $n-1$ distinct labels. Then we have, for all $n \geq 0$,*

$$\frac{n}{2} \leq \mathbb{E}[H_n^{\mathcal{S}}] \leq n-1.$$

Proof. Define the sequence of random trees $(t_n)_{n\geq 0}$ recursively as:
- t_1 is a single leaf.
- Given t_{n-1}, we define t_n as the tree obtained by choosing a leaf uniformly at random among all leaves of t_{n-1}, replacing it by an internal nodes to which two leaves are attached, and, for each of the other leaves of t_{n-1}, choose with probability 1/2 (independently from the rest) whether to leave it unchanged or to replace it by a unary node to which one leaf is attached.

One can prove by induction on n that for all $n \geq 1$, t_n is uniformly distributed in $\mathcal{S}_n$. We denote by $H_n^{\mathcal{F}}$ the height of t_n. Since the height of t_n is at most the height of t_{n-1} plus 1 for all $n \geq 2$, we get that $H_n^{\mathcal{S}} \leq n-1$ almost surely.

For the upper bound, we note that, for the height of t_n to be larger than the height of t_{n-1}, we need to have replaced at least one of the maximal-height leaves in t_{n-1}. There is at least one leaf of t_{n-1} which is at height $H_{n-1}^{\mathcal{S}}$ and this leaf is replaced by an internal node with probability

$$\frac{1}{2}\left(1-\frac{1}{n-1}\right)+\frac{1}{n-1} \geq \frac{1}{2}.$$

Therefore, for all $n \geq 1$, we have

$$\mathbb{P}(H_n^{\mathcal{S}} = H_{n-1}^{\mathcal{S}}+1) \geq \frac{1}{2},$$

which implies, since $H_n^{\mathcal{S}} \in \{H_{n-1}^{\mathcal{S}}, H_{n-1}^{\mathcal{S}}+1\}$ almost surely,

$$\mathbb{E}[H_n^{\mathcal{S}}] = \mathbb{E}[H_{n-1}^{\mathcal{S}}] + \mathbb{P}(H_n^{\mathcal{S}} = H_{n-1}^{\mathcal{S}}+1) \geq \mathbb{E}[H_{n-1}^{\mathcal{S}}] + \frac{1}{2}.$$

Therefore, for all $n \geq 1$, we have $\mathbb{E}[H_n^{\mathcal{S}}] \geq \mathbb{E}[H_0^{\mathcal{S}}] + n/2 = n/2$, as claimed. $\square$

Proof of Theorem 4.3.4. By Definition 4.1.1, it is straightforward to see that the height of a tree built in ℓ steps is at most ℓ since the height increases by at most one per construction step. Since a tree of size n is built in at most $n-1$ steps, we get that $H_n^{\mathcal{F}} \leq n-1$ almost surely, which implies, in particular, that $\mathbb{E}[H_n^{\mathcal{F}}] \leq n-1$.

For the lower bound, note that, if t_n is a tree taken uniformly at random in $\mathcal{F}_n$ and $H_n^{\mathcal{F}}$ is its height, then

$$\mathbb{E}[H_n^{\mathcal{F}}] \geq \mathbb{E}[H_n^{\mathcal{F}} | t_n \in \mathcal{S}_n] \cdot \mathbb{P}(X \in \mathcal{S}_n) \geq \frac{2}{3}\, \mathbb{E}[H_n^{\mathcal{S}}],$$

where we have used Proposition 4.2.1 and the fact that t_n conditioned on being in $\mathcal{S}_n$ is uniformly distributed in this set and thus $E[H_n^{\mathcal{F}} \mid t_n \in \mathcal{S}_n] = \mathbb{E}H_n^{\mathcal{S}}$. By Proposition 4.3.5, we thus get $\mathbb{E}[H_n^{\mathcal{F}}] \geq n/3$, as claimed. $\square$

4.3.4. *Quantitative analysis of the depth of the leftmost leaf.*

Theorem 4.3.6. *Let us denote by $D_n^{\mathcal{F}}$ the height of a tree taken uniformly at random in $\mathcal{F}_n$, the set of all strict monotonic general trees of size n. Then we have, for all $n \geq 0$,*

$$\frac{n}{3} \leq \mathbb{E}[H_n^{\mathcal{F}}] \leq n-1.$$

Proposition 4.3.7. *Let us denote by $D_n^{\mathcal{S}}$ the depth of the leftmost leaf of a tree taken uniformly at random in $\mathcal{S}_n$, the set of all strict monotonic general trees of size n that have $n-1$ distinct labels. Then we have, for all $n \geq 0$,*

$$\frac{n}{2} \leq \mathbb{E}[D_n^{\mathcal{S}}] \leq n-1.$$

Proof. Given the uniform process of trees t_n presented in 4.3.5. The depth of the leftmost leaf is always smaller than $n-1$. Let X_n be a Bernoulli variable taking value 1 if the leftmost leaf of t_n has been expanded at iteration n and the value 0 otherwise. Then for $n \geq 1$,

$$\mathbb{P}(X_n = 1) = \frac{1}{n} + \frac{(n-1)}{n}\frac{1}{2} = \frac{n+1}{2n} \geq \frac{1}{2}.$$

Since at each iteration step either the leftmost leaf expand to make a binary node which gives $\frac{1}{n}$ or it has not created a binary and then it has $\frac{1}{2}$ probability to make a unary node. The depth of the leftmost leaf is $D_n^{\mathcal{S}} = \sum_{k=1}^{n} X_k$. Therefore for $n \geq 1$,

$$\mathbb{E}[D_n^{\mathcal{S}}] \geq \frac{n}{2}.$$

Which concludes the proof. $\square$

Proof of Theorem 4.3.6. By the same arguments as in Theorem 4.3.4 the result follows directly since we have the same bounds on the depth of leftmost leaf as we had in the height of the tree. $\square$

4.4. **Correspondence with labelled graphs.** In Section 4.2 we defined $f_{n,k}$ the number of strict monotonic general trees of size n that have k distinct node-labels then we have shown that, for all $n \geq 1$,

$$f_{n,n-1} = (n-1)!\, 2^{\frac{(n-1)(n-2)}{2}}.$$

The factor $2^{(n-1)(n-2)/2} = 2^{\binom{n-1}{2}}$ in graphs of $(n-1)$ vertices counts the different combinations of edges (not directed) between vertices. The factor $(n-1)!$ accounts for all possible permutations of vertices. We will denote $\mathcal{S}_n$ to be the trees that $f_{n,n-1}$ counts and exhibit a bijection between strict monotonic general trees of $\mathcal{S} = \cup_{n\geq 1}\mathcal{S}_n$ with a class of labelled graphs with $n-1$ vertices defined in the following. Let us define the subclass of strict monotonic general trees $\mathcal{S} = \cup_{n\geq 1}\mathcal{S}_n$.

For all $n \geq 1$, we denote by $\mathcal{G}_n$ the set of all labelled graphs (V, ℓ, E) such that $V = \{1, \dots, n\}$, $E \subseteq \{\{i, j\} \colon i \neq j \in V\}$ and $\ell = (\ell_1, \dots, \ell_n)$ is a permutation of V (see Figure 10 for an example). We set $\mathcal{G} = \cup_{n=0}^{\infty}\mathcal{G}_n$. Choosing a graph in $\mathcal{G}_n$ is equivalent to (1) choosing ℓ (there are $n!$ choices) and (2) for each of the $\binom{n}{2}$ possible edges, choose whether it belongs to E or not (there are $2^{\binom{n}{2}}$ choices in total). In total, we thus get that $|\mathcal{G}_n| = n!\, 2^{\binom{n}{2}}$.

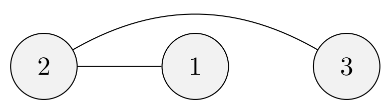


FIGURE 10. The graph $\mathcal{G}_3$ graph. In this representation, the vertices $V = \{1, \dots, n\}$ are drawn from left to right (node 1 is the leftmost, node n is the rightmost), and their label is their image by ℓ: in this example $\ell = (2, 1, 3)$.

We recall the definitions used in Section 2.4. A size-n permutation σ is denoted by $(\sigma_1, \dots, \sigma_n)$, and σ_i is its i-th element (the image of i), while $\sigma^{-1}(k)$ is the preimage of k (the position of k in the permutation).

Another important bijection that we will use is the bijection between binary increasing trees and permutations, see [FS09, page143].

We define $\mathcal{M}'' : \mathcal{S} \to \mathcal{G}$ recursively on the size of the tree it takes as an input: first, if t is the tree of size 1 (which contains only one leaf) or the tree of size 2 (one internal node attached to two leaves), then we set $\mathcal{M}''(t)$ to be the graph $(\{1\}, (1), \varnothing)$ (the graph with one vertex labelled 1 and no edge). Now assume we have defined $\mathcal{M}''$ on $\cup_{\ell=1}^{n-1}\mathcal{S}_\ell$, and consider a tree $t \in \mathcal{S}_n$. By Definition 4.1.1 and since $t \in \mathcal{S}_n$, then there exists a unique binary node in t labelled by $n-1$, and this node is attached to two leaves. Consider $\hat{t}$ the tree obtained when removing all internal nodes labelled by $n-1$ (and all the leaves attached to them) from t and replacing them by leaves. Denote by v_n the position (in, e.g., depth-first order) of the leaf of $\hat{t}$ that previously contained the binary node labelled by $n-1$ in t. Denote by $u_1, \dots, u_m$ the positions or the leaves of $\hat{t}$ that previously contained unary nodes labelled by $n-1$ in t. We set $\mathcal{M}''(\hat{t}) = (\{1, \dots, n-1\}, \hat{\ell}, \hat{E})$ and define

$\mathcal{M}''(t) = (\{1,\dots,n\}, \ell, E)$ where

$$\ell_i = \begin{cases} v_n & \text{if } i = n \\ \hat{\ell}_i & \text{if } \hat{\ell}_i < v_n \\ \hat{\ell}_i + 1 & \text{if } \hat{\ell}_i \geq v_n, \end{cases}$$

$E = \hat{E} \cup \{\{\hat{\ell}^{-1}(u_j), n\} : 1 \leq j \leq m\}$. An example of the bijection is depicted in Figure 11.

Theorem 4.4.1. *The mapping $\mathcal{M}''$ is bijective, and $\mathcal{M}''(\mathcal{S}_n) = \mathcal{G}_{n-1}$.*

Proof. From the definition, it si clear that two different trees have two distinct images by $\mathcal{M}''$, thus implying that $\mathcal{M}''$ is injective; this is enough to conclude since $|\mathcal{G}_{n-1}| = |\mathcal{S}_n|$ (see Theorem 4.2.1 for the cardinality of $\mathcal{S}_n$). □

*R*emark: It is interesting to note that this graph model is a labelled version of the binomial random graph $\mathcal{G}_n(1/2) = (V, E)$ defined as follows: $V = \{1,\dots,n\}$ and each edge belong to E with probability 1/2, independently from the other edges. This model, also called the Erdös-Renyi random graph was originally introduced by Erdös and Renyi [ER59], and simultaneously by Gilbert [Gil59], and has been since then extensively studied in the probability and combinatorics literature (see, for example, the books [Bol01] and [Dur06] for introductory surveys).

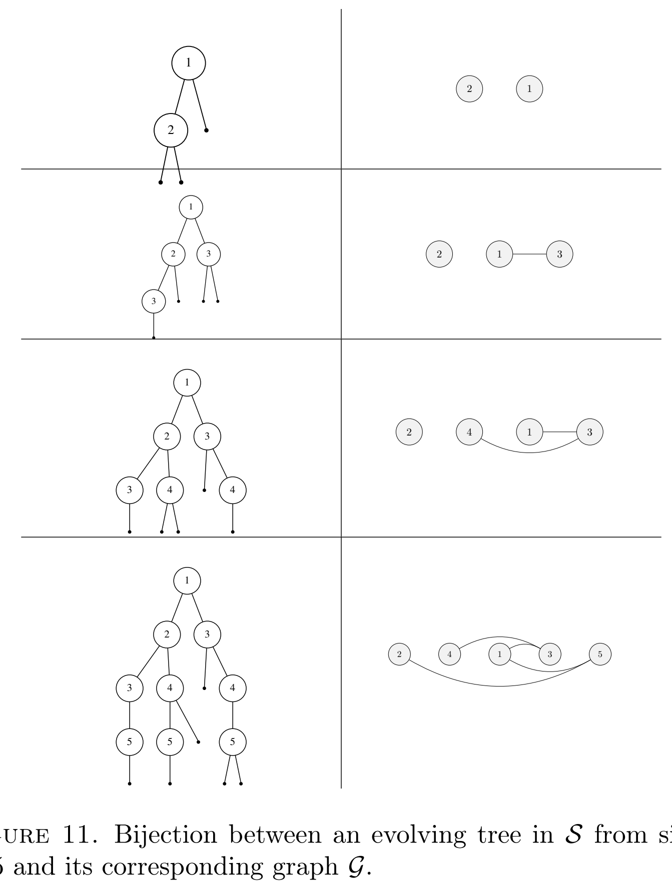


FIGURE 11. Bijection between an evolving tree in $\mathcal{S}$ from size 3 to 5 and its corresponding graph $\mathcal{G}$.

4.5. **Uniform random sampling.** In this section we exhibit a very efficient way for the uniform sampling of the tree model using the evolution process.

Once again when r grows, the sequence $(f_{n-r})_r$ decreases extremely fast. Thus for the uniform random sampling, it will appear more efficient to read Equation (28) in the following way:

$$\begin{aligned} f_n = &\binom{n-1}{1} 2^{n-2} f_{n-1} + \sum_{i=1}^{2} \binom{n-2}{i} 2^{n-2-i} f_{n-2} \\ &+ \sum_{i=1}^{3} \binom{n-3}{i} 2^{n-3-i} \binom{2}{i-1} f_{n-3} + \cdots + f_1. \end{aligned} \tag{30}$$

Using the latter decomposition the algorithm can now be described as Algorithm 4.

Algorithm 4 Strict Monotonic General Tree Unranking

1: **function** UnrankTree(n, s)
2:     **if** $n = 1$ **then**
3:         **return** the tree reduced to a single leaf
4:     $\ell := 1$
5:     $r := s$
6:     $i := 1$
7:     **while** $r >= 0$ **do**
8:         $t := \binom{n-\ell}{i} 2^{n-\ell-i} \binom{\ell-1}{i-1}$
9:         $r := r - t \cdot f_{n-\ell}$
10:         $i := i + 1$
11:         **if** $i > min(\ell, n-\ell)$ **then**
12:             $i := 1$
13:             $\ell := \ell + 1$
14:     **if** $i > 1$ **then**
15:         $i := i - 1$
16:     **else**
17:         $\ell := \ell - 1$
18:         $i := min(\ell, n-\ell)$
19:     $r := r + t \cdot f_{n-\ell}$
20:     $T :=$UnrankTree$(n-\ell, r \mod f_{n-\ell})$
21:     $r := r // f_{n-\ell}$
22:     $B :=$UnrankBinomial$(n-\ell, i, r // \binom{n-\ell}{i})$
23:     $r := r \mod \binom{n-\ell}{i}$
24:     $F := r // \binom{\ell-1}{i-1}$
25:     $C :=$UnrankComposition$(\ell, i, r \mod \binom{\ell-1}{i-1})$
26:     Substitute in T, using traversal $\mathcal{T}$, the leaves selected with B with internal nodes
27:     and new leaves according to C and the other leaves are changed or not based on F
28:     **return** the tree T

The sequences $(f_\ell)_{\ell \le n}$ and $(\ell!)_{\ell \in \{1,\dots,n\}}$ have been pre-computed and stored.

Algorithm 4 is very similar to the one corresponding to strict monotonic trees. In fact this new one is a bit more involved than the previous one because of the recurrence formula for enumerating the tree. However both algorithm cores are very close. First the `While` loop allows to determine the values for ℓ, i and r. Then the recursive call is done using the adequate rank $r \mod f_{n-\ell}$. The last lines of the algorithm (for 21 to 27) are necessary to modify the tree T of size $n - \ell$ that has just been built. In line 22 we determine which leaves T will be substituted by internal nodes (of arity at most 2) with new leaves. It is based on the unranking of combinations that is very close to the unranking of compositions. Then for the other leaves that are either kept as they are of replaced by unary internal nodes attached to a leaf we use the integer F seen as a $n - \ell - i$-bit integer: if the bit $\#s$ is 0 then the corresponding leaf is kept, and if it is 1 then the leaf si substituted. And finally the composition unranking allows to determine how many leaves are attached to the nodes selected with B.

Theorem 4.5.1. *The function* UnrankTree *is an unranking algorithm and calling it with the parameters n and a uniformly-sampled integer s in $\{0, \dots, f_n - 1\}$ gives as output a uniform strict monotonic general tree of size n .*

The correctness of the algorithm follows directly from the total order over the trees deduced from the decomposition (30).

Theorem 4.5.2. *Once the pre-computations have been done, the function* UnrankTree *needs in average $\Theta(n)$ arithmetic operations to construct a tree of size n.*

The proof for this theorem is analogous to the one for Theorem 3.5.5 corresponding to the complexity of the tree builder for strict monotonic Schröder trees.

5. Conclusion

As a conclusion, we comment our main analytical results (summarised in Table 1) in the light of the simulations obtained using the different random samplers designed in the paper (see the right-hand sides of Figures 2, 5 and 9), and compare on the similarities and difference of our three models. Recall that in the representations no label is represented but the length of an edge between two internal nodes is proportional to the difference of the labels of the nodes it connects.

A few of our analytical results can be observed looking at the simulations: for example, the fact that a large proportion of the nodes are binary in a large monotonic Schröder tree, which we have confirmed by a rigorous analysis (see Theorem 2.3.6), is visible on Figure 2. From Figure 5, one could conjecture this is also true in the case of strict monotonic Schröder trees, but this question remains open.

From Figures 2, 5 and 9 it seems clear that the model of strict monotonic general Schröder trees behaves drastically differently from the two other models, which are quite similar. This is indeed what we have proved in our analysis: for example, the height of a typical strict monotonic general tree of size n is of order $\Theta(n)$ (see Theorem 4.3.4), while we have shown that a in the monotonic case, the height is of order $\Theta(\log n)$ (see Theorem 2.6.1). Another huge difference is that the number of internal nodes in a large typical monotonic general Schröder tree is of order $\Theta(n^2)$ (see Theorem 4.3.1) while, in the two other models, this parameter is of order n (see Theorems 2.3.1 and 3.4.2).

Proving results on the height of different families of random trees is often a challenging question, and we have seen that it is indeed one of the most intricate parameters to study in our three models: in the case of monotonic and strict monotonic general trees, we obtain a Θ-estimate but we only obtain a $\ln n$ lower bound in the case of strict monotonic Schröder trees (see Proposition 3.4.4). A natural conjecture, based on the fact that monotonic and strictly monotonic Schröder trees seem to behave similarly, is that the height of a typical strictly monotonic Schröder tree is also of order $\Theta(\ln n)$.

Acknowledgment. We thank Hsien-Kuei Hwang for his help with the proof of Theorem 2.3.6 and for telling us about the links of this proof with [CHY00].

References

[Ald96] D. Aldous. Probability distributions on cladograms. In D. Aldous and R. Pemantle, editors, *Random Discrete Structures*, pages 1–18. Springer New York, 1996.

[BBHT17] A. Bacher, O. Bodini, H.-K. Hwang, and T.-H. Tsai. Generating random permutations by coin tossing: Classical algorithms, new analysis, and modern implementation. *ACM Trans. Algorithms*, 13(2):24:1–24:43, 2017.

[BDF+16] O. Bodini, M. Dien, X. Fontaine, A. Genitrini, and H.-K. Hwang. Increasing Diamonds. In *LATIN 2016: Theoretical Informatics - 12th Latin American Symposium*, pages 207–219, 2016.

[BDGV18] O. Bodini, M. Dien, A. Genitrini, and A. Viola. Beyond series-parallel concurrent systems: the case of arch processes. In *29th International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, page to appear, 2018.

[Ben73] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A*, 15(1):91–111, 1973.

[BFS92] F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP*, pages 24–48, 1992.

[BGGW20] O. Bodini, A. Genitrini, B. Gittenberger, and S. Wagner. On the number of increasing trees with label repetitions. *Discrete Mathematics*, 343(8):111722, 2020.

[BGN19] O. Bodini, A. Genitrini, and M. Naima. Ranked schröder trees. In *Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2019*, pages 13–26, 2019.

[BGNS20] O. Bodini, A. Genitrini, M. Naima, and A. Singh. Families of Monotonic Trees: Combinatorial Enumeration and Asymptotics. In *15th International Computer Science Symposium in Russia (CSR)*, page To appear, 2020.

[BL77] B. P. Buckles and M. Lybanon. Algorithm 515: Generation of a vector from the lexicographical index [g6]. *ACM Trans. Math. Softw.*, 3(2):180–182, June 1977.

[BO99] C. Bender and S. Orszag. *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory*, volume 1. 1999.

[Bol01] B. Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.

[BRS12] O. Bodini, O. Roussel, and M. Soria. Boltzmann samplers for first-order differential specifications. *Discrete Applied Mathematics*, 160(18):2563–2572, 2012.

[CHY00] H.-H. Chern, H.-K. Hwang, and Y.-N. Yeh. Distribution of the number of consecutive records. *Random Structures & Algorithms*, 17(3-4):169–196, 2000.

[Dev90] L. Devroye. On the height of random m-ary search trees. *Random Struct. Algorithms*, 1(2):191–204, 1990.

[DGH20] C. Donnot, A. Genitrini, and Y. Herida. Unranking combinations lexicographically: an efficient new strategy compared with others. https://hal.archives-ouvertes.fr/hal-02462764v1, 2020.

[Drm09] M. Drmota. *Random trees*. Springer, Vienna-New York, 2009.

[Dur06] R. Durrett. *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2006.

[ER59] P. Erdös and A. Rényi. On random graphs. *Publicationes mathematicae*, 6(26):290–297, 1959.

[FS09] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[FZVC94] P. Flajolet, P. Zimmermann, and B. Van Cutsem. A calculus for the random generation of labelled combinatorial structures. *Theoretical Computer Science*, 132(1-2):1–35, 1994.

[GGKW20] A. Genitrini, B. Gittenberger, M. Kauers, and M. Wallner. Asymptotic enumeration of compacted binary trees of bounded right height. *to appear in Journal of Combinatorial Theory, Series A*, 172:105177, 2020.

[Gil59] E.Ñ. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[Hwa98] H.-K. Hwang. On Convergence Rates in the Central Limit Theorems for Combinatorial Structures. *Eur. J. Comb.*, 19(3):329–343, 1998.

[Jan10] S. Janson. A divertent generating function that can be summed and analysed analytically. *Discret. Math. Theor. Comput. Sci.*, 12(2):1–22, 2010.

[MM03] C. Martínez and X. Molinero. Generic algorithms for the generation of combinatorial objects. In *28th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 572–581. Springer Berlin Heidelberg, 2003.

[Mol05] X. Molinero. *Ordered Generation of Classes of Combinatorial Structures*. Phd thesis, Universitat Politècnica de Catalunya, 2005.

[Moo74] J. W. Moon. *The distance between nodes in recursive trees*, page 125–132. London Mathematical Society Lecture Note Series. Cambridge University Press, 1974.

[NW75] A. Nijenhuis and H. S. Wilf. *Combinatorial algorithms*. Computer science and applied mathematics. Academic Press, New York, NY, 1975.

[Pip10] N. Pippenger. The hypercube of resistors, asymptotic expansions, and preferential arrangements. *Mathematics Magazine*, 83(5):331–346, 2010.

[Pit94] B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures & Algorithms*, 5(2):337–347, 1994.

[Rus03] F. Ruskey. Combinatorial generation. 2003.

[Sch70] E. Schröder. Vier Combinatorische Probleme. *Z. Math. Phys.*, 15:361–376, 1870.

[SM01] M. Steel and A. McKenzie. Properties of phylogenetic trees generated by yule-type specification models. *Mathematical Biosciences*, 170(1):91–112, 2001.

[Yul25] G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213:21–87, 1925.

Olivier Bodini and Mehdi Naima. Université Sorbonne Paris Nord, Laboratoire d'Informatique de Paris Nord, CNRS, UMR 7030, F-93430, Villetaneuse, France.
Email address: `{Olivier.Bodini, Mehdi.Naima}@lipn.univ-paris13.fr`

Antoine Genitrini. Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6 -LIP6- UMR 7606, F-75005 Paris, France.
Email address: `Antoine.Genitrini@lip6.fr`

Cécile Mailler. University of Bath, Department of Mathematical Sciences, Claverton Down, BA2 7AY Bath, UK.
Email address: `C.Mailler@bath.ac.uk`