CONTRIBUTIONS TO THE ANALYSIS OF INCREASING TREES AND OTHER FAMILIES OF TREES

Katherine Morris

A research report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, 2005

Declaration

I declare that this research report is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy in Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Katherine Morris

21 October 2005

Abstract

Increasing trees are labelled rooted trees in which labels along any branch from the root appear in increasing order. They have numerous applications in tree representations of permutations, data structures in computer science and probabilistic models in a multitude of problems.

We use a generating function approach for the computation of parameters arising from such trees. The generating functions for some parameters are shown to be related to ordinary differential equations. Singularity analysis is then used to analyze several parameters of the trees asymptotically.

Various classes of trees are considered. Parameters such as depth and path length for heap ordered trees have been analyzed in [35]. We follow a similar approach to determine grand averages for such trees. The model is that p of the n nodes are labelled at random in $\binom{n}{p}$ ways, and the characteristic parameters depend on these labelled nodes. Also, we will attempt to look at the limiting distributions involved. Often, when they are Gaussian, Hwang's quasi power theorem, from [18], can be employed.

Spanning tree size and the Wiener index for binary search trees have been computed in [33]. The Wiener index is the sum of all distances between pairs of nodes in a tree. A related parameter of interest is the Steiner distance which generalises, to sets of k nodes, the Wiener index (k=2). Furthermore, the distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees is presented in [30]. The same procedure is followed to obtain the Steiner distance for heap ordered trees and for other varieties of increasing trees.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Helmut Prodinger, for his time, help and encouragement. It has been an honour and privilege to work with him.

Also, many thanks go to the School of Mathematics and the University of the Witwatersrand for the opportunity to undertake my degree.

Contents

1

Preliminaries

You can't be suspicious of a tree, or accuse a bird or a squirrel of subversion or challenge the ideology of a violet.

Sundial of the Seasons
Hal Borland

What are trees and what does one do with them? The answers to these questions can be summed up with the following quote from the book of Flajolet and Sedgewick, [11]:

> *Trees are fundamental structures that arise implicitly and explicitly in many practical algorithms and it is important to understand their properties in order to be able to analyze these algorithms. Many algorithms construct trees explicitly; in other cases trees assume significance as models of programs, especially recursive programs. Indeed, trees are the quintessential non trivial recursively defined objects: a tree is either empty or a root node connected to a sequence (or a multiset) of trees.*

We begin with the following formal definition of a tree:

1.1 Definition [Knuth [22]] A *tree* is a finite set T of one or more nodes such that
(a) there is one specially designated node called the *root* of the tree, $\mathrm{root}(T)$ and;
(b) the remaining nodes (excluding the root) are partitioned into $m \geq 0$ disjoint sets $T_1, T_2, \ldots, T_m$ and each of these sets in turn is a tree. The trees $T_1, T_2, \ldots, T_m$ are called the *subtrees* of the root.

It follows that every node of a tree is the root of some subtree contained in the whole tree. The number of subtrees of a node is called the *degree* of that node.

A *labelled tree* is one in which all the nodes are distinguishable, in other words each node has its own identity. We speak of "labels" as a combinatorial device to distinguish nodes.

There are many parameters that arise in tree analysis, among them the height of nodes and path length being the well known ones. Some recent research papers [35], [36] deal with statistics of the height of the nodes in heap ordered trees. The altitude of nodes in random trees was also studied in [24].

One often encounters the *level* of a node in a tree: the root is at level 0, descendants of the root are at level 1 and, in general, descendants of a node at level k are at level $k+1$. Thus the level is the distance traversed to get from the root to the node.

1.2 Definition [11] For a tree T, the *path length* of the tree is the sum of the levels of each of the nodes in T and the *height* of the tree is the maximum level among all the nodes of the tree.

Now the height of a node in a tree T is defined as the number of nodes lying on the unique path from the root to this node. In this thesis we consider a simple generalization of this height: for p given nodes in a heap ordered tree T we consider the size of the ancestor tree of these selected nodes.

1.3 Definition [30] The *ancestor tree* is the subtree of T which is spanned by the root and the p chosen nodes and hence it is defined as the tree containing all ascendants of the p given nodes.

Spanning tree size and the Wiener index for binary search trees have been computed in [23], [28] and [33]. This index was introduced by the chemist H. Wiener in 1947, [42], in the study of organic compounds and their molecular graphs and has numerous applications in chemistry and combinatorics.

1.4 Definition [19] The *Wiener index*, $W(G)$, of a connected graph G is the sum of all the distances between pairs of vertices of G.

A related parameter of interest which will be analysed throughout the thesis is the Steiner distance which is a generalisation of the Wiener index.

1.5 Definition [32] Consider a tree T and choose p random nodes in it. The *Steiner distance* is the size of the smallest subtree generated by the p nodes.

So, the Steiner distance is a scaled down version of the Wiener index; in a sense they behave roughly like path length versus (insertion) depth. For expectations, the concepts are equivalent, but not for higher moments and the limiting distribution. We consider a natural generalization: instead of selecting two random nodes and looking at the distance, we consider p randomly chosen nodes and look at the size of the subtree spanned by these nodes. If all of the p selected nodes lie in one subtree of the root then the Steiner distance is smaller than the size of the ancestor tree. Otherwise the two parameters are the same. A different generalization of the Steiner distance can, for example, be found in [4].

How does one count trees? Given an infinite class of finite sets T_i, where i ranges over some index set, we need to count the number $t(i)$ of elements of each T_i "simultaneously". The answer is eloquently provided by R. Stanley, [40]:

> *The most useful but most difficult to understand method for evaluating $t(i)$ is to give its generating function.*

Results from Flajolet and Sedgewick, [11], [13]

Given the nature of trees, one needs to derive expressions for the value of terms in a sequence of quantities $a_0, a_1, a_2, a_3, \ldots$ which usually measure some parameter. It is very useful to work with a single mathematical object which represents the whole sequence.

1.6 Definition [11] Given a sequence $a_0, a_1, a_2, \ldots, a_k, \ldots$, the function

$$A(z) = \sum_{k \geq 0} a_k z^k \, , \tag{1}$$

is called the *ordinary generating function* of the sequence. One uses the notation $[z^k]A(z)$ to refer to the coefficient a_k.

Some sequences are dealt with in a more convenient way by a generating function that involves a normalising factor.

1.7 Definition [11] Given a sequence $a_0, a_1, a_2, \ldots, a_k, \ldots$, the function

$$A(z) = \sum_{k \geq 0} a_k \frac{z^k}{k!} \, , \tag{2}$$

is called the *exponential generating function* of the sequence. One uses the notation $k![z^k]A(z)$ to refer to the coefficient a_k.

Exponential generating functions are central to the analysis of the trees in this thesis since we deal with labelled structures: say k nodes are labelled such that each has a distinct identity. Then the factor $k!$ accounts for all of the arrangements of the labelled nodes.

An application of generating functions is their use for computing probabilities, averages and variances.

1.8 Definition [11] Given a random variable X that takes on only non-negative integer values, with $p_k \equiv \mathbb{P}\{X = k\}$, the function

$$P(u) = \sum_{k \geq 0} p_k u^k \,, \tag{3}$$

is called the *probability generating function* for the random variable.

Generally one is not only interested in counting tree structures of a given size, but also in finding values of parameters relating to the trees. Bivariate generating functions are used for this purpose. These are functions of two variables that represent doubly indexed sequences: one index for the problem size and one for the value of the parameter under analysis.

1.9 Definition [11] Given a doubly indexed sequence $\{a_{nk}\}$, the function

$$A(u, z) = \sum_{n \geq 0} \sum_{k \geq 0} a_{nk} u^k z^n \,, \tag{4}$$

is called the *bivariate generating function* of the sequence. We use the notation $[u^k z^n]A(u, z)$ to refer to a_{nk}, $[z^n]A(u, z)$ to refer to $\sum_{k \geq 0} a_{nk} u^k$ and $[u^k]A(u, z)$ to refer to $\sum_{n \geq 0} a_{nk} z^n$.

One uses bivariate generating functions to count parameter values in trees. For $p \in \mathcal{P}$, where $\mathcal{P}$ is a certain class of trees, let $cost(p)$ be a function that gives the value of some parameter defined for each tree. Then we have the following generating function for the class of trees

$$P(u, z) = \sum_{p \in \mathcal{P}} u^{\{cost(p)\}} z^{|p|} = \sum_{n \geq 0} \sum_{k \geq 0} p_{nk} u^k z^n \,, \tag{5}$$

where $|p|$ denotes the cardinality of p and p_{nk} is the number of trees of size n and cost k. One also writes

$$P(u, z) = \sum_{n \geq 0} p_n(u) z^n \,, \tag{6}$$

where

$$p_n(u) = [z^n]A(u, z) = \sum_{k \geq 0} p_{nk} u^k \,, \tag{7}$$

to separate all the costs for the trees of size n and

$$P(u, z) = \sum_{k \geq 0} q_k(z) u^k \,, \tag{8}$$

where

$$q_k(z) = [u^k]A(u,z) = \sum_{n\geq 0} p_{nk} z^n\,, \tag{9}$$

to separate all the trees of cost k. It is useful to note that

$$P(1,z) = \sum_{p\in\mathcal{P}} z^{|p|} = \sum_{n\geq 0} p_n(1) z^n = \sum_{k\geq 0} q_k(z)\,, \tag{10}$$

is the ordinary generating function that enumerates $\mathcal{P}$.

Of great interest for this thesis is the fact that $\frac{p_n(u)}{p_n(1)}$ is the probability generating function for the random variable representing cost, if all trees of size n are equally likely. Thus knowing $p_n(u)$ and $p_n(1)$ allows us to compute the average cost and other moments. Differentiating with respect to u and evaluating at $u = 1$, one finds that

$$p_n'(1) = \sum_{k\geq 0} k p_{nk}\,. \tag{11}$$

The partial derivative with respect to u of $P(u,z)$ evaluated at $u = 1$ is the generating function for this quantity. We know that $p_n(1)$ is the number of trees in $\mathcal{P}$ of size n. If we consider the latter to be equally likely, then the probability that a tree of size n has cost k is $\frac{p_{nk}}{p_n(1)}$, the average cost of a tree of size n is $\frac{p_n'(1)}{p_n(1)}$ and the variance is $\frac{p_n''(1)}{p_n(1)} + \frac{p_n'(1)}{p_n(1)} - \left(\frac{p_n'(1)}{p_n(1)}\right)^2$.

1.10 Definition [11] Let $\mathcal{P}$ be a class of trees with bivariate generating function $P(u,z)$. Then the function

$$\left.\frac{\partial P(u,z)}{\partial u}\right|_{u=1} = \sum_{p\in\mathcal{P}} cost(p) z^{|p|}\,, \tag{12}$$

is defined to be the *cumulative generating function* for the class. Also, let $\mathcal{P}_n$ denote the class of all trees of size n in $\mathcal{P}$. Then the sum

$$\sum_{p\in\mathcal{P}_n} cost(p)\,, \tag{13}$$

is defined to be the *cumulated cost* for the trees of size n.

1.11 Theorem [11] *Consider a bivariate generating function $P(u,z)$ for a class of trees. Then the average cost for all trees of a particular size is given by the cumulated cost divided by the number of trees, or*

$$\frac{[z^n]\left.\frac{\partial P(u,z)}{\partial u}\right|_{u=1}}{[z^n]P(1,z)}\,. \tag{14}$$

The importance of the use of bivariate generating functions is that the average cost can be calculated by extracting coefficients independently from $\left.\frac{\partial P(u,z)}{\partial u}\right|_{u=1}$ and $P(1,z)$ and then dividing as outlined in the above theorem.

In this thesis we mainly encounter trivariate generating functions but the results outlined above are analogous. There are many techniques for computing the required coefficients and they appear, for example, in [15], [37] and [43]. When these coefficients have no simple closed form one needs to employ special techniques to find them asymptotically. This is described in various sources such as [10], [12] and [16].

Many applications rely on determining the asymptotic coefficients of a function which is analytic at the origin. It is well-known that the function's dominant singularities (the ones with smallest modulus) contain a lot of information about these coefficients. To this end, singularity analysis of generating functions is employed. In [9], this is defined to be

> *a class of methods by which one can translate, on a term by term basis, an asymptotic expansion of a function around a dominant singularity into a corresponding asymptotic expansion for the Taylor coefficients of the function. This approach is based on contour integration using Cauchy's formula and Hankel-like contours.*

Let $\mathcal{G}$ be the class of functions $g_\alpha = K(1-z)^\alpha$, for α a real number and K a constant. The Taylor coefficients of any member of $\mathcal{G}$ are known both exactly

$$[z^n](1-z)^\alpha = \binom{n-\alpha-1}{n} = \frac{\Gamma(n-\alpha)}{\Gamma(-\alpha)\Gamma(n+1)}, \qquad (15)$$

and asymptotically from Stirling's formula ($\alpha \neq \{0,1,2,\ldots\}$)

$$[z^n](1-z)^\alpha \sim \frac{n^{\alpha-1}}{\Gamma(-\alpha)}\left(1+\frac{\alpha(\alpha+1)}{2n}+\frac{\alpha(\alpha+1)(\alpha+2)(3\alpha+1)}{24n^2}+\cdots\right). \qquad (16)$$

The general form of the asymptotic coefficients above, as it appears in [9], is presented below.

1.12 Proposition [Flajolet, Odlyzko, [9]] *The binomial coefficients expressing* $[z^n](1-z)^\alpha$ *have an asymptotic expansion as* $n \to \infty$,

$$[z^n](1-z)^\alpha \sim \frac{n^{-\alpha-1}}{\Gamma(-\alpha)}\left(1+\sum_{k\geq 1}\frac{e_k^{(\alpha)}}{n^k}\right), \quad \alpha \notin \{0,1,2,\cdots\}, \quad \textit{where}$$

$$e_k^{(\alpha)} = \sum_{l=k}^{2k}(-1)^l\lambda_{k,l}(\alpha+1)(\alpha+2)\cdots(\alpha+l), \quad \textit{with} \quad \sum_{k,l\geq 0}\lambda_{k,l}v^k t^l = e^t(1+vt)^{-1-1/v}. \qquad (17)$$

The authors in [9] were then able to come up with the following very important transfer theorem:

1.13 Theorem [9] *Assume that, with the sole exception of the singularity* $z = 1$, $f(z)$ *is analytic in the domain* $\Delta = \Delta(\phi, \eta)$, *where* $\eta > 0$ *and* $0 < \phi < \frac{\pi}{2}$. *Assume further that as* z *tends to* 1 *in* Δ

$$f(z) = \mathcal{O}(|1 - z|^{\alpha}), \tag{18}$$

for some real number α. *Then the* n*-th Taylor coefficient of* $f(z)$ *satisfies*

$$f_n = [z^n]f(z) = \mathcal{O}(n^{-\alpha-1}). \tag{19}$$

1.14 Note *We recall two useful formulas which will be used in the thesis. Firstly, Cauchy's formula is*

$$f_n = \frac{1}{2\pi i} \int_{\mathcal{O}^+} f(z) \frac{dz}{z^{n+1}}, \tag{20}$$

where f_n *and* $f(z)$ *are as described in the theorem above. Secondly, Stirling's formula is*

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{5140n^3} + \cdots\right). \tag{21}$$

Often one needs to analyse the coefficients of a function defined implicitly by an equation

$$y(z) = z\phi(y(z)), \tag{22}$$

and the procedures involved can be found for instance in [5], [6] and [13]. The next result gives us the idea of what is required.

1.15 Proposition [13] *Let* ϕ *be a function analytic at zero having non negative Taylor coefficients with* $\phi(0) = 0$ *and such that there exists a positive solution* τ *to the characteristic equation*

$$\phi(\tau) - \tau\phi'(\tau) = 0, \tag{23}$$

strictly within the disc of convergence of ϕ. *Let* $y(z)$ *be the solution analytic at the origin of* $y(z) = z\phi(y(z))$. *Then* $y(z)$ *has a dominant singularity at*

$$z = \rho \quad \text{where} \quad \rho = \frac{\tau}{\phi(\tau)}. \tag{24}$$

The singular expansion of y *at* ρ *is of the form*

$$y(z) = \tau + \sum_{j=1}^{\infty} d_j^* \left(1 - \frac{z}{\rho}\right)^{j/2}, \tag{25}$$

for some computable constants d_j^**. In particular one has*

$$d_1^* = -\sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\,. \tag{26}$$

We note that, for all the problems analysed in the chapters that follow, the dominant singularity is always of square-root type. Thus the asymptotic form of the coefficients turns out to be

$$[z^n]y(z) \sim C\,\rho^{-n}\,n^{-3/2}\,. \tag{27}$$

Results from Harary and Palmer, [17]

Let

$$T(x) = \sum_{p=1}^{\infty} T_p x^p\,, \tag{28}$$

be the generating function for rooted trees. Thus T_p is the number of rooted trees of order p. The following result of Pólya, [34], can be employed to calculate the coefficients of $T(x)$.

1.16 Theorem [34] *The counting series* $T(x)$ *for rooted trees satisfies*

$$T(x) = x \exp\left\{\sum_{k=1}^{\infty} \frac{T(x^k)}{k}\right\}. \tag{29}$$

Pólya found asymptotic formulas for chemical compounds by regarding the generating functions as ordinary analytic functions such that the coefficients could be estimated by means of the Cauchy integral formula. Otter, [29], observed that the same method could be applied to trees and in what follows we present some of his results.

1.17 Lemma [17] *The power series* $T(x)$ *for rooted trees converges in a circle of radius* $\eta \geq \frac{1}{4}$.

All the coefficients of $T(x)$ are positive so η is a singularity of $T(x)$. However, $T(x)$ converges with $x = \eta$, which can be proved by using the next result.

1.18 Lemma [17] *The limit of* $T(x)$ *as* $x \to \eta-$ *exists and is equal to* $\sum_{k=1}^{\infty} T_k \eta^k$.

Proof: Since $T(x)$ satisfies the functional equation (29) we have for all x in $(0, \eta)$

$$\log\left(\frac{T(x)}{x}\right) = T(x) + \sum_{k=2}^{\infty} \frac{T(x^k)}{k}\,. \tag{30}$$

From this it follows that

$$\frac{\frac{T(x)}{x}}{\log\left(\frac{T(x)}{x}\right)} \leq \frac{1}{x}\,, \tag{31}$$

and hence $T(x)$ is bounded on the interval $(0, \eta)$. Since $T(x)$ is monotone, the left hand side limit at η exists and we let

$$b_0 = \lim_{x \to \eta-} T(x)\,. \tag{32}$$

It now follows immediately that $b_0 = T(\eta)$. ∎

The value of b_0 is determined by the next lemma.

1.19 Lemma [17] *The series $T(x)$ for rooted trees has the property that*

$$T(\eta) = 1\,. \tag{33}$$

Proof: First we define the complex valued function $F(x, y)$ for complex x and y by

$$F(x, y) = x \exp\left\{y + \sum_{k=2}^{\infty} \frac{T(x^k)}{k}\right\} - y\,, \tag{34}$$

and consider the equation

$$F(x, y) = 0\,. \tag{35}$$

From (29) we can show that $y = T(x)$ is the unique analytic solution of (35) and we know it has a singularity at $x = \eta$. The preceding lemma implies that $F(\eta, b_0) = 0$ and furthermore it is easy to see that $F(x, y)$ is analytic in each variable separately in neighbourhoods of η and b_0.

On differentiating (34) with respect to y we find

$$\frac{\partial F}{\partial y} = F(x, y) + y - 1\,. \tag{36}$$

Since $F(\eta, b_0) = 0$, we know that this partial derivative at (η, b_0) is given by

$$\frac{\delta F}{\delta y}(\eta, b_0) = b_0 - 1\,. \tag{37}$$

Furthermore, this partial derivative must be zero at (η, b_0), in other words $b_0 = 1$. Otherwise, by the implicit function theorem, there is a unique solution $y = f(x)$ of (35) which is analytic in a neighbourhood of η, in particular at η itself. But such a solution would have to be $y = T(x)$ and we know that the latter is not analytic at $x = \eta$, proving (33). ∎

Often one needs to know the asymptotic expansions of various functions around their dominant singularity. The next theorem (which is a combination of the implicit function theorem and observations of Polya, Otter, Ford and Uhlenbeck) is pivotal in this respect.

1.20 Theorem [17] *Let $F(x,y)$ be analytic in each variable separately in some neighbourhood of (x_0,y_0) and suppose that the following conditions are satisfied:*
(i) $F(x_0,y_0)=0$;
(ii) $y=f(x)$ is analytic in $|x|<|x_0|$ and x_0 is the unique singularity on the circle of convergence;
(iii) if $f(x)=\sum_{n=0}^{\infty} f_n x^n$ is the expansion of f at the origin, then $y_0=\sum_{n=0}^{\infty} f_n x_0^n$;
(iv) $F(x,f(x))=0$ for $|x|<x_0$;
(v) $\frac{\partial F(x_0,y_0)}{\partial y}=0$;
(vi) $\frac{\partial^2 F(x_0,y_0)}{\partial y^2}\neq 0$.
Then $f(x)$ may be expanded about x_0:

$$f(x)=f(x_0)+\sum_{k=1}^{\infty} a_k(x_0-x)^{k/2}\,, \tag{38}$$

and if $a_1\neq 0$,

$$f(x)\sim -\frac{a_1}{2\sqrt{\pi}}x_0^{-n+1/2}n^{-3/2}\,, \tag{39}$$

and if $a_1=0$ but $a_3\neq 0$

$$f_n\sim \frac{3a_3}{4\sqrt{\pi}}x_0^{-n+3/2}n^{-5/2}\,. \tag{40}$$

To apply this theorem, one notes that the function in (34) satisfies all the conditions with $(x_0,y_0)=(\eta,1)$ and $f(x)=T(x)$. Hence $T(x)$ can be expanded as in (38) and if $a_1\neq 0$, then the coefficients behave as in (39). It remains to be shown that if

$$T(x)=1-b_1(\eta-x)^{1/2}+b_2(\eta-x)+b_3(\eta-x)^{3/2}+\cdots\,, \tag{41}$$

then $b_1\neq 0$ and $b_3\neq 0$ and one also requires approximations to b_1 and η. After differentiating (41) we obtain

$$T'(x)=\frac{1}{2}b_1(\eta-x)^{-1/2}-b_2+\cdots\,, \tag{42}$$

where the terms omitted contain $(\eta-x)^{1/2}$ to the first power at least. By multiplying both sides of $T'(x)$ with $1-T(x)$ as obtained from (41) we have

$$T'(x)(1-T(x))=\frac{1}{2}b_1^2+\cdots\,, \tag{43}$$

where once again the terms omitted contain $(\eta - x)^{1/2}$ to at least the first power. Hence

$$\lim_{x \to \eta-} T'(x)(1 - T(x)) = \frac{1}{2} b_1^2 \,. \tag{44}$$

On the other hand, by differentiating (29), one obtains

$$T'(x) = \frac{T(x)}{x} + T(x) \sum_{k=1}^{\infty} T'(x^k) x^{k-1} \,, \tag{45}$$

and therefore

$$T'(x)(1 - T(x)) = \frac{T(x)}{x} + T(x) \sum_{k=2}^{\infty} T'(x^k) x^{k-1} \,. \tag{46}$$

Thus the limit in (44) can also be obtained from the above and we have

$$\frac{1}{2} b_1^2 = \frac{1}{\eta} + \sum_{k=2}^{\infty} T'(\eta^k) \eta^{k-1} \,. \tag{47}$$

Using (29) and (33), Otter estimated that $\eta = 0.3383219$. Then from an equation similar to the one above he found that

$$\frac{b_1 \eta^{1/2}}{2\sqrt{\pi}} = 0.4399237 \ldots \,, \tag{48}$$

with $b_1 = 2.681127$.

Results from Bergeron, Flajolet and Salvy, [2]

1.21 Definition [2] A *labelled tree* of size n is a rooted tree comprising n nodes that are labelled by distinct integers from the set $\{1, 2, 3, \ldots, n\}$.

1.22 Definition [2] An *increasing tree* is a labelled rooted tree in which labels along any branch from the root appear in increasing order.

These trees appear in problems involving tree representations of permutations, data structures in computer science and probabilistic models in diverse applications. An interesting approach to the enumeration of parameters in these trees was developed in [2], whereby the generating functions are related to a simple ordinary differential equation

$$\frac{d}{dz} Y(z) = \phi(Y(z)) \,, \tag{49}$$

which is non linear and autonomous. Then one applies singularity analysis in order to analyse asymptotically the required parameters.

There are two important types of increasing trees:
(i) *non-plane trees*, which are defined in the graph theoretic sense so that subtrees stemming from a node are not ordered among themselves;
(ii) *plane trees*, where a plane embedding is specified so that subtrees stemming from a node are ordered among themselves.

1.23 Definition [2] Let $\{s_r\}_{r=0}^{\infty}$ be a sequence of non negative integers, such that $s_0 \neq 0$ and $s_r \neq 0$ for some $r \geq 2$. The *variety of trees* associated to $\{s_r\}$ and the specification of an element of {Plane, Non-Plane} is the collection of all increasing trees (plane or non-plane depending on the specification) with s_r sorts of nodes of outdegree r for all r.

The *degree function* of a variety of trees associated with $\{s_r\}$ is defined as follows

$$\text{in the plane case:} \qquad \phi(\omega) = \sum_{r\geq 0} s_r \omega^r \,,$$
$$\text{in the non-plane case:} \qquad \phi(\omega) = \sum_{r\geq 0} s_r \frac{\omega^r}{r!} \,. \tag{50}$$

Now, the coefficients of the degree function $\phi(\omega)$ are denoted by ϕ_r such that

$$\phi(\omega) = \sum_{r=0}^{\infty} \phi_r \omega^r \,, \quad \text{with} \begin{cases} \phi_r = s_r & \text{in the plane case and} \\ \phi_r = \frac{s_r}{r!} & \text{in the non-plane case}\,. \end{cases} \tag{51}$$

Whenever the collection of node types allowed is finite, $\phi(\omega)$ is a polynomial and we call the variety a *polynomial variety*. The degree of $\phi(\omega)$ is denoted by d, an integer, which is the maximum node degree allowed in the variety.

The main result is as follows: fix a variety of trees $\mathcal{Y}$, that is fix the degree function $\phi(\omega)$. Let Y_n be the number of trees of size n in the variety. The exponential generating function of the variety of trees

$$Y(z) = \sum_{n=1}^{\infty} Y_n \frac{z^n}{n!} \,, \tag{52}$$

is defined implicitly by

$$\int_0^{Y(z)} \frac{d\omega}{\phi(\omega)} = z \,. \tag{53}$$

Under general conditions (in particular whenever $\phi(\omega)$ is a polynomial, which means a finite set of allowed degrees) the equation (53) can be analysed near its dominant singularity, ρ. For polynomial varieties this leads to an asymptotic counting result of the form

$$\frac{Y_n}{n!} \sim K\rho^{-n} n^{-(d-2)/(d-1)} \quad \text{with} \quad \rho = \int_0^\infty \frac{d\omega}{\phi(\omega)}, \tag{54}$$

where $K = K_\phi$ is a constant that depends on ϕ alone and d is the degree of the variety. The quantity ρ appears to be always a logarithmic form in algebraic numbers.

2

Ordered Trees

The chestnut's proud, and the lilac's pretty,
The poplar's gentle and tall,
But the plane tree's kind to the poor dull city -
I love him best of all.

Child's Song in Spring
Edith Nesbit

2.1 Introduction

In this chapter we consider ordered trees, a special class of simply generated trees. What is presented here constitutes a specialisation of the results obtained by Panholzer in [30]. Although the general results are not new, they serve to illustrate the methods and ideas employed in subsequent chapters of the this thesis.

Simply generated trees include several tree families such as binary trees and planted plane (planar, ordered or rooted) trees. Let $\mathcal{T}$ be the set of ordered labelled trees with exponential generating function

$$T(z) = \sum_{t \in \mathcal{T}} \frac{z^{|t|}}{|t|!} = \sum_{n \geq 0} T_n \frac{z^n}{n!}, \tag{1}$$

where T_n is the number of ordered trees with n nodes. Then the generating function for ordered trees satisfies

$$T(z) = \frac{z}{1 - T(z)}, \quad \text{thus} \quad T(z) = \frac{1 - \sqrt{1 - 4z}}{2}. \tag{2}$$

2.1 Theorem [11] *Let t_n denote the number of ordered trees with n nodes. Then t_n is equal to the number of binary trees with $n - 1$ nodes and is given by the Catalan numbers*

$$t_n = \frac{1}{n}\binom{2n-2}{n-1} = \frac{1}{2n-1}\binom{2n-1}{n} = \frac{(2n-2)!}{n!(n-1)!}. \tag{3}$$

A class $\mathcal{T}$ of simply generated trees can also be defined in the following way.

2.2 Definition [30] The *weight*, $w(T)$, of any simply generated tree T is defined by a sequence of non-negative numbers $(\phi_k)_{k\geq 0}$ with $\phi_0 > 0$ such that

$$w(T) = \prod_{v} \phi_{d(v)}\,, \tag{4}$$

where v ranges over all vertices of T and $d(v)$ is the number of descendants of v. Then the *family* $\mathcal{T}$ consists of all trees T and its weights $w(T)$.

Furthermore,

$$T_n = \sum_{|T|=n} w(T)\,, \tag{5}$$

where $|T|$ is the size of the tree and has the generating function

$$T(z) = \sum_{n\geq 0} T_n z^n\,. \tag{6}$$

This generating function satisfies the functional equation

$$T(z) = z\phi(T(z))\,, \tag{7}$$

where $\phi(t)$ is given as the formal power series

$$\phi(t) = \sum_{k\geq 0} \phi_k t^k\,. \tag{8}$$

If all ϕ_k are non-negative integers then T_n counts the number of trees in $\mathcal{T}$ with size n.

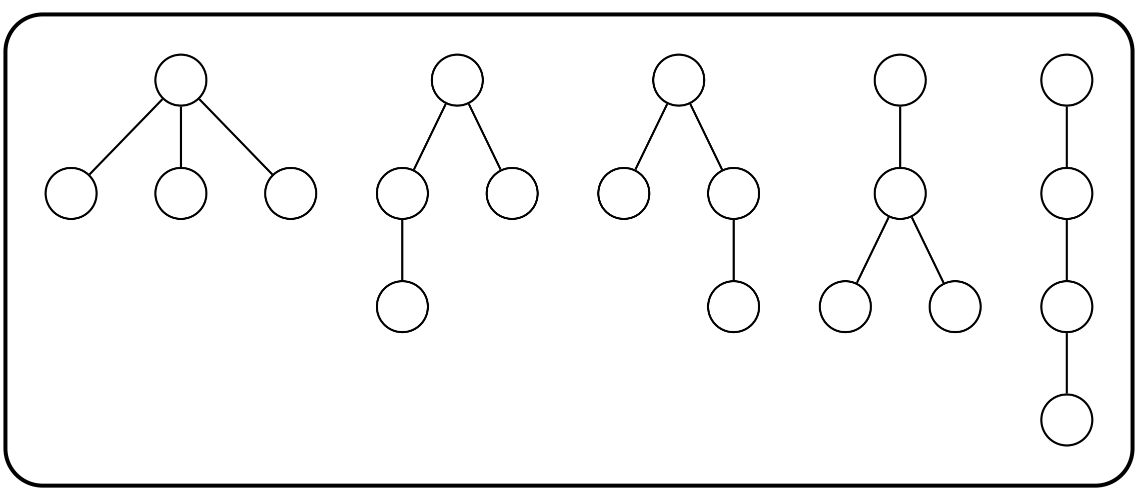


Figure 2.1: All 5 ordered trees of size 4.

2.2 Size of the ancestor tree in ordered trees

We consider the results of Panholzer from [30] and derive them explicitly for ordered trees. Let $X_{n,p}$ be the random variable which counts the size of the ancestor tree of p randomly chosen nodes in an ordered tree of size n. We denote with $g_{n,p,m}$ the sum of the weights $w(T)$ coming from all different pairs (T, S) of a simply generated tree T of size n and a subset S of p nodes of T such that the ancestor tree of these p nodes in T has size m. It follows that

$$g_{n,p,m} = \mathbb{P}(X_{n,p} = m) T_n \binom{n}{p}, \tag{9}$$

and the trivariate generating function

$$G(z, u, v) = \sum_{n \geq 1} \sum_{0 \leq p \leq n} \sum_{m \geq 0} g_{n,p,m} z^n u^p v^m, \tag{10}$$

satisfies the following functional equation from [30]

$$G(z, u, v) = zv(1 + u)\phi(G(z, u, v)) + (1 - v)T(z), \tag{11}$$

where the term $(1 - v)T(z)$ results from the case where no nodes are selected in the tree. We look at a special case of (11), namely the planar trees, for which the functional equation becomes

$$G(z, u, v) = \frac{zv(1 + u)}{1 - G(z, u, v)} + (1 - v)T(z). \tag{12}$$

However, for planar trees, we know $T(z)$ and its solution. Since there is no tree of size 0 we have

$$T(z) = \frac{1 - \sqrt{1 - 4z}}{2} = \sum_{n \geq 1} \frac{1}{n} \binom{2n - 2}{n - 1} z^n.$$

Substituting this value in (12) we obtain:

2.3 Theorem [30] *The generating function for the size of the ancestor tree in ordered trees satisfies the functional equation*

$$G(z, u, v) = \frac{zv(1 + u)}{1 - G(z, u, v)} + \frac{(1 - v)(1 - \sqrt{1 - 4z})}{2}. \tag{13}$$

First, the solution of (13) is computed

$$\begin{aligned} G(z, u, v) =& \frac{3 - v}{4} + \frac{\sqrt{1 - 4z}(v - 1)}{4} \\ &- \frac{\sqrt{2 + 2\sqrt{1 - 4z} - 4z - 8zv + 2v^2 - 2v^2\sqrt{1 - 4z} - 4v^2 z - 16zvu}}{4}. \end{aligned} \tag{14}$$

Our aim is to find the expectation and variance of the size of the ancestor tree with the help of the probabilities

$$\mathbb{P}\{X_{n,p} = m\} = \frac{[z^n u^p v^m] G(z,u,v)}{\binom{n}{p} T_n} ,$$

for fixed p and $n \to \infty$, and thus one needs to produce $[z^n u^p v^m] G(z,u,v)$. For the expectation we differentiate (14) with respect to v and evaluate at $v = 1$, namely

$$\frac{\partial}{\partial v} G(z,u,v)\Big|_{v=1} = \frac{\sqrt{1-4z}-1}{4} + \frac{z(1+u)}{\sqrt{1-4z(1+u)}} + \frac{\sqrt{1-4z}-1}{4\sqrt{1-4z(1+u)}} . \qquad (15)$$

Now we can easily find the coefficients $[z^n u^p]$ from (15), as follows

$$\begin{aligned}
[z^n u^p] \frac{z(1+u)}{\sqrt{1-4z(1+u)}} &= \binom{n}{p}\binom{2n-2}{n-1} , \\
[z^n u^p] \frac{\sqrt{1-4z}}{4\sqrt{1-4z(1+u)}} &= 4^{n-p-1}\binom{2p}{p}\binom{n-1}{p-1} , \\
[z^n u^p] \frac{1}{4\sqrt{1-4z(1+u)}} &= \frac{1}{4}\binom{n}{p}\binom{2n}{n} .
\end{aligned} \qquad (16)$$

The final step is to divide these coefficients by the normalising constant $\binom{n}{p}\frac{1}{n}\binom{2n-2}{n-1}$ and then the expectation becomes

$$E_{n,p} = \frac{[z^n u^p] \frac{\partial}{\partial v} G(z,u,v)\big|_{v=1}}{\binom{n}{p}\frac{1}{n}\binom{2n-2}{n-1}} = \frac{1}{2} + 4^{n-p-1}\frac{(2p)!(n-1)!(n-1)!}{p!(p-1)!(2n-2)!} , \qquad (17)$$

and moreover, by standard methods (using Stirling's formula) we obtain the following asymptotic equivalents

$$E_{n,p} = \begin{cases} n - \frac{l}{2} - \frac{l(l-1)}{8n} + \mathcal{O}\left(\frac{1}{n^2}\right) , & \text{for } l = n-p \text{ fixed} , \\ 4^{-p-1}\, p\binom{2p}{p}\sqrt{\pi n} + \mathcal{O}(1) , & \text{for } p \text{ fixed} , \\ \frac{1}{2} + 2^{-\rho n - 4}\frac{\sqrt{\pi n}}{(\rho n - 1)!} + \mathcal{O}(1) , & \text{for } p = \rho n \text{ and } 0 < \rho < 1 . \end{cases} \qquad (18)$$

To find the variance of the size of the ancestor tree one differentiates (14) twice with respect to v and lets $v = 1$, thus

$$\frac{\partial^2}{\partial v^2} G(z,u,v)\Big|_{v=1} = \frac{2z^2 u(1+u)}{(1-4z(1+u))^{3/2}} . \qquad (19)$$

Then we compute the coefficient $[z^n u^p]$ in (19)

$$[z^n u^p]\frac{z^2u(1+u)}{(1-4z(1+u))^{3/2}} = 2(2n-3)\binom{n-1}{p-1}\binom{2n-4}{n-2}.$$

Upon dividing this coefficient by the normalising constant we obtain the variance

$$\begin{aligned} V_{n,p} &= \frac{[z^n u^p]\frac{\partial^2}{\partial v^2}G(z,u,v)\big|_{v=1}}{\binom{n}{p}\frac{1}{n}\binom{2n-2}{n-1}} + E_{n,p} - (E_{n,p})^2 \\ &= p(n-1) + \frac{1}{2} + \frac{4^{n-p-1}(2p)!\,((n-1)!)^2}{p!(p-1)!(2n-2)!} - \left(\frac{1}{2} + \frac{4^{n-p-1}(2p)!\,((n-1)!)^2}{p!(p-1)!(2n-2)!}\right)^2 . \end{aligned} \tag{20}$$

We can also find the first two moments for the size of the ancestor tree by using singularity analysis. Proceeding from (15), one can use MAPLE to compute the coefficients a_i of its series expansion in u and the first few are

$$\begin{aligned} a_1 &= -\frac{z(4z-\sqrt{1-4z}-1)}{2(1-4z)^{3/2}}, \\ a_2 &= \frac{z^2(4z-3\sqrt{1-4z}-1)}{2(1-4z)^{5/2}}, \\ a_3 &= -\frac{z^3(4z-5\sqrt{1-4z}-1)}{(1-4z)^{7/2}}, \\ a_4 &= \frac{5z^4(4z-7\sqrt{1-4z}-1)}{2(1-4z)^{9/2}}, \\ a_5 &= -\frac{7z^5(4z-9\sqrt{1-4z}-1)}{(1-4z)^{11/2}}. \end{aligned} \tag{21}$$

We observe that $z = \frac{1}{4}$ is the only singularity appearing in the a_i. By computing series expansions around this singularity and considering the first two terms only, we are able to compute the coefficients of u^p and then z^n in (15).

2.4 Lemma *The coefficients of u^p in the first derivative of $G(z,u,v)$ are*

$$[u^p]\frac{\partial G(z,u,v)}{\partial v}\bigg|_{v=1} \sim \frac{1}{2^{2p+2}}\binom{2p}{p}(1-4z)^{-p}, \ z\to\frac{1}{4}. \tag{22}$$

Proof: This is done via straight forward computations of the coefficient of u^p in the dominant term $\frac{\sqrt{1-4z}}{4\sqrt{1-4z(1+u)}}$ from (15). ■

Then the coefficient of z^n in the first derivative of $G(z,u,v)$ turns out to be

$$[z^n u^p]\frac{\partial}{\partial v}G(z,u,v)\Big|_{v=1} \sim \binom{2p}{p}\frac{4^n n^{p-1}}{2^{2p+2}\Gamma(p)}\,, \tag{23}$$

and the expectation is found by normalising the above with $\binom{n}{p}\frac{1}{n}\binom{2n-2}{n-1} \sim \frac{4^n n^{-3/2}}{4\sqrt{\pi}}\frac{n^p}{\Gamma(p+1)}$

$$E_{n,p} \sim \frac{p\sqrt{\pi}}{2^{2p}}\binom{2p}{p}\sqrt{n}\,, \tag{24}$$

which agrees with the result in (17). A similar method works for the second moment.

2.5 Lemma *The coefficients of u^p in the second derivative of $G(z,u,v)$ have the asymptotic form*

$$[u^p]\frac{\partial^2}{\partial v^2}G(z,u,v)\Big|_{v=1} \sim \frac{p}{2^{2p+2}}\binom{2p}{p}(1-4z)^{-p-1/2}\,,\ z\to\frac{1}{4}\,. \tag{25}$$

Proof: One proceeds by extracting the desired coefficient of the dominant term from (19) as follows

$$\begin{aligned}[u^p]\frac{\partial^2}{\partial v^2}G(z,u,v)\Big|_{v=1} &\sim [u^p]\frac{2z^2u}{(1-4z(1+u))^{3/2}} = 2z^2[u^{p-1}]\frac{1}{(1-4z)^{3/2}\left(1-\frac{4zu}{1-4z}\right)^{3/2}}\\ &= 2z^2(-4z)^{p-1}\binom{-3/2}{p-1}(1-4z)^{-p-1/2}\\ &= \frac{p}{2^{2p+2}}(1-4z)^{-p-1/2}\,,\ z\to\frac{1}{4}\,,\end{aligned} \tag{26}$$

which proves the result. ■

Next, the coefficient of z^n in the second derivative of $G(z,u,v)$ was computed

$$[z^n u^p]\frac{\partial^2}{\partial v^2}G(z,u,v)\Big|_{v=1} \sim \frac{p4^n}{2^{2p+2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{27}$$

and its normalisation leads to the second moment. Hence, the asymptotic formula for the variance of the ancestor tree size is

$$\begin{aligned}V_{n,p} &\sim \frac{[z^n u^p]\frac{\partial^2}{\partial v^2}G(z,u,v)\big|_{v=1}}{\frac{4^n n^{-3/2}}{4\sqrt{\pi}}\frac{n^p}{\Gamma(p+1)}} + E_{n,p} - (E_{n,p})^2\\ &\sim \frac{p\sqrt{\pi}}{2^{2p}}\binom{2p}{p}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})}n - \frac{p^2\pi}{2^{4p}}\binom{2p}{p}^2 n\\ &= pn\left(1-\frac{p\pi}{2^{4p}}\binom{2p}{p}^2\right),\end{aligned} \tag{28}$$

where $1 - \frac{p\pi}{2^{4p}}\binom{2p}{p}^2 \to 0$ as $p \to \infty$.

2.3 The Steiner distance in ordered trees

We turn our attention to the random variable $Y_{n,p}$ which counts the size of the Steiner distance of p randomly chosen nodes in an ordered tree of size n and specialise the results of Panholzer, [30]. Let $f_{n,p,m}$ be the sum of the weights $w(T)$ coming from all different pairs (T, S) of a simply generated tree T of size n and a subset S of p nodes of T such that the ancestor tree of these p nodes in T has size m. It follows that

$$f_{n,p,m} = \mathbb{P}(Y_{n,p} = m)T_n\binom{n}{p}, \tag{29}$$

and the trivariate generating function

$$F(z,u,v) = \sum_{n\geq 1}\sum_{0\leq p\leq n}\sum_{m\geq 0} f_{n,p,m}z^n u^p v^m, \tag{30}$$

satisfies the following functional equation which appears in [30]

$$F(z,u,v) = G(z,u,v) - zv\phi'(T(z))G(z,u,v) + z\phi'(T(z))F(z,u,v) - z(1-v)\phi'(T(z))T(z). \tag{31}$$

However, this equation for $F(z,u,v)$ can be obtained by translating the equation (7). The difference between the parameters $X_{n,p}$ and $Y_{n,p}$ coming from the case where the root is not selected and all selected nodes appear on the same subtree leads to the given relation between both generating functions. Thus (31) becomes

$$F(z,u,v) = \frac{G(z,u,v)(1 - zv\phi'(T(z))) - z(1-v)\phi'(T(z))T(z)}{1 - z\phi'(T(z))}. \tag{32}$$

We apply this result to the case of planar trees. Since $\phi'(t) = \frac{1}{(1-t)^2}$ we have

$$\phi'(T(z)) = \frac{1}{(1-T(z))^2}. \tag{33}$$

Substituting this value of $\phi'(T(z))$ into the functional equation (32) we obtain the following result.

2.6 Theorem [30] *The Steiner distance in ordered trees has the generating function*

$$F(z,u,v) = \frac{(1+\sqrt{1-4z} - 2z(1+v))G(z,u,v) - z(1+v+\sqrt{1-4z}(1-v))}{1+\sqrt{1-4z} - 4z}. \tag{34}$$

One can find the expectation and variance of the size of the Steiner distance with the help of the probabilities

$$\mathbb{P}\{Y_{n,p} = m\} = \frac{[z^n u^p v^m]F(z,u,v)}{\binom{n}{p} T_n},$$

for fixed p and $n \to \infty$, and thus the first step is to compute $[z^n u^p v^m]F(z,u,v)$.

Now we substitute the value of $G(z,u,v)$ from (14) into the lemma above, then we differentiate with respect to v and set $v = 1$, which gives

$$\frac{\partial}{\partial v}F(z,u,v)\Big|_{v=1} = \frac{z(2+u(1+\sqrt{1-4z}) - 2\sqrt{1-4z}\sqrt{1-4z(1+u)} - 8z(1+u))}{\sqrt{1-4z(1+u)}(1+\sqrt{1-4z}-4z)}. \tag{35}$$

We proceed with finding the coefficients of $z^n u^p$ in (35). It is advantageous to first find the coefficients of u^p and then use

$$\frac{1}{1+\sqrt{1-4z}-4z} = \frac{1}{4z\sqrt{1-4z}} - \frac{1}{4z}, \tag{36}$$

obtained by multiplying top and bottom by $1-4z-\sqrt{1-4z}$, to find the coefficients of z^n. It follows that

$$\begin{aligned}
[z^n u^p]\frac{\partial}{\partial v}F(z,u,v)\Big|_{v=1} =& \frac{1}{2}4^{n-p}\binom{2p}{p}\binom{n}{n-p} - \frac{1}{2}4^{n-p}\binom{2p}{p}\binom{n-\frac{1}{2}}{n-p}\\
&+4^{n-p}\binom{2p-2}{p-1}\binom{n}{n-p+1} - 4^{n-p}\binom{2p-2}{p-1}\binom{n-\frac{1}{2}}{n-p+1}\\
&+4^{n-p}\binom{2p-2}{p-1}\binom{n-\frac{1}{2}}{n-p+1} - 4^{n-p}\binom{2p-2}{p-1}\binom{n-1}{n-p+1}\\
&+2\cdot 4^{n-p-1}\binom{2p}{p}\binom{n-\frac{3}{2}}{n-p-1} - 2\cdot 4^{n-p-1}\binom{2p}{p}\binom{n-1}{n-p-1}\\
&+2\cdot 4^{n-p}\binom{2p-2}{p-1}\binom{n-\frac{3}{2}}{n-p} - 2\cdot 4^{n-p}\binom{2p-2}{p-1}\binom{n-1}{n-p}.
\end{aligned} \tag{37}$$

Next, (37) is divided by the normalising factor $\frac{1}{n}\binom{n}{p}\binom{2n-2}{n-1}$ which yields the expectation for the Steiner distance

$$\begin{aligned}
E_{n,p} &= 4^{n-p}\frac{(p-1)(2p-2)!\,((n-1)!)^2}{((p-1)!)^2\,(2n-2)!)} + 4^{n-p}\frac{(2p-1)!(n-1)!(n-\frac{3}{2})!}{2(p-\frac{1}{2})!(p-1)!(2n-2)!}\\
&= 4^{n-p}\frac{(p-1)(2p-2)!\,((n-1)!)^2}{((p-1)!)^2\,(2n-2)!)} + 1\\
&= \begin{cases} n, & p = n,\\ 1, & p = 1. \end{cases}
\end{aligned} \tag{38}$$

To compute the second moment, we differentiate $F(z,u,v)$ twice with respect to v and then put $v=1$ to obtain

$$\frac{\partial^2 F(z,u,v)}{\partial v^2}\Big|_{v=1} = \frac{z}{(1-4z(1+u))^{3/2}(1-4z+\sqrt{1-4z})}\Big[\sqrt{1-4z}(-1+2z(1+u)(2+u)) + (1-\sqrt{1-4z})(1-4z(1+u))^{3/2} + 2z(1+u)(u(1+4z)-4(1+2z))\Big]. \tag{39}$$

Now the coefficients of u^p and z^n need to be extracted. It is obvious that the methods employed to compute the expectation become quite cumbersome here. To this end, singularity analysis proves to be a better approach. Before proceeding with the second moment, it is interesting to see to see how one obtains the expectation for the Steiner distance using singularity analysis. The series expansion in u of (35) reveals our old singularity $z = \frac{1}{4}$.

2.7 Lemma *The coefficients of u^p in the first derivative of $F(z,u,v)$ are*

$$[u^p]\frac{\partial}{\partial v}F(z,u,v)\Big|_{v=1} \sim \frac{p-1}{(2p-1)2^{2p+1}}\binom{2p}{p}(1-4z)^{-p}, \; z \to \frac{1}{4}. \tag{40}$$

Furthermore, the coefficients of z^n in the first derivative of $F(z,u,v)$ can be easily computed now

$$[z^n u^p]\frac{\partial}{\partial v}F(z,u,v)\Big|_{v=1} \sim \frac{(p-1)4^n}{(2p-1)2^{2p+1}}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}, \tag{41}$$

and their normalisation yields the same expectation obtained in (38)

$$E_{n,p} \sim \frac{p(p-1)\sqrt{\pi}}{(2p-1)2^{2p-1}}\binom{2p}{p}\sqrt{n}. \tag{42}$$

2.8 Note *For $p = n$ in the above expectation, one obtains*

$$E_{n,n} \sim n - \frac{5}{8} - \frac{23}{128}n^{-1} + O\left(n^{-2}\right). \tag{43}$$

For the second moment, the coefficients extracted from the dominant term in (39) are given in the next result.

2.9 Lemma *The coefficients of u^p in the second derivative of $F(z,u,v)$ have the form*

$$[u^p]\frac{\partial^2}{\partial v^2}F(z,u,v)\Big|_{v=1} \sim \frac{p-1}{2^{2p+2}}\binom{2p}{p}(1-4z)^{-p-1/2}, \; z \to \frac{1}{4}. \tag{44}$$

So it follows that the coefficients of z^n in the second derivative of $F(z,u,v)$ are given asymptotically by

$$[z^n u^p]\frac{\partial^2}{\partial v^2}F(z,u,v)\Big|_{v=1} \sim \frac{(p-1)4^n}{2^{2p+2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{45}$$

moreover, the second moment is

$$\frac{[z^n u^p]\frac{\partial^2}{\partial v^2}F(z,u,v)\Big|_{v=1}}{\frac{4^n}{4\sqrt{\pi}}n^{-3/2}\frac{n^p}{\Gamma(p+1)}} \sim (p-1)n\,, \tag{46}$$

and the variance for the Steiner distance becomes

$$V_{n,p} \sim (p-1)n\Big(1-\frac{\pi p^2(p-1)}{(2p-1)^2 2^{4p-2}}\binom{2p}{p}^2\Big)\,. \tag{47}$$

2.10 Note *For $p=1$ the variance above is zero and for $p=n$ we obtain*

$$V_{n,n} \sim \frac{1}{4}n - \frac{1}{32} - \frac{5}{128}n^{-1} + \mathcal{O}(n^{-2})\,. \tag{48}$$

Also, $1-\frac{\pi p^2(p-1)}{(2p-1)^2 2^{4p-2}}\binom{2p}{p}^2 \to 0$ *as* $p\to\infty$.

We started the analysis of tree structures with one of the simplest constructions: the ordered (planar) tree. While it was possible to obtain closed form results in some cases, computing asymptotics turned out to be a better approach.

3

Heap Ordered Trees

> If what I say resonates with you, it is merely because we are both branches on the same tree.
>
> William Butler Yeats

3.1 Introduction

3.1 Definition [35] A *heap ordered tree* with n nodes ("size n") can be described as a *planted plane tree* together with a bijection from the nodes to the set $\{1, \ldots, n\}$ which is *monotonically increasing* when going from the root to the leaves.

The enumeration of heap ordered trees appeared in [38] and parameters such as depth, path length and level of nodes have been analysed in [35] and [36].

3.2 Theorem [35] *Let a_n be the number of heap ordered trees of size n. Then a_n satisfies the recursion*

$$a_{n+1} = \sum_{k \geq 1} \sum_{j_1+\ldots+j_k=n} \binom{n}{j_1, \ldots, j_k} a_{j_1} \cdots a_{j_k}\,, \ \textit{for } n \geq 1\,,\ a_1 = 1\,. \tag{1}$$

It follows that the exponential generating function for heap ordered trees

$$A(z) = \sum_{n \geq 0} a_n \frac{z^n}{n!}\,, \tag{2}$$

fulfills the differential equation

$$A'(z) = \frac{1}{1 - A(z)} \quad \text{with} \quad A(0) = 0\,. \tag{3}$$

Its solution is

$$A(z) = 1 - \sqrt{1 - 2z}\,, \tag{4}$$

and it follows that

$$a_n = n! 2^{1-n} \mathcal{C}_n\,, \tag{5}$$

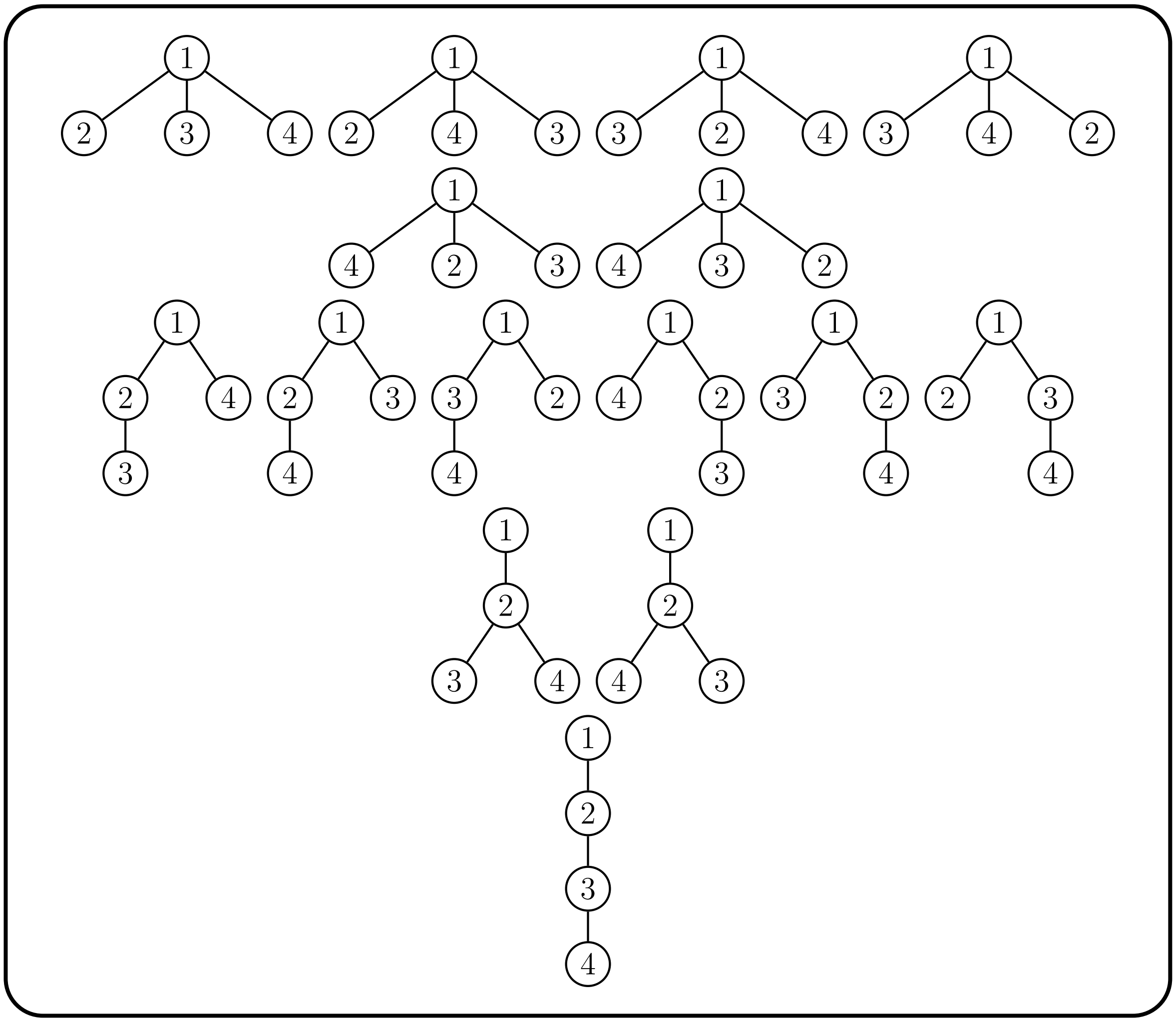


Figure 3.1: All 15 heap ordered trees with 4 nodes.

where $\mathcal{C}_n$ represents the shifted Catalan number $\frac{1}{n}\binom{2n-2}{n-1}$.

The aim is to compute the expectation and variance for the size of the ancestor tree and the Steiner distance in heap ordered trees. Also, we will consider the limiting distributions involved. For the parameters discussed the distributions turn out to be Gaussian and we will use Hwang's *quasi power theorem*, see [18], to determine them. All the results presented in this chapter have appeared in [26].

3.3 Theorem [H. K. Hwang, [18]] *Let* $\{\Omega_n\}_{n\geq 1}$ *be a sequence of integral random variables. Suppose that the moment generating function satisfies the asymptotic expression*

$$M_n(s) = \mathbb{E}(e^{\Omega_n s}) = \sum_{m\geq 0} \mathbb{P}\{\Omega_n = m\} e^{ms} = e^{H_n(s)}(1 + \mathcal{O}(\kappa_n^{-1}))\,,$$

the $\mathcal{O}$*-term being uniform for* $|s| \leq \tau$, $s \in \mathbb{C}$, $\tau > 0$, *where*

(i) $H_n = u(s)\phi(n) + v(s)$, *with* $u(s)$ *and* $v(s)$ *analytic for* $|s| \leq \tau$ *and independent of* n*;*

$u''(0) \neq 0$,

(ii) $\phi(n) \to \infty$,

(iii) $\kappa_n \to \infty$.

Under these assumptions the distribution of Ω_n *is asymptotically Gaussian:*

$$\mathbb{P}\Big\{\frac{\Omega_n - u'(0)\phi(n)}{\sqrt{u''(0)\phi(n)}} < x\Big\} = \Phi(x) + \mathcal{O}\bigg(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\phi(n)}}\bigg),$$

uniformly with respect to x*,* $x \in \mathbb{R}$*. Here* $\Phi(x)$ *denotes the distribution function of the standard normal distribution* $\mathcal{N}(0,1)$*. Moreover, the mean and variance of* Ω_n *satisfy*

$$\mathbb{E}(\Omega_n) = u'(0)\phi(n) + v'(0) + \mathcal{O}(\kappa_n^{-1}), \quad \mathbb{V}(\Omega_n) = u''(0)\phi(n) + v''(0) + \mathcal{O}(\kappa_n^{-1})\,.$$

(We will use the letters u and v also in a different context in the chapter, but there is no chance of confusion.)

For fixed p and $n \to \infty$, the expected value of both, the ancestor tree, and the Steiner distance, are asymptotic to $\frac{p}{2}\log n$, the difference being in the smaller order terms. To apply the quasi power theorem, an inductive process (with respect to p) is used. Part of the difficulty is that a certain trivariate generating function is only implicitly given, and sufficient information must be "pumped out" of this implicit equation.

3.2 Size of the ancestor tree in heap ordered trees

For a given tree family let $X_{n,p}$ denote the random variable that counts the size of the ancestor tree of p randomly chosen nodes in a heap ordered tree of size n and T_n the number of such trees of size n.

A simple family of increasing trees (which includes heap ordered trees) is defined by labelled rooted trees in which labels along any branch from the root go in increasing order, see [2]. For this type of problem, it is natural to consider exponential generating functions. First, we introduce the generating functions

$$T(z) = \sum_{n\geq 0} \frac{T_n}{n!} z^n \qquad \text{and} \qquad G(z,u,v) = \sum_{n\geq 0,\, p\geq 0,\, m\geq 0} \mathbb{P}\Big\{X_{n,p} = m\Big\} T_n \binom{n}{p} \frac{z^n}{n!} u^p v^m\,, \tag{6}$$

which lead to the functional equations

$$T'(z) = \varphi(T(z)) \quad \text{and} \quad \frac{\partial}{\partial z} G(z,u,v) = v(1+u)\varphi(G(z,u,v)) + (1-v)\varphi(T(z)), \tag{7}$$

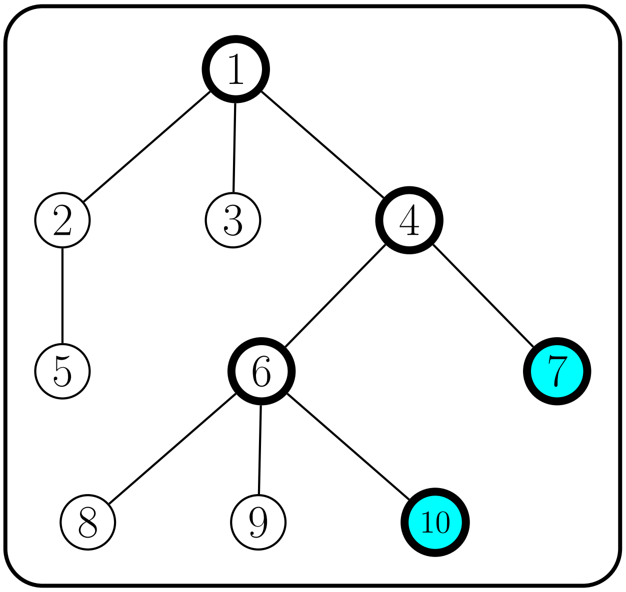


Figure 3.2: A heap ordered tree of size 10 with the ancestor tree under consideration; nodes 7 and 10 are labelled.

with initial values $T(0) = 0$ and $G(0, u, v) = 0$. The first term in (7) takes care of the instance where the root is labelled and the second term accounts for a non–labelled root. Here the *degree generating function* $\varphi(t) = \sum_{n\geq 0} \varphi_n t^n$ satisfies $\varphi_i \geq 0$ for $i \geq 1$ and $\varphi_0 > 0$. This function is responsible for the recursive generation of these trees. However, we are only concerned with the case where each degree can occur with weight one which gives the case of heap ordered trees.

Thus, we have $\varphi(t) = \frac{1}{1-t}$, and we obtain the differential equation

$$T'(z) = \frac{1}{1 - T(z)}, \quad T(0) = 0, \tag{8}$$

which gives the well-known formula

$$T(z) = 1 - \sqrt{1 - 2z}, \tag{9}$$

for the exponential generating function $T(z)$. By extracting coefficients the number of heap ordered trees is obtained,

$$T_n = \prod_{k=1}^{n-1} (2k - 1) = \frac{(n-1)!}{2^{n-1}} \binom{2n-2}{n-1}. \tag{10}$$

The differential equation of interest for $G(z, u, v)$ in the case of heap ordered trees is as follows:

3.4 Theorem [26] *The generating function for the size of the ancestor tree in heap ordered trees satisfies*

$$\frac{\partial}{\partial z}G(z,u,v) = \frac{v(1+u)}{1-G(z,u,v)} + \frac{1-v}{\sqrt{1-2z}}\,, \tag{11}$$

where $G(0,u,v)=0, \quad G(z,u,1) = 1-\sqrt{1-2z(1+u)}$.

It turns out that it is advantageous to make the following substitution in (11)

$$H(z,u,v) = \frac{1-G(z,u,v)}{\sqrt{1-2z}}\,, \tag{12}$$

so the differential equation becomes

$$H(z,u,v) - \frac{v(1+u)}{H(z,u,v)} - 1 + v = (1-2z)\frac{\partial}{\partial z}H(z,u,v), \quad H(0,u,v)=1\,. \tag{13}$$

Using separation of variables we get the implicit solution

$$\frac{1}{2}\log\frac{1}{1-2z} = \int_{x=1}^{H(z,u,v)} \frac{x\,dx}{x^2-(1-v)x-v(1+u)}\,, \tag{14}$$

and by integration it follows that

$$\begin{aligned}\log\frac{1}{1-2z} = \log\Big(1 - &\frac{(H(z,u,v)-1)(H(z,u,v)+v)}{vu}\Big)\\ &- \frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1+\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+2-(1-v)}\Big)\\ &+ \frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1-\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+(1-v)-2}\Big).\end{aligned} \tag{15}$$

Now $H(z,u,v)$ is replaced with $\frac{1-G(z,u,v)}{\sqrt{1-2z}}$ in (15) and is differentiated with respect to v. In the resulting equation we let $v=1$ and solve for $\frac{\partial}{\partial v}G(z,u,v)\big|_{v=1}$ which gives

$$\begin{aligned}\frac{\partial}{\partial v}G(z,u,v)\Big|_{v=1} =& \frac{1}{2}\sqrt{1-2z} - \frac{1}{2}\sqrt{1-2z(1+u)}\\ &- \frac{1}{4}\frac{u\Big(\log(2+u-4z(1+u)+2\sqrt{(1-2z(1+u))(1-2z)(1+u)}\Big)}{\sqrt{(1+u)(1-2z(1+u))}}\\ &- \frac{1}{4}\frac{2u\log(1+\sqrt{1+u}\,)}{\sqrt{(1+u)(1-2z(1+u))}}\,.\end{aligned} \tag{16}$$

From (16) one can find $\frac{\partial^2}{\partial z \partial v}G(z,u,v)\big|_{v=1}$ as well which will be used in the next section to compute the expectation for the Steiner distance, see (120). We differentiate equation

(15) to get

$$
\begin{aligned}
&\frac{\partial^2}{\partial z \partial v} G(z,u,v)\Big|_{v=1} \\
&= -\frac{1}{2\sqrt{1-2z}} - \frac{1+u}{2\sqrt{1-2z(1+u)}} \\
&\quad + \left(4(1+u) - \frac{2(1+u)(1-2z)(1+u) + 2(1-2z(1+u))(1+u)}{\sqrt{(1-2z(1+u))(1-2z)(1+u)}}\right) \\
&\quad \times \frac{u}{4(2+u-4z(1+u))\sqrt{(1-2z(1+u))(1+u)} + 8(1-2z(1+u))(1+u)\sqrt{1-2z}} \\
&\quad - \Bigg(\log\left(2+u-4z(1+u) + 2\sqrt{(1-2z(1+u))(1-2z)(1+u)}\,\right) \\
&\quad - \log(2+u+2\sqrt{1+u}\,)\Bigg) \frac{u\sqrt{1+u}}{4(1-2z(1+u))^{3/2}}\,.
\end{aligned} \tag{17}
$$

Next the (formal) expansions are considered

$$
G(z,u,v) = \sum_{p\geq 0} G_p(z,v) u^p \quad \text{respectively} \quad H(z,u,v) = \sum_{p \geq 0} H_p(z,v)u^p\,, \tag{18}
$$

where our aim is to describe the limiting behaviour of $[z^n]G_p(z,v)$ uniformly in a neighbourhood of $v = 1$ and then apply a central limit theorem (Hwang's quasi power theorem) to find the Gaussian limiting distribution of $X_{n,p}$ for fixed $p \geq 1$.

Obviously one has

$$
\begin{aligned}
G_p(z,v) &= \sum_{n\geq 0, m\geq 0} \mathbb{P}\Big\{X_{n,p} = m\Big\} T_n \binom{n}{p} \frac{z^n}{n!} v^m\,, \\
H_p(z,v) &= -\frac{G_p(z,v)}{\sqrt{1-2z}}\,, \quad p \geq 1\,, \\
H_0(z,v) &= \frac{1-G_0(z,v)}{\sqrt{1-2z}}\,.
\end{aligned} \tag{19}
$$

Since $\mathbb{P}\big\{X_{n,0} = m\big\} = \delta_{m,n}$, we immediately get that

$$
\begin{aligned}
G_0(z,v) &= T(z) = 1 - \sqrt{1-2z}\,, \\
H_0(z,v) &= 1\,.
\end{aligned} \tag{20}
$$

The required expansion for $p \geq 1$ is stated as the following lemma.

3.5 Lemma [26] *The coefficients $H_p(z,v)$ have for $p \geq 1$ around their (only) dominant sin-*

gularity $z = \frac{1}{2}$ uniformly for $|v-1| \le \varepsilon$ and $\varepsilon > 0$ the expansion

$$H_p(z,v) = h_p(v)\frac{1}{(1-2z)^{\frac{p(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{(p-1)(v+1)}{2}}}\Big). \tag{21}$$

The coefficient generating function $C(v,x) = \sum_{p\ge 1} h_p(v)x^p$ of the $h_p(v)$ is given implicitly by the equation

$$\frac{C(v,x)(1+v+C(v,x))}{vx} = -\Big(\frac{1+\frac{C(v,x)}{1+v}}{-\frac{1+v}{v}\frac{C(v,x)}{x}}\Big)^{\frac{1-v}{1+v}} \tag{22}$$

and it holds for

$$h_p(1) = [x^p]C(1,x) = -\frac{2}{4^p p}\binom{2(p-1)}{p-1}, \tag{23}$$

where

$$\begin{aligned} C(1,x) &= -1+\sqrt{1-x}\,, \\ C_v(1,x) &= \frac{C(1,x)}{2} + \frac{x}{4}\frac{1}{1+C(1,x)}\log\Big(\frac{1+\frac{C(1,x)}{2}}{-2\frac{C(1,x)}{x}}\Big). \end{aligned} \tag{24}$$

Thus the expansion for the $G_p(z,v)$ for $p \ge 1$ is given by

$$G_p(z,v) = -h_p(v)\frac{1}{(1-2z)^{\frac{p(v+1)-1}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{(p-1)(v+1)-1}{2}}}\Big). \tag{25}$$

Proof: To obtain $H_1(z,v)$ and thus $G_1(z,v)$ we consider (15) and compare coefficients at u^0. It follows that

$$\begin{aligned} &[u^0]\log\Big(1-\frac{(H(z,u,v)-1)(H(z,u,v)+v)}{vu}\Big) \\ &= [u^0]\log\Big(1-\frac{(H_1(z,v)u+\mathcal{O}(u^2))(1+v+\mathcal{O}(u))}{vu}\Big) = \log\Big(1-\frac{1+v}{v}H_1(z,v)\Big), \end{aligned} \tag{26}$$

$$\begin{aligned} &[u^0]\frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1+\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+2-(1-v)}\Big) \\ &= [u^0]\frac{1-v}{1+v}\frac{1}{\sqrt{1+\frac{4v}{(1+v)^2}u}}\log\Big(1+\frac{2(H_1(z,v)u+\mathcal{O}(u^2))}{(1+v)\sqrt{1+\frac{4v}{(1+v)^2}u}+1+v}\Big) \\ &= [u^0]\frac{1-v}{1+v}(1+\mathcal{O}(u))\log(1+\mathcal{O}(u)) = 0, \end{aligned} \tag{27}$$

$$\begin{aligned} &[u^0]\frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1-\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+(1-v)-2}\Big) \\ &= [u^0]\frac{1-v}{1+v}(1+\mathcal{O}(u))\log\Big(1-\frac{2(H_1(z,v)u+\mathcal{O}(u^2))}{(1+v)\sqrt{1+\frac{4v}{(1+v)^2}u}-1-v}\Big) \end{aligned}$$

$$= [u^0]\frac{1-v}{1+v}(1+\mathcal{O}(u))\log(1-\frac{1+v}{v}H_1(z,v)+\mathcal{O}(u)) = \frac{1-v}{1+v}\log\Big(1-\frac{1+v}{v}H_1(z,v)\Big), \tag{28}$$

and further

$$\log\Big(\frac{1}{1-2z}\Big) = \frac{2}{1+v}\log\Big(1-\frac{1+v}{v}H_1(z,v)\Big), \tag{29}$$

which gives

$$H_1(z,v) = \frac{v}{1+v}\Big(1-\frac{1}{(1-2z)^{\frac{v+1}{2}}}\Big) \ \text{ and } \ G_1(z,v) = \frac{\sqrt{1-2z}\,v}{1+v}\Big(\frac{1}{(1-2z)^{\frac{v+1}{2}}}-1\Big). \tag{30}$$

Therefore the asymptotic expansion given above holds for $p = 1$ (although the bound for the remainder term is not tight here) with $h_1(v) = -\frac{v}{1+v}$ and thus the stated formula for $h_p(1)$ is also valid for $p = 1$.

Now we assume that the lemma for $H_l(z,v)$, respectively $G_l(z,v)$ is true for all $1 \le l \le p$ and we will show that it then also holds for $p+1$. To prove the result for $H_{p+1}(z,v)$, we will consider the coefficients of u^p in the equation (15).

For the first term in (15), one uses the expansion

$$\log\Big(1-\frac{(H(z,u,v)-1)(H(z,u,v)+v)}{vu}\Big) = \log\Big(1-\frac{1+v}{v}H_1(z,v)\Big)+\log\Big(1-\widetilde{H}(z,u,v)\Big), \tag{31}$$

with

$$\begin{aligned}\widetilde{H}(z,u,v) &= \sum_{l\ge1}\widetilde{H}_l(z,v)u^l \\ &= \frac{1}{1-\frac{1+v}{v}H_1(z,v)}\Big(\frac{(H(z,u,v)-1)(H(z,u,v)+v)}{vu}-\frac{1+v}{v}H_1(z,v)\Big).\end{aligned} \tag{32}$$

Then we get

$$[u^p]\log\Big(1-\frac{(H(z,u,v)-1)(H(z,u,v)+v)}{vu}\Big) = -\sum_{j=1}^{p}\frac{1}{j}\sum_{\substack{p_1+\cdots+p_j=p\\ p_i\ge1}}\prod_{i=1}^{j}\widetilde{H}_{p_i}(z,v)$$

$$= -\frac{\frac{1+v}{v}H_{p+1}(z,v)}{1-\frac{1+v}{v}H_1(z,v)} - \frac{\frac{1}{v}\sum_{k=1}^{p}H_k(z,v)H_{p+1-k}(z,v)}{1-\frac{1+v}{v}H_1(z,v)} - \sum_{j=2}^{p}\frac{1}{j}\sum_{\substack{p_1+\cdots+p_j=p\\ p_i\ge1}}\prod_{i=1}^{j}\widetilde{H}_{p_i}(z,v)\,, \tag{33}$$

where

$$\widetilde{H}_l(z,v) = \frac{1}{1-\frac{1+v}{v}H_1(z,v)}\frac{1}{v}\Big((1+v)H_{l+1}(z,v) + \sum_{k=1}^{l} H_k(z,v)H_{l+1-k}(z,v)\Big). \qquad (34)$$

Under the assumptions of the lemma we now obtain, for $1 \le l \le p-1$, around the dominant singularity $z = \frac{1}{2}$ in a neighbourhood of $v = 1$, the uniform expansion

$$\begin{aligned}\widetilde{H}_l(z,v) &= (1-2z)^{\frac{v+1}{2}}\Big(\frac{\frac{1+v}{v}h_{l+1}(v)}{(1-2z)^{\frac{(l+1)(v+1)}{2}}} + \frac{\frac{1}{v}\sum_{k=1}^{l} h_k(v)h_{l+1-k}(v)}{(1-2z)^{\frac{(l+1)(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{l(v+1)}{2}}}\Big)\Big)\\ &= \widetilde{h}_l(v)\frac{1}{(1-2z)^{\frac{l(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{(l-1)(v+1)}{2}}}\Big), \qquad (35)\end{aligned}$$

where

$$\widetilde{h}_l(v) = \frac{1}{v}\Big((1+v)h_{l+1}(v) + \sum_{k=1}^{l} h_k(v)h_{l+1-k}(v)\Big). \qquad (36)$$

With the abbreviations

$$\begin{aligned}\widehat{H}(z,u,v) &= \sum_{l\ge1}\widehat{H}_l(z,v)u^l = \frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+2-(1-v)},\\ \widehat{a}_l(v) &= [u^l]\frac{1}{\sqrt{1+\frac{4v}{(1+v)^2}u}},\\ \widehat{b}_l(v) &= [u^l]\frac{2}{\sqrt{4vu+(1+v)^2}+2-(1-v)}, \qquad (37)\end{aligned}$$

the following expansion for the coefficients of the second term in (15) is obtained

$$\begin{aligned}&[u^p]\frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1+\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+2-(1-v)}\Big)\\ &= \frac{1-v}{1+v}\sum_{k=1}^{p}\widehat{a}_{p-k}(v)\sum_{j=1}^{k}\frac{(-1)^{j+1}}{j}\sum_{\substack{k_1+\cdots+k_j=k\\ k_i\ge1}}\prod_{i=1}^{j}\widehat{H}_{k_i}(z,v)\\ &= \frac{1-v}{1+v}\sum_{j=1}^{p}\frac{(-1)^{j+1}}{j}\sum_{\substack{p_1+\cdots+p_j=p\\ p_i\ge1}}\prod_{i=1}^{j}\widehat{H}_{p_i}(z,v)\\ &\quad+\frac{1-v}{1+v}\sum_{k=1}^{p-1}\widehat{a}_{p-k}(v)\sum_{j=1}^{k}\frac{(-1)^{j+1}}{j}\sum_{\substack{k_1+\cdots+k_j=k\\ k_i\ge1}}\prod_{i=1}^{j}\widehat{H}_{k_i}(z,v), \qquad (38)\end{aligned}$$

where

$$\widehat{H}_l(z,v) = \sum_{k=1}^{l} H_k(z,v)\widehat{b}_{l-k}(v)\,. \tag{39}$$

Under the assumptions of the lemma we obtain, for $1 \le l \le p$, the uniform expansion

$$\begin{aligned}\widehat{H}_l(z,v) &= \sum_{k=1}^{l}\Big(\frac{h_k(v)}{(1-2z)^{\frac{k(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{(k-1)(v+1)}{2}}}\Big)\Big)\widehat{b}_{l-k}(v)\\ &= \widehat{h}_l(v)\frac{1}{(1-2z)^{\frac{l(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{(l-1)(v+1)}{2}}}\Big)\,,\end{aligned} \tag{40}$$

where

$$\widehat{h}_l(v) = \frac{1}{1+v}h_l(v)\,. \tag{41}$$

Finally, for the third term in (15) we use the expansion

$$\begin{aligned}&\frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1-\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+(1-v)-2}\Big)\\ &= \frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1-\frac{1+v}{v}H_1(z,v)\Big) + \frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1-\overline{H}(z,u,v)\Big)\,,\end{aligned} \tag{42}$$

with

$$\begin{aligned}\overline{H}(z,u,v) &= \sum_{l\ge 1}\overline{H}_l(z,v)u^l\\ &= \frac{1}{1-\frac{1+v}{v}H_1(z,v)}\Big(\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+(1-v)-2} - \frac{1+v}{v}H_1(z,v)\Big)\,.\end{aligned} \tag{43}$$

In what follows, the abbreviations below will be employed

$$\overline{a}_l(v) = [u^l]\frac{1}{\sqrt{1+\frac{4v}{(1+v)^2}u}}, \quad \overline{b}_l(v) = [u^l]\frac{2u}{\sqrt{4vu+(1+v)^2}+(1-v)-2}\,. \tag{44}$$

We get the expansion

$$\begin{aligned}&[u^p]\frac{1-v}{\sqrt{4vu+(1+v)^2}}\log\Big(1-\frac{2(H(z,u,v)-1)}{\sqrt{4vu+(1+v)^2}+(1-v)-2}\Big)\\ &= \frac{1-v}{1+v}\overline{a}_p(v)\log\Big(1-\frac{1+v}{v}H_1(z,v)\Big)\end{aligned}$$

$$- \frac{1-v}{1+v} \sum_{k=1}^{p} \overline{a}_{p-k}(v) \sum_{j=1}^{k} \frac{(-1)^{j+1}}{j} \sum_{\substack{k_1+\cdots+k_j=k \\ k_i \geq 1}} \prod_{i=1}^{j} \overline{H}_{k_i}(z,v)$$

$$= \frac{1-v}{1+v}\overline{a}_p(v) \log\Big(1 - \frac{1+v}{v} H_1(z,v)\Big) - \frac{1-v}{1+v} \frac{\frac{1+v}{v} H_{p+1}(z,v)}{1-\frac{1+v}{v}H_1(z,v)}$$

$$- \frac{1-v}{1+v} \frac{\sum_{k=0}^{p-1} H_{k+1}(z,v)\overline{b}_{p-k}(v)}{1-\frac{1+v}{v}H_1(z,v)}$$

$$- \frac{1-v}{1+v} \sum_{j=2}^{p} \frac{1}{j} \sum_{\substack{p_1+\cdots+p_j=p \\ p_i \geq 1}} \prod_{i=1}^{j} \overline{H}_{p_i}(z,v)$$

$$- \frac{1-v}{1+v} \sum_{k=1}^{p-1} \overline{a}_{p-k}(v) \sum_{j=1}^{k} \frac{1}{j} \sum_{\substack{k_1+\cdots+k_j=k \\ k_i \geq 1}} \prod_{i=1}^{j} \overline{H}_{k_i}(z,v)\,, \tag{45}$$

where

$$\overline{H}_l(z,v) = \frac{1}{1-\frac{1+v}{v}H_1(z,v)} \sum_{k=0}^{l} H_{k+1}(z,v)\overline{b}_{l-k}(v)\,. \tag{46}$$

Now, under the assumptions of the lemma, for $1 \leq l \leq p-1$, the following uniform expansion is obtained

$$\overline{H}_l(z,v) = (1-2z)^{\frac{v+1}{2}} \sum_{k=0}^{l} \Big(\frac{h_{k+1}(v)}{(1-2z)^{\frac{(k+1)(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{k(v+1)}{2}}} \Big) \Big) \overline{b}_{l-k}(v)$$

$$= \overline{h}_l(v) \frac{1}{(1-2z)^{\frac{l(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{(l-1)(v+1)}{2}}} \Big)\,, \tag{47}$$

where

$$\overline{h}_l(v) = \frac{1+v}{v} h_{l+1}(v)\,. \tag{48}$$

Comparing coefficients leads to the following equation for $H_{p+1}(z,v)$:

$$\frac{2}{v} \frac{1}{1-\frac{1+v}{v}H_1(z,v)} H_{p+1}(z,v) =$$

$$- \frac{\frac{1}{v}\sum_{k=1}^{p} H_k(z,v) H_{p+1-k}(z,v)}{1-\frac{1+v}{v}H_1(z,v)} - \sum_{j=2}^{p} \frac{1}{j} \sum_{\substack{p_1+\cdots+p_j=p \\ p_i \geq 1}} \prod_{i=1}^{j} \widetilde{H}_{p_i}(z,v)$$

$$- \frac{1-v}{1+v} \sum_{j=1}^{p} \frac{(-1)^{j+1}}{j} \sum_{\substack{p_1+\cdots+p_j=p \\ p_i \geq 1}} \prod_{i=1}^{j} \widehat{H}_{p_i}(z,v)$$

$$
\begin{aligned}
&- \frac{1-v}{1+v} \sum_{k=1}^{p-1} \widehat{a}_{p-k}(v) \sum_{j=1}^{k} \frac{(-1)^{j+1}}{j} \sum_{\substack{k_1+\dots+k_j=k\\ k_i\ge 1}} \prod_{i=1}^{j} \widehat{H}_{k_i}(z,v) \\
&+ \frac{1-v}{1+v} \overline{a}_p(v) \log\Big(1 - \frac{1+v}{v} H_1(z,v)\Big) - \frac{1-v}{1+v} \frac{\sum_{k=0}^{p-1} H_{k+1}(z,v) \overline{b}_{p-k}(v)}{1 - \frac{1+v}{v} H_1(z,v)} \\
&- \frac{1-v}{1+v} \sum_{j=2}^{p} \frac{1}{j} \sum_{\substack{p_1+\dots+p_j=p\\ p_i\ge 1}} \prod_{i=1}^{j} \overline{H}_{p_i}(z,v) \\
&- \frac{1-v}{1+v} \sum_{k=1}^{p-1} \overline{a}_{p-k}(v) \sum_{j=1}^{k} \frac{1}{j} \sum_{\substack{k_1+\dots+k_j=k\\ k_i\ge 1}} \prod_{i=1}^{j} \overline{H}_{k_i}(z,v)\,. \qquad (49)
\end{aligned}
$$

The asymptotic expansion

$$
H_{p+1}(z,v) = h_{p+1}(v) \frac{1}{(1-2z)^{\frac{(p+1)(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{p(v+1)}{2}}}\Big), \qquad (50)
$$

follows by inspection, where

$$
\begin{aligned}
h_{p+1}(v) = \frac{v}{2}\bigg[&- \frac{1}{v} \sum_{k=1}^{p} h_k(v) h_{p+1-k}(v) - \sum_{j=2}^{p} \frac{1}{j} \sum_{\substack{p_1+\dots+p_j=p\\ p_i\ge 1}} \prod_{i=1}^{j} \widetilde{h}_{p_i}(v) \\
&- \frac{1-v}{1+v} \sum_{j=1}^{p} \frac{(-1)^{j+1}}{j} \sum_{\substack{p_1+\dots+p_j=p\\ p_i\ge 1}} \prod_{i=1}^{j} \widehat{h}_{p_i}(v) - \frac{1-v}{1+v} \sum_{j=2}^{p} \frac{1}{j} \sum_{\substack{p_1+\dots+p_j=p\\ p_i\ge 1}} \prod_{i=1}^{j} \overline{h}_{p_i}(v) \bigg], \qquad (51)
\end{aligned}
$$

and this part of the lemma is proved. The expansion of $G_p(z,v)$ given in (25) follows immediately. It should be remarked, that this detailed description of $H_{p+1}(z,v)$ shown inductively also shows that the assumptions necessary for the application of singularity analysis are satisfied. The logarithmic remainder term appears for $p = 2$ due to

$$
\log\Big(1 - \frac{1+v}{v} H_1(z,v)\Big) = -\frac{v+1}{2} \log(1-2z)\,, \qquad (52)
$$

and thus it also arises for $p \ge 2$. To get an equation for the coefficient generating function

$$
C(v,x) = \sum_{p\ge 1} h_p(v) x^p\,, \qquad (53)
$$

one could of course use equation (51), but it follows much easier directly from (15), when considering which terms give contributions to the main term of $H_p(z,v)$. Then one gets

$$\log\Big(1-\frac{\frac{C(v,x)}{x}(1+v+C(v,x))-(v+1)h_1(v)}{v}\Big)$$
$$-\frac{1-v}{1+v}\log\Big(1+\frac{C(v,x)}{1+v}\Big)+\frac{1-v}{1+v}\log\Big(1-\frac{1+v}{v}\Big(\frac{C(v,x)}{x}-h_1(v)\Big)\Big)=0, \quad (54)$$

or

$$\frac{C(v,x)(1+v+C(v,x))}{vx}=-\Big(\frac{1+\frac{C(v,x)}{1+v}}{-\frac{1+v}{v}\frac{C(v,x)}{x}}\Big)^{\frac{1-v}{1+v}}. \quad (55)$$

From (55), one obtains the equation

$$\frac{C(1,x)(2+C(1,x))}{x}=-1\,, \quad (56)$$

which gives

$$C(1,x)=-1+\sqrt{1-x} \quad \text{and} \quad (57)$$
$$h_p(1)=[x^p]C(1,x)=-\frac{2}{4^p p}\binom{2(p-1)}{p-1}, \quad \text{for } p\geq 1. \quad (58)$$

This completes the proof of the lemma. ■

Using singularity analysis, from the above lemma we obtain the following expansion, which is uniform for $|v-1|\leq\varepsilon$ and $\varepsilon>0$,

$$\begin{aligned}\sum_{m\geq 0}\mathbb{P}\{X_{n,p}=m\}v^m &= \frac{n!}{\binom{n}{p}T_n}[z^n]G_p(z,v)\\ &= -\frac{p!h_p(v)2\sqrt{\pi}}{\Gamma\Big(\frac{p(v+1)-1}{2}\Big)}n^{\frac{p(v-1)}{2}}\Big(1+\mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big)\Big)\\ &= \exp\Big[\frac{p(v-1)}{2}\log n+\log\Big(\frac{-2\sqrt{\pi}p!h_p(v)}{\Gamma\big(\frac{p(v+1)-1}{2}\big)}\Big)\Big]\Big(1+\mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big)\Big),\end{aligned} \quad (59)$$

where we used the asymptotic expansion for the number T_n of heap ordered trees

$$T_n=\frac{n!2^{n-1}n^{-\frac{3}{2}}}{\sqrt{\pi}}\Big(1+\mathcal{O}\Big(\frac{1}{n}\Big)\Big). \quad (60)$$

With the notations of the quasi power theorem, one obtains

$$u(s)=\frac{p(e^s-1)}{2} \qquad \text{and} \qquad v(s)=\log\left(\frac{-2\sqrt{\pi}p!h_p(e^s)}{\Gamma\big(\frac{p(e^s+1)-1}{2}\big)}\right). \quad (61)$$

To apply the quasi power theorem, $v(s)$ has to be analytic around $s = 0$. But this is true since

$$h_p(1) = -\frac{2}{4^p p}\binom{2(p-1)}{p-1} \neq 0\,. \tag{62}$$

Moreover, we find

$$u'(s) = \frac{p}{2}e^s, \quad u''(s) = \frac{p}{2}e^s, \qquad \text{thus} \qquad u'(0) = \frac{p}{2}, \quad u''(0) = \frac{p}{2}. \tag{63}$$

Therefore the limiting distribution for the size of the ancestor tree is obtained:

3.6 Theorem [26] *The distribution of the random variable $X_{n,p}$, which counts the size of the ancestor tree of p randomly chosen nodes in a random heap ordered tree of size n is for $p \geq 1$ asymptotically Gaussian, where the convergence rate is of order $\mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big)$:*

$$\mathbb{P}\left\{\frac{X_{n,p} - \frac{p}{2}\log n}{\sqrt{\frac{p}{2}\log n}} < x\right\} = \Phi(x) + \mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big), \tag{64}$$

and the expectation $E_{n,p} = \mathbb{E}(X_{n,p})$ and the variance $V_{n,p} = \mathbb{V}(X_{n,p})$ satisfy

$$\begin{aligned} E_{n,p} &= \frac{p}{2}\log n + v'(0) + \mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big)\,, \\ V_{n,p} &= \frac{p}{2}\log n + v''(0) + \mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big). \end{aligned} \tag{65}$$

Remark. By inspection we can get the following expansions from the first derivative of $G(z, u, v)$

$$\begin{aligned} [u^p]\frac{\partial}{\partial v}G(z,u,v)\Big|_{v=1} &= \sum_{i=1}^{p}(-1)^{p+i}(p-1)^{\underline{i-1}}\,\frac{(2i-2)!}{(i-1)!\,4^i}\,\frac{1}{(1-2z)^{i-1/2}}\,\log\frac{1}{1-2z} \\ &\quad + \sum_{i=0}^{p-1} b_i(p)\frac{1}{(1-2z)^{p-i-1/2}}. \end{aligned} \tag{66}$$

The computation of the $b_i(p)$'s is cumbersome as they become increasingly involved. However, we were able to obtain $b_1(p)$ and $b_2(p)$ explicitly:

$$\begin{aligned} b_1(p) &= 2^{-2p-1}\binom{2p}{p}\Big(H_{2p} - H_p\Big)\,, \\ b_2(p) &= -H_{2p-1}\left(2^{2p-3} + \frac{1}{2}\binom{2p-2}{p} + \binom{2p-2}{p-1}\right) + \sum_{k=0}^{p}(p+1-k)\binom{2p-2}{k}H_{2p-1-k}\,. \end{aligned} \tag{67}$$

The constant $v'(0)$ in the expectation can also be computed. One gets

$$v'(s) = \frac{h_p'(e^s)e^s}{h_p(e^s)} - \frac{p}{2}e^s\Psi\Big(\frac{p(e^s+1)-1}{2}\Big), \quad \text{thus} \quad v'(0) = \frac{h_p'(1)}{h_p(1)} - \frac{p}{2}\Psi\Big(\frac{2p-1}{2}\Big). \tag{68}$$

Here $\Psi(x)$ denotes the digamma function $\Psi(x) = (\log\Gamma(x))'$. For properties of this function we refer the reader to [1]. There remains the calculation of $h_p'(1) = [x^p]C_v(1,x)$. One gets the equation

$$\begin{aligned} C_v(1,x) &= \frac{C(1,x)}{2} + \frac{x}{4(1+C(1,x))}\log\Big(\frac{1+\frac{C(1,x)}{2}}{-2\frac{C(1,x)}{x}}\Big) \\ &= \frac{\sqrt{1-x}-1}{2} + \frac{x}{4\sqrt{1-x}}\log\Big(\frac{1+\frac{\sqrt{1-x}-1}{2}}{-2\frac{\sqrt{1-x}-1}{x}}\Big). \end{aligned} \tag{69}$$

To extract coefficients, we consider

$$\begin{aligned} [x^p]\log\Big(\frac{1+\frac{\sqrt{1-x}-1}{2}}{-2\frac{\sqrt{1-x}-1}{x}}\Big) &= \frac{1}{p}[x^{p-1}]\Big[\log\Big(\frac{1+\frac{\sqrt{1-x}-1}{2}}{-2\frac{\sqrt{1-x}-1}{x}}\Big)\Big]' \\ &= \frac{1}{p}[x^{p-1}]\Big(-\frac{1}{x\sqrt{1-x}}+\frac{1}{x}\Big) = -\frac{1}{4^p p}\binom{2p}{p}, \end{aligned} \tag{70}$$

and one finds with Lemma 3.7 (below)

$$\begin{aligned} h_p'(1) &= -\frac{1}{4^p p}\binom{2(p-1)}{p-1} - \frac{1}{4^p}\sum_{j=1}^{p-1}\frac{1}{j}\binom{2j}{j}\binom{2(p-1-j)}{p-1-j} \\ &= -\frac{1}{4^p p}\binom{2(p-1)}{p-1} - \frac{2}{4^p}\binom{2(p-1)}{p-1}(H_{2p-2}-H_{p-1}) \\ &= -\frac{1}{4^p}\binom{2(p-1)}{p-1}\Big(\frac{1}{p}+2(H_{2p-2}-H_{p-1})\Big). \end{aligned} \tag{71}$$

However, the way (69) is expressed is ungainly and the substitution $x = \frac{4t}{(1+t)^2}$ is useful for the following computations

$$\begin{aligned} C_v(1,x) &= \frac{2t}{1-t^2}\log\Big(\frac{1}{1+t}\Big) - \frac{t}{1+t}, \\ C_{vv}(1,x) &= -\frac{2t(t^2+1)}{(1-t)^3(1+t)}\log^2\Big(\frac{1}{1+t}\Big) + \frac{2t}{(1-t)^2}\log\Big(\frac{1}{1+t}\Big) + \frac{t}{1-t^2}. \end{aligned} \tag{72}$$

3.7 Lemma [26]

$$(i)\ \sum_{j\geq 1}\frac{1}{j}\binom{2j}{j}z^j = 2\log\Big(\frac{1-\sqrt{1-4z}}{2z}\Big)\,,$$
$$(ii)\ \sum_{j=1}^{p-1}\frac{1}{j}\binom{2j}{j}\binom{2(p-1-j)}{p-1-j} = \binom{2(p-1)}{p-1}(H_{2p-2}-H_{p-1})\,. \tag{73}$$

Proof: (i) It is easier to prove the equivalent result

$$\sum_{j\geq 1}\binom{2j}{j}z^{j-1} = 2\frac{d}{dz}\Big[\log\Big(\frac{1-\sqrt{1-4z}}{2z}\Big)\Big]$$
$$= \frac{4}{\sqrt{1-4z}(1-\sqrt{1-4z})} - \frac{2}{z} = \frac{1}{z}\Big[\frac{1}{\sqrt{1-4z}}-1\Big]. \tag{74}$$

Now, it is well known that

$$\sum_{j\geq 0}\binom{2j}{j}z^j = \frac{1}{\sqrt{1-4z}}, \tag{75}$$

and thus

$$\sum_{j\geq 1}\binom{2j}{j}z^{j-1} = \frac{1}{z}\Big[\frac{1}{\sqrt{1-4z}}-1\Big], \tag{76}$$

which proves the first part of the lemma.

(ii) We use the substitution

$$z=\frac{u}{(1+u)^2},\quad dz=\frac{1-u}{(1+u)^3}du,\quad \sqrt{1-4z}=\frac{1-u}{1+u}, \tag{77}$$

to simplify the given summation as follows

$$\sum_{j=1}^{p-1}\frac{1}{j}\binom{2j}{j}\binom{2(p-1-j)}{p-1-j} = [z^{p-1}]\frac{1}{\sqrt{1-4z}}\log\Big(\frac{1-\sqrt{1-4z}}{2z}\Big)^2 \tag{78}$$
$$= \frac{1}{2\pi i}\oint\frac{(1+u)^{2p-2}}{u^p}2\log(1+u)du$$
$$= [u^{p-1}]2(1+u)^{2p-2}\log(1+u)$$
$$= (-1)^p[u^{p-1}]2(1-u)^{2p-2}\log\Big(\frac{1}{1-u}\Big)$$
$$= 2(-1)^p\binom{-p}{p-1}(H_{-p}-H_{-2p+1})$$
$$= 2\binom{2p-2}{p-1}(H_{2p-2}-H_{p-1})\,. \tag{79}$$

■

We can determine the constant term $v'(0)$ in the asymptotic expansion of the expectation $E_{n,p}$ given above:

$$\begin{aligned} v'(0) &= \frac{1}{2} + p(H_{2p-2} - H_{p-1}) - \frac{p}{2}\Psi\Big(\frac{2p-1}{2}\Big) \\ &= \frac{1}{2} + p(H_{2p-2} - H_{p-1}) - \frac{p}{2}\Big(2H_{2p-2} - H_{p-1} + \Psi\Big(\frac{1}{2}\Big)\Big) = -\frac{p}{2}H_p + \frac{p}{2}\gamma + p\log 2\,. \end{aligned} \tag{80}$$

Next $v''(0)$ in the variance is computed. We obtain

$$\begin{aligned} v''(s) &= \frac{h_p''(e^s)e^{2s}}{h_p(e^s)} + \frac{h_p'(e^s)e^s}{h_p(e^s)} - \frac{(h_p'(e^s))^2e^{2s}}{h_p^2(e^s)} \\ &\quad - \frac{p}{2}e^s\Psi\left(\frac{p(e^s+1)-1}{2}\right) - \frac{p^2}{4}e^{2s}\Psi'\left(\frac{p(e^s+1)-1}{2}\right), \\ v''(0) &= \frac{h_p''(1)}{h_p(1)} + \frac{h_p'(1)}{h_p(1)} - \frac{(h_p'(1))^2}{h_p^2(1)} - \frac{p}{2}\Psi\left(\frac{2p-1}{2}\right) - \frac{p^2}{4}\Psi'\left(\frac{2p-1}{2}\right). \end{aligned} \tag{81}$$

Firstly, we are required to calculate $h_p''(1) = [x^p]C_{vv}(1,x)$, namely

$$[x^p]\Bigg(- \frac{2t(t^2+1)}{(1-t)^3(1+t)}\log^2\left(\frac{1}{1+t}\right) + \frac{2t}{(1-t)^2}\log\left(\frac{1}{1+t}\right) + \frac{t}{1-t^2}\Bigg). \tag{82}$$

We confine ourselves to considering the first few terms only. From the series expansion of (83) one can produce the local expansion around the dominant singularity $x = 1$ and use singularity analysis [9]:

$$\begin{aligned} h_p''(1) &= [x^p]\Bigg(- \frac{\log^2 2}{4}(1-x)^{-3/2} + \Big(\frac{1}{2} - \frac{\log 2}{4}\Big)(1-x)^{-1/2} + \mathcal{O}(1)\Bigg) \\ &= -\frac{\log^2 2}{4}\binom{-3/2}{p} + \Big(\frac{1}{2} - \frac{\log 2}{4}\Big)\binom{-1/2}{p} + \mathcal{O}(1). \end{aligned} \tag{83}$$

From this it follows that

$$v''(0) = -\frac{1}{2}\,p\log p + p\Big(\log 2 - \frac{5}{4}\Big) + \frac{1}{8}\log 2 + \frac{15}{16} - \frac{1}{4}\log^2 2 + \mathcal{O}(p^{-1})\,. \tag{84}$$

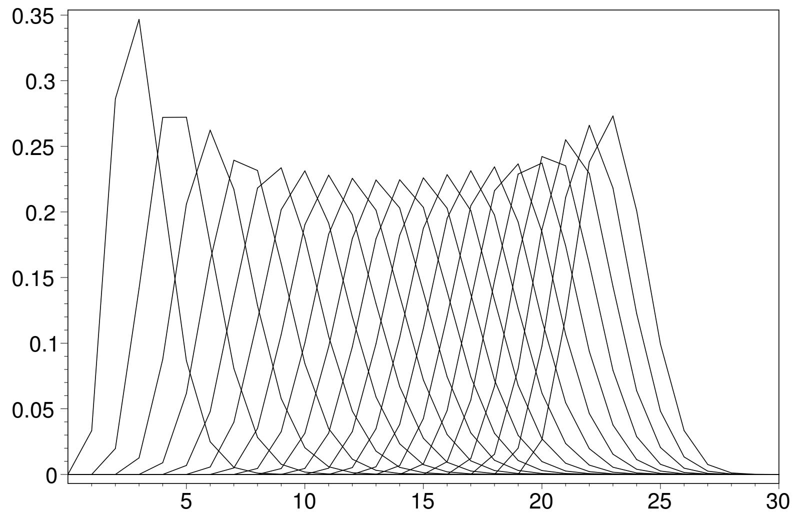


Figure 3.3: The probability distributions of the ancestor tree for $n = 30$, $p = 1, \ldots, 20$.

3.2.1 Some numerical experiments

Using numerical methods, we find an asymptotic value for the expectation of the ancestor tree. We differentiate $G(z, u, v)$ in (11) with respect to v, put $v = 1$ and use the substitutions

$$\frac{\partial}{\partial v} G(z, u, v)\Big|_{v=1} = H(z)\,, \quad G(z, u, 1) = 1 - \sqrt{1 - 2z(1+u)}\,. \tag{85}$$

The resulting equation is

$$\frac{\partial}{\partial z} H(z) = \frac{1+u}{\sqrt{1-2z(1+u)}} + \frac{(1+u)H(z)}{1-2z(1+u)} - \frac{1}{\sqrt{1+2z}}\,, \tag{86}$$

with initial condition $H(0, u) = 0$. Its solution (as computed by MAPLE) has the form

$$\begin{aligned} H(z) =& \frac{1}{2}\sqrt{1-2z} - \frac{1}{2}\sqrt{1-2z(1+u)} \\ &- \frac{1}{4}\frac{u(\log(2+u-4z(1+u) + 2\sqrt{(-1+2z(1+u))(-1+2z)(1+u)}\,)}{\sqrt{(1+u)(1-2z(1+u))}} \\ &- \frac{1}{4}\frac{2u\log(1+\sqrt{1+u}\,)}{\sqrt{(1+u)(1-2z(1+u))}}\,. \end{aligned} \tag{87}$$

Then we produce a series expansion, S, in u up to 16 terms and compute the (simplified) coefficients, a_p, of the series S, in u, for $0 \le p \le 15$ (with the following substitution in a_p:

csgn$(-1+2z)=-1$). The first few terms in the expansion of a_p are

$$
\begin{aligned}
a_2 &= \frac{-1}{8}\frac{2z^2+2z-4z\log(1-2z)+\log(1-2z)}{(-1+2z)\sqrt{1-2z}}\,,\\
a_3 &= \frac{1}{32}\frac{16z\log(1-2z)-3\log(1-2z)-32z^2\log(1-2z)-6z+26z^2+8z^3}{(-1+2z)^2\sqrt{1-2z}}\,,\\
a_4 &= \frac{1}{192}\frac{108z\log(1-2z)-15\log(1-2z)-288z^2\log(1-2z)-30z+384z^3\log(1-2z)}{(-1+2z)^3\sqrt{1-2z}}\\
&\quad+\frac{1}{192}\frac{186z^2-400z^3-60z^4}{(-1+2z)^3\sqrt{1-2z}}\,.
\end{aligned}
\tag{88}
$$

The behaviour of the a_i's can be analysed by looking at their log and non-log parts.

The Log Term. One can estimate the generating function for the log-terms in the a_i's by computing the coefficients, b_i, from the series expansion of a_i in $\log(1-2z)$ around 1,

$$
\begin{aligned}
b_2 &= -\frac{1}{8}\frac{-4z+1}{(-1+2z)\sqrt{1-2z}}\,,\\
b_3 &= \frac{1}{32}\frac{16z-3-32z^2}{(-1+2z)^2\sqrt{1-2z}}\,,\\
b_4 &= \frac{1}{192}\frac{108z-15-288z^2+384z^3}{(-1+2z)^3\sqrt{1-2z}}\,,\\
b_5 &= -\frac{1}{1536}\frac{6144z^4-960z+105+3456z^2-6144z^3}{(-1+2z)^4\sqrt{1-2z}}\,.
\end{aligned}
\tag{89}
$$

The b_i's will be simplified by multiplying them with a factor of $(1-2z)^{p-1/2}$ and then we substitute z with $\frac{Z}{2}$. Our newly obtained coefficients, c_i, have the form

$$
\begin{aligned}
c_2 &= \frac{1}{8}-\frac{1}{4}Z\,,\\
c_3 &= \frac{1}{4}Z-\frac{3}{32}-\frac{1}{4}Z^2\,,\\
c_4 &= -\frac{9}{32}Z+\frac{5}{64}+\frac{3}{8}Z^2-\frac{1}{4}Z^3\,,\\
c_5 &= -\frac{1}{4}Z^4+\frac{5}{16}Z-\frac{35}{512}-\frac{9}{16}Z^2+\frac{1}{2}Z^3\,,\\
c_6 &= -\frac{1}{4}Z^5+\frac{5}{8}Z^4-\frac{175}{512}Z+\frac{63}{1024}+\frac{25}{32}Z^2-\frac{15}{16}Z^3\,.
\end{aligned}
\tag{90}
$$

Then Z is substituted with $w+1$ in the series expansion of c_i and this yields a new series with coefficients e_i of the form

$$
e_2 = -\frac{1}{8}-\frac{1}{4}w\,,
$$

$$\begin{aligned}
e_3 &= -\frac{1}{4}w - \frac{3}{32} - \frac{1}{4}w^2\,,\\
e_4 &= -\frac{9}{32}w - \frac{5}{64} - \frac{3}{8}w^2 - \frac{1}{4}w^3\,,\\
e_5 &= -\frac{1}{4}w^4 - \frac{1}{2}w^3 - \frac{9}{16}w^2 - \frac{5}{16}w - \frac{35}{512}\,,\\
e_6 &= -\frac{1}{4}w^5 - \frac{5}{8}w^4 - \frac{15}{16}w^3 - \frac{25}{32}w^2 - \frac{175}{512}w - \frac{63}{1024}\,.
\end{aligned} \tag{91}$$

Let f_i be the coefficients of w^0 in e_i. Then, for each f_i we consider two sequences: l_1, say, containing the numbers $i = 2, \ldots, 15$ and l_2, say, containing the corresponding f_i coefficients. By using the function 'interp' in MAPLE, we compute polynomials of deg $\leq i$, in some variable X, say, which interpolate the points

$$(l_1(2), l_2(2)), (l_1(3), l_2(3)), \ldots, (l_1(i-1), l_2(i-1))\,.$$

The first few polynomials obtained with this method are:

$$\begin{aligned}
i = 2 &: -\frac{1}{8}(X-1)\,,\\
i = 3 &: -\frac{3}{64}(X-1)(X-2)\,,\\
i = 4 &: -\frac{5}{384}(X-1)(X-2)(X-3)\,,\\
i = 5 &: -\frac{35}{12288}(X-1)(X-2)(X-3)(X-4)\,,\\
i = 6 &: -\frac{21}{40960}(X-1)(X-2)(X-3)(X-4)(X-5)\,.
\end{aligned}$$

Then the list, L, containing the coefficients of X^0 in these polynomials is:

$$L = \Big[\frac{1}{4}, \frac{1}{8}, \frac{3}{64}, \frac{5}{384}, \frac{35}{12288}, \frac{21}{40960}, \frac{77}{983040}, \frac{143}{13762560}, \frac{143}{117440512}, \frac{2431}{19025362944},$$
$$\frac{46189}{3805072588800}, \frac{4199}{3986266521600}, \frac{96577}{1148044758220800}, \frac{7429}{45921790328832}, \frac{7429}{95245194756096}\Big],$$

and thus the generating function that incorporates them is given by

$$\frac{((i-1)!)^3 4^i}{(2i-2)!}\,. \tag{92}$$

Now we find the coefficients l_p from $l_p \frac{1}{(1-2z)^{p-1/2}} \log \frac{1}{1-2z}$. Since

$$l_p = \sum_{i\geq 0}^{p-1} C_{p,i}(1-2z)^i\,, \tag{93}$$

our coefficient formula becomes

$$\sum_{i\geq 0}^{p-1} \frac{C_{p,i}}{(1-2z)^{p-i-1/2}} \log\frac{1}{1-2z}\,. \tag{94}$$

The dominant term occurs when $i = 0$. The first few coefficients are

$$\begin{aligned}
C_{p,p-1} &: (-1)^p \frac{1}{4}\\
C_{p,p-2} &: (-1)^p \frac{p-1}{8}\\
C_{p,p-3} &: (-1)^p \frac{-3}{64}(p-1)^{\underline{2}}\\
C_{p,p-4} &: (-1)^p \frac{5}{384}(p-1)^{\underline{3}}\\
&\ \vdots\\
C_{p,p-i} &: (-1)^{p+i}\,(p-1)^{\underline{i-1}}\frac{(2i-2)!}{((i-1)!)^3 4^i}.
\end{aligned} \tag{95}$$

So the log term is

$$\sum_{i=1}^{p}(-1)^{p+i}(p-1)^{\underline{i-1}}\frac{(2i-2)!}{(i-1)!^3 4^i}(1-2z)^{p-i}\frac{1}{(1-2z)^{p-1/2}}\log\frac{1}{1-2z}\,. \tag{96}$$

The explicit coefficients of $[z^n]$ in the above are

$$\sum_{i=1}^{p}(-1)^{p+i}(1-p)^{\underline{i-1}}\frac{(2i-2)!}{((i-1)!)^3 4^i}\,2^n\binom{n+i-\frac{3}{2}}{n}[H_{n+i-3/2}-H_{i-3/2}]\,, \tag{97}$$

and the asymptotic ones from $i = p$ have the form

$$(p-1)^{\underline{p-1}}\frac{(2p-2)!}{((p-1)!)^3 4^p}2^n\binom{n+p-3/2}{n}[H_{n+p-3/2}-H_{p-3/2}]\,. \tag{98}$$

With MAPLE we find the leading coefficient for $i = p$

$$\frac{(p-1)!(2p-2)!}{(p-1)!^3 4^p}2^n\frac{n^{p-3/2}}{\Gamma(p-\frac{1}{2})}\log n\,, \tag{99}$$

which has to be divided by the normalising factor $\frac{1}{\binom{n}{p}}\frac{n!}{1.3\ldots(2n-3)}$. Since

$$\frac{n!}{1.3\ldots(2n-3)} \sim 2^{1-n}n^{3/2}\sqrt{\pi}\,, \tag{100}$$

and $\binom{n}{p} \sim \frac{n^{\underline{p}}}{p!} \sim \frac{n^p}{p!}$, the leading term becomes

$$\frac{(2p-2)!\,2p\,\sqrt{\pi}}{(p-1)!\,4^p\,\Gamma(p-\frac{1}{2})}\log n\,. \tag{101}$$

Now,

$$\Gamma(2z) = \frac{1}{\sqrt{2\pi}}\,2^{2z-1/2}\,\Gamma(z)\,\Gamma(z+\frac{1}{2})\,, \tag{102}$$

and thus

$$\frac{1}{\Gamma(p-\frac{1}{2})} = \frac{2^{2p-5/2}\,\Gamma(p-1)}{\sqrt{2\pi}\,\Gamma(2p-2)}\,. \tag{103}$$

Substituting into (101) gives $\frac{p}{2}$ and we conclude that the leading term of the average (parameter p) is of form

$$\frac{p}{2}\log n + K_p + \mathcal{O}\big(\frac{\log n}{n}\big)\,, \tag{104}$$

which agrees with the result obtained from the limiting distribution of the ancestor tree.

The Non-Log Term. Now we turn our attention to obtaining the generating function for the non-log terms which appeared a_i. Using a procedure similar to the one for the log-term we obtain the list of coefficients

$$L = \Big[0, \frac{1}{8}, \frac{7}{64}, \frac{37}{384}, \frac{533}{6144}, \frac{1627}{20480}, \frac{18107}{245760}, \frac{237371}{3440640}, \frac{477745}{7340032}, \frac{8161705}{132120576}, \frac{155685007}{2642411520}, \frac{156188887}{2768240640}, \frac{3602044091}{66437775360}, \frac{18051406831}{345476431872}, \frac{7751493599}{153545080832}, \frac{225175759291}{4606352424960}, \frac{13981692518567}{294806555197440}, \frac{14000078506967}{303740087173120}, \frac{98115155543129}{2186928627646464}, \frac{3634060848592973}{83103287850565632}\Big].$$

Then we use 'listtorec' and 'rectodiffeq' (from the MAPLE package GFUN, [39]) to obtain the following differential equation satisfied by the coefficients from L,

$$\begin{aligned} T'(z) =& 2f(z) + (-28+34z)\mathrm{D}(f)(z) + (16+40z^2-56z)D^2(f)(z) \\ &+ (-16z^2+8z^3+8z)D^3(f)(z)\,. \end{aligned} \tag{105}$$

The first term in the solution to the above equation is

$$T_1(u) = \frac{1}{2}\Big[\frac{\log 2}{\sqrt{1-u}} - \frac{\log(1+\sqrt{1-u})}{\sqrt{1-u}}\Big]. \tag{106}$$

To find the coefficients of u^p in this term the following substitutions will be used

$$u = \frac{4v}{(1+v)^2}, \quad du = \frac{4(1-v)}{(1+v)^3}\,dv, \quad \sqrt{1-u} = \frac{1-v}{1+v}\,, \tag{107}$$

and then the contour integration method is employed as follows

$$\begin{aligned}
&[u^p]\frac{1}{2}\Big[\frac{\log 2}{\sqrt{1-u}} - \frac{\log(1+\sqrt{1-u})}{\sqrt{1-u}}\Big] \\
&= \frac{1}{2}\frac{1}{2\pi i}\oint \frac{du}{u^{p+1}}\frac{1}{2}\Big[\frac{\log 2}{\sqrt{1-u}} - \frac{\log(1+\sqrt{1-u})}{\sqrt{1-u}}\Big] \\
&= \frac{1}{4\pi i}\oint \frac{4(1-v)}{(1+v)^3}\,dv\frac{(1+v)^{2p+2}}{4^{p+1}}\Big[\log 2\Big(\frac{1+v}{1-v}\Big) - \log\Big(\frac{2}{1+v}\Big)\frac{1+v}{1-v}\Big] \\
&= \frac{1}{4}[v^p]4^{-p}(1+v)^{2p}\log(1+v) \\
&= 2^{-2p-1}\sum_{k=1}(-1)^{k-1}\binom{2p}{p-k}. \qquad (108)
\end{aligned}$$

Alternatively, from (108), one can proceed with

$$\begin{aligned}
&\frac{1}{4}[v^p]4^{-p}(1+v)^{2p}\log(1+v) = 2^{-2p-1}(-1)^{p-1}[v^p](1-v)^{2p}\log\frac{1}{1-v} \\
&\Big(\text{since } \alpha+1 = -2p, \alpha = -2p-1, p+\alpha = -p-1, \text{for } \frac{1}{(1-v)^{\alpha+1}}\Big) \\
&= 2^{-2p-1}(-1)^{p-1}\binom{p+\alpha}{p}[H_{p+\alpha} - H_\alpha] \\
&= 2^{-2p-1}(-1)^{p-1}\binom{-p-1}{p}\Big[\frac{1}{-2p} + \cdots + \frac{1}{-(p+1)}\Big] \\
&= 2^{-2p-1}\binom{2p}{p}\Big[\frac{1}{2p} + \cdots + \frac{1}{(p+1)}\Big] \\
&= 2^{-2p-1}\binom{2p}{p}[H_{2p} - H_p]. \qquad (109)
\end{aligned}$$

Thus we can express the coefficients of u^p in (106) as

$$2^{-2p-1}\sum_{k=1}(-1)^{k-1}\binom{2p}{p-k} \quad \text{or as} \quad 2^{-2p-1}\binom{2p}{p}[H_{2p} - H_p]. \qquad (110)$$

The next term in the differential equation is

$$\frac{(-\frac{1}{8} + \frac{1}{4}\log(2))u(u-2)}{(1-u)^{3/2}} - \frac{1}{8}\frac{u(1+2\sqrt{1-u} + 2(u-2)\log(\sqrt{1-u}+1))}{(1-u)^{3/2}}. \qquad (111)$$

The list of coefficients from which (111) is obtained consists of

$$L = \Big[0, -\frac{1}{8}, -\frac{5}{16}, -\frac{3}{8}, -\frac{163}{384}, -\frac{2875}{6144}, -\frac{1299}{2560}, -\frac{133679}{245760}, -\frac{994529}{1720320}, -\frac{4479885}{7340032}, -\frac{42340055}{66060288}, -\frac{1770733217}{2642411520}, -\frac{40259683}{57671680}, -\frac{48165130403}{66437775360} - \frac{129706240867}{172738215936}, -\frac{119140740885}{153545080832}\Big].$$

Now we are in a position to compute the coefficients of u^p in (111)

$$\begin{aligned} &[u^p]\frac{(-\frac{1}{8}+\frac{1}{4}\log(2))u(u-2)}{(1-u)^{3/2}} - \frac{1}{8}\,\frac{u(1+2\sqrt{1-u}+2(u-2)\log(\sqrt{1-u}+1))}{\big(1-u\big)^{3/2}} \\ &= 2^{-2p}[v^p]\frac{(1+v)^{2p-2}(1+v^2)}{(1-v)^2}\log\frac{1}{1+v} - 2^{2p+2}[v^p]\frac{(1+v)^{2p-2}(3v+1)v}{1-v}\,, \end{aligned} \tag{112}$$

where the substitutions from (107) have been used once again. Next, the coefficients of v^p are extracted for each term above. The first one is

$$\begin{aligned} &[v^p]\frac{(1+v)^{2p-2}}{(1-v)^2}\log\frac{1}{1+v} \\ &= (-1)^p[v^p]\frac{1}{(1+v)^2}(1-v)^{2p-2}\log\frac{1}{1-v} \\ &= (-1)^p\sum_{k=0}^{p}(-1)^{p-k}(p+1-k)\binom{k-2p+1}{k}[H_{k-2p+1}-H_{-2p+1}] \\ &= (-1)^p\sum_{k=0}^{p}(-1)^{p-k}(p+1-k)\binom{-k+2p-1+k-1}{k}(-1)^k(-1)[H_{2p-1}-H_{2p-1-k}] \\ &= -\sum_{k=0}^{p}(p+1-k)\binom{2p-2}{k}[H_{2p-1}-H_{2p-1-k}] \\ &= -H_{2p-1}\sum_{k=0}^{p}(p+1-k)\binom{2p-2}{k} + \sum_{k=0}^{p}(p+1-k)\binom{2p-2}{k}H_{2p-1-k} \\ &= -H_{2p-1}\Big(2^{2p-3}+\frac{1}{2}\Big[\binom{2p-2}{p}+2\binom{2p-2}{p-1}\Big]\Big) + \sum_{k=0}^{p}(p+1-k)\binom{2p-2}{k}H_{2p-1-k} \\ &= -\Big(\sum_{k=0}^{2p-1}\frac{1}{2k-1}\Big)\Big(2^{2p-3}+\frac{1}{2}\Big[\binom{2p-2}{p}+2\binom{2p-2}{p-1}\Big]\Big) \\ &\quad + \sum_{k=0}^{p}(p+1-k)\binom{2p-2}{k}H_{2p-1-k}\,. \end{aligned} \tag{113}$$

Furthermore, the coefficients of v^p in the second term of (112) turn out to be

$$\begin{aligned}[v^p]\frac{(1+v)^{2p-2}(3v+1)v}{1-v} &= 3[v^{p-2}]\frac{1}{1-v}(1+v)^{2p-2} + [v^{p-1}]\frac{1}{1-v}(1+v)^{2p-2} \\ &= 3\sum_{i=0}^{p-2}\binom{2p-2}{i} + \sum_{i=0}^{p-1}\binom{2p-2}{i} \\ &= \frac{3}{2}\Big(2^{2p-2} - \binom{2p-2}{p-1}\Big) + \frac{1}{2}\Big(2^{2p-2} + \binom{2p-2}{p-1}\Big) \\ &= 2^{2p-1} - \binom{2p-2}{p-1}. \end{aligned} \tag{114}$$

This is substituted back into (112) to obtain the coefficients of u^p in the second term of the differential equation

$$\begin{aligned}[u^p]T_2(u) =& \frac{1}{2} - \frac{1}{2^{p+1}}\binom{2p-2}{p-1} \\ &- \Big(\sum_{k=0}^{2p-1}\frac{1}{2k-1}\Big)\Big(2^{4p-1} + 2^{2p+1}\binom{2p-2}{p} + 2^{p+2}\binom{2p-2}{p-1}\Big) \\ &+ 2^{2p+2}\sum_{k=0}^{p}(p+1-k)\binom{2p-2}{k}H_{2p-1-k}\,. \end{aligned} \tag{115}$$

3.3 The Steiner distance in heap ordered trees

Once again we are interested in finding the expectation, variance and limiting distribution for the Steiner distance. Here $Y_{n,p}$ will denote the random variable that counts the Steiner distance of p randomly chosen nodes in a heap ordered tree of size n.

For increasing trees we introduce the generating function

$$F(z,u,v) = \sum_{n\geq 0, p\geq 0, m\geq 0} \mathbb{P}\{Y_{n,p} = m\}T_n\binom{n}{p}\frac{z^n}{n!}u^p v^m\,, \tag{116}$$

which gives the functional equation

$$\begin{aligned}\frac{\partial}{\partial z}F(z,u,v) =& \varphi'(T(z))F(z,u,v) + \frac{\partial}{\partial z}G(z,u,v) - v\varphi'(T(z))G(z,u,v) \\ &- (1-v)\varphi'(T(z))T(z)\,, \end{aligned} \tag{117}$$

with initial value $F(0,u,v) = 0$. The generating functions $T(z)$ and $G(z,u,v)$ are as defined in the earlier section. The first two terms in (117) arise when the root is labelled

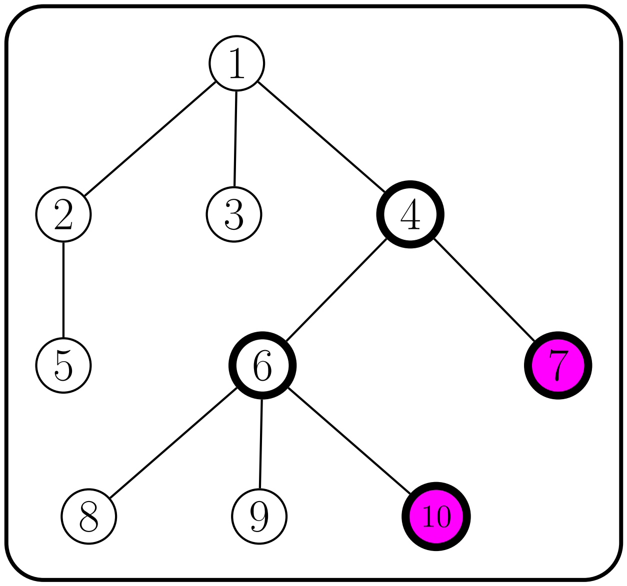


Figure 3.4: A heap ordered tree of size 10 with the Steiner distance under consideration; nodes 7 and 10 are labelled.

and the last two terms represent the corrections arising when the root is not labelled.

Again, we only look at $\varphi(t) = \frac{1}{1-t}$ which is the special case of heap ordered trees.

3.8 Lemma [26] *The generating functions $G(z,u,v)$ of the Steiner distance in heap ordered trees satisfies the following functional-differential equation*

$$\frac{\partial}{\partial z}F(z,u,v) = \frac{\partial}{\partial z}G(z,u,v) + F(z,u,v)\frac{1}{1-2z} - G(z,u,v)\frac{v}{1-2z} - \frac{1-v}{1-2z}(1-\sqrt{1-2z}\,). \tag{118}$$

This is a first order differential equation which will be solved for $F(z,u,v)$

$$F(z,u,v) = \frac{1}{\sqrt{1-2z}}\int_0^z \sqrt{1-2t}\Big[\frac{\partial}{\partial t}G(t,u,v) - G(t,u,v)\frac{v}{1-2t} - \frac{1-v}{1-\sqrt{1-2t}}\Big]dt\,. \tag{119}$$

For the expectation, $F(z,u,v)$ is differentiated with respect to v and evaluated at $v=1$

$$\frac{\partial}{\partial v}F(z,u,v)\Big|_{v=1} = \frac{1}{\sqrt{1-2z}}\int_0^z \sqrt{1-2t}\Big[\frac{\partial^2}{\partial v\partial t}G(t,u,v)\Big|_{v=1} - \frac{\partial}{\partial v}G(t,u,v)\Big|_{v=1}\frac{1}{1-2t} - \frac{1-\sqrt{1-2t(1+u)}}{1-2t} + \frac{1}{1-\sqrt{1-2t}}\Big]dt\,, \tag{120}$$

since $G(z,u,v)\big|_{v=1} = 1-\sqrt{1-2z(1+u)}$. This integration is cumbersome, so instead of performing it we find the coefficients u^p in (120) and then consider the dominant term only

$$\begin{aligned}
&[u^p]\frac{\partial}{\partial v}F(z,u,v)\Big|_{v=1}\\
&= [u^p]\frac{1}{\sqrt{1-2z}}\int_0^z \sqrt{1-2t}\Big[\frac{\partial^2}{\partial v\partial t}G_p(t,v)\Big|_{v=1} - \frac{\partial}{\partial v}G_p(t,v)\Big|_{v=1}\frac{1}{1-2t}\\
&\qquad - \frac{1-\sqrt{1-2t(1+u)}}{1-2t} + \frac{1}{1-\sqrt{1-2t}}\Big]dt\\
&= \Big(\frac{ph_p(1)\log(1-2z)}{(1-2z)^{p-1/2}} - \frac{h_p'(1)}{(1-2z)^{p-1/2}} + \frac{h_p'(1)}{(1-2z)^{1/2}}\Big) + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{p-3/2}}\Big)\\
&\qquad - [u^p]\frac{1}{\sqrt{1-2z}}\int_0^z \frac{1-\sqrt{1-2t(1+u)}}{\sqrt{1-2t}}dt\,, \qquad (121)
\end{aligned}$$

where $h_p(1)$ and $h_p'(1)$ were computed in (58) and (71) respectively. It is not difficult to see that

$$[u^p]\frac{1}{\sqrt{1-2z}}\int_0^z \frac{1-\sqrt{1-2t(1+u)}}{\sqrt{1-2t}}dt = \mathcal{O}\Big(\frac{1}{(1-2z)^{p-1/2}}\Big), \qquad (122)$$

therefore the main contribution comes from $ph_p(1)\frac{\log(1-2z)}{(1-2z)^{p-1/2}}$.

The expected value of the Steiner distance, $\mathbb{E}(Y_{n,p})$, is found by normalising (121) with $\binom{n}{p}^{-1}\frac{n!}{1\cdot 3\cdots(2n-3)}$ and then reading off the coefficient of z^n in the resulting equation. Firstly, looking at the dominant term in (121) one sees that

$$\begin{aligned}
E_{n,p} &= [z^n]\frac{\binom{n}{p}1\cdot 3\cdots(2n-3)}{n!}\,\frac{ph_p(1)\log(1-2z)}{(1-2t)^{p-1/2}}\\
&= [z^n]\frac{\binom{n}{p}1\cdot 3\cdots(2n-3)}{n!}\,\frac{-p\frac{2}{p4^p}\binom{2(p-1)}{p-1}\log(1-2z)}{(1-2z)^{p-1/2}}\\
&\sim \frac{p}{2}\log n\,, \qquad (123)
\end{aligned}$$

since we have the following

$$\begin{aligned}
[z^n]\frac{1}{(1-2z)^{p-1/2}}\log(1-2z) &= -2^n[z^n]\frac{1}{(1-z)^{p-1/2}}\log\frac{1}{1-z}\\
&= -2^n\binom{n+p-3/2}{n}(H_{n+p-3/2} - H_{p-3/2})\\
&\sim -2^n\frac{n^{p-3/2}}{\Gamma(p-\frac{1}{2})}\log n \quad (n\to\infty,\ p \text{ fixed})\,, \qquad (124)
\end{aligned}$$

as well as $\frac{n!}{1\cdot 3\cdots(2n-3)} \sim 2^{1-n}n^{3/2}\sqrt{\pi}$ and $\binom{n}{p} \sim \frac{n^p}{p!}$.

To obtain limiting theorems for the distribution of $Y_{n,p}$, we want to apply the quasi power theorem again and will therefore require for $|v-1| \le \varepsilon$ a uniform expansion of $F_p(z,v) = [u^p]F(z,u,v)$ around the dominant singularity $z = \frac{1}{2}$. From equation (119) one obtains immediately

$$F_p(z,v) = \frac{1}{\sqrt{1-2z}} \int_{t=0}^{z} \sqrt{1-2t}\Big(\frac{\partial}{\partial t} G_p(t,v) - \frac{v}{1-2t} G_p(t,v)\Big) dt. \qquad (125)$$

We will now use the following more detailed expansion of $G_p(z,v)$ which follows from the proof of Lemma 3.5:

$$G_p(z,v) = -h_p(v) \frac{1}{(1-2z)^{\frac{p(v+1)-1}{2}}} + \sum_{\substack{1 \le k \le p-1,\\ 0 \le j \le p-k}} \alpha_{p,k,j}(v) \frac{\log^j(1-2z)}{(1-2z)^{\frac{k(v+1)-1}{2}}} + \alpha_{p,0,0}(v)\sqrt{1-2z}. \qquad (126)$$

This is also used to obtain the bound for the remainder term given below. The integrand in (125) is then given by

$$\begin{aligned}
&\sqrt{1-2t}\Big(\frac{\partial}{\partial t} G_p(t,v) - \frac{v}{1-2t} G_p(t,v)\Big) \\
&= \sqrt{1-2t}\Big(\frac{-h_p(v)(p(v+1)-1)}{(1-2t)^{\frac{p(v+1)+1}{2}}} + \frac{v h_p(v)}{(1-2t)^{\frac{p(v+1)+1}{2}}} + \mathcal{O}\Big(\frac{\log(1-2t)}{(1-2t)^{\frac{(p-1)(v+1)+1}{2}}}\Big)\Big) \\
&= -\frac{h_p(v)(p-1)(v+1)}{(1-2t)^{\frac{p(v+1)}{2}}} + \mathcal{O}\Big(\frac{\log(1-2t)}{(1-2t)^{\frac{(p-1)(v+1)}{2}}}\Big),
\end{aligned} \qquad (127)$$

and for $p \ge 2$ the following expansion arises

$$F_p(z,v) = -\frac{h_p(v)(p-1)(v+1)}{p(v+1)-2} \frac{1}{(1-2z)^{\frac{p(v+1)-1}{2}}} + \mathcal{O}\Big(\frac{\log(1-2z)}{(1-2z)^{\frac{(p-1)(v+1)-1}{2}}}\Big).$$

Using singularity analysis to extract coefficients leads to

$$[z^n]F_p(z,v) = -\frac{h_p(v)(p-1)(v+1)}{p(v+1)-2} \frac{2^n n^{\frac{p(v+1)-1}{2}-1}}{\Gamma(\frac{p(v+1)-1}{2})} \Big(1 + \mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big)\Big), \qquad (128)$$

and furthermore

$$\begin{aligned}
\sum_{m \ge 0} \mathbb{P}(Y_{n,p} = m) v^m &= \frac{n!}{\binom{n}{p} T_n} [z^n] F_p(z,v) \\
&= -\frac{2\sqrt{\pi} p! (p-1)(v+1) h_p(v)}{\Gamma(\frac{p(v+1)-1}{2})(p(v+1)-2)} n^{\frac{p(v-1)}{2}} \Big(1 + \mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big)\Big).
\end{aligned} \qquad (129)$$

With the notations used in the quasi power theorem we have

$$u(s) = \frac{p(e^s-1)}{2}, \qquad v(s) = \log\left(\frac{-2\sqrt{\pi}p!(p-1)(e^s-1)h_p(e^s)}{\Gamma(\frac{p(e^s+1)-1}{2})(p(e^s+1)-2)}\right), \tag{130}$$

which gives

$$u'(0) = \frac{p}{2}, \quad u''(0) = \frac{p}{2}, \tag{131}$$

for $p \geq 2$, $v(1) \neq 0$ since $h_p(1) < 0$ and thus the quasi power theorem is applicable. On the other hand, for $p = 1$ one knows, *a priori*, from the combinatorial description, that $\mathbb{P}\{Y_{n,1} = 1\} = 1$ for $n \geq 1$.

For the constant $v'(0)$ in the expectation $E_{n,p} = \mathbb{E}(Y_{n,p})$ we compute

$$\begin{aligned} v'(s) &= \left[\log(e^s+1) + \log\big(h_p(e^s)\big) - \log\big(p(e^s+1)-2\big) - \log\Gamma\Big(\frac{p(e^s+1)-1}{2}\Big)\right]' \\ &= \frac{e^s}{e^s+1} + \frac{h_p'(e^s)e^s}{h_p(e^s)} - \frac{pe^s}{p(e^s+1)-2} - \frac{pe^s}{2}\Psi\Big(\frac{p(e^s+1)-1}{2}\Big), \end{aligned} \tag{132}$$

and further

$$v'(0) = \frac{h_p'(1)}{h_p(1)} - \frac{p}{2}\Psi\Big(\frac{2p-1}{2}\Big) - \frac{1}{2(p-1)} = -\frac{p}{2}H_p + \frac{p}{2}\gamma + p\log 2 - \frac{1}{2(p-1)}. \tag{133}$$

We note that this gives us the expected value with a higher accuracy than (123) and it leads to the following theorem.

3.9 Theorem [26] *The distribution of the random variable $Y_{n,p}$, which counts the Steiner distance of p randomly chosen nodes in a random heap ordered tree of size n is for $p \geq 2$ asymptotically Gaussian, where the convergence rate is of order $\mathcal{O}\big(\frac{1}{\sqrt{\log n}}\big)$:*

$$\mathbb{P}\left\{\frac{Y_{n,p} - \frac{p}{2}\log n}{\sqrt{\frac{p}{2}\log n}} < x\right\} = \Phi(x) + \mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big), \tag{134}$$

and the expectation $E_{n,p} = \mathbb{E}(Y_{n,p})$ and the variance $V_{n,p} = \mathbb{V}(X_{n,p})$ satisfy

$$\begin{aligned} E_{n,p} &= \frac{p}{2}\log n - \frac{p}{2}H_p + \frac{p}{2}\gamma + p\log 2 - \frac{1}{2(p-1)} + \mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big), \\ V_{n,p} &= \frac{p}{2}\log n + v''(0) + \mathcal{O}\Big(\frac{1}{n^{1-\varepsilon}}\Big). \end{aligned} \tag{135}$$

For the proof, it remains to discuss the variance. Since we have obtained the variance of the size of the ancestor tree in (84), we can easily get the variance of the Steiner distance.

It follows that

$$
\begin{aligned}
v''(s) &= \frac{e^s}{e^s+1} - \frac{e^{2s}}{e^s+1} + \frac{h_p''(e^s)e^s}{h_p(e^s)} + \frac{h_p'(e^s)e^s}{h_p(e^s)} - \frac{(h_p'(e^s))^2 e^{2s}}{h_p^2(e^s)} \\
&\quad - \frac{pe^s}{p(e^s+1)-2} + \frac{p^2e^{2s}}{(p(e^s+1)-2)^2} \\
&\quad - \frac{pe^s}{2}\Psi\left(\frac{p(e^s+1)-1}{2}\right) - \frac{p^2e^{2s}}{4}\Psi'\left(\frac{p(e^s+1)-1}{2}\right), \qquad (136)
\end{aligned}
$$

where $h_p''(1)$ is given by (83) and moreover,

$$
\begin{aligned}
v''(0) &= \frac{3}{4} + \frac{h_p''(1)}{h_p(1)} + \frac{h_p'(1)}{h_p(1)} - \frac{(h_p'(1))^2}{h_p^2(1)} - \frac{p}{2(p-1)} + \frac{p^2}{4(p-1)^2} \\
&\quad - \frac{p}{2}\Psi\left(\frac{2p-1}{2}\right) - \frac{p^2}{4}\Psi'\left(\frac{2p-1}{2}\right) \\
&= -\frac{p}{2}\log p + p\left(\log 2 - \frac{5}{4}\right) + \frac{1}{8}\log 2 + \frac{23}{16} - \frac{1}{4}\log^2 2 + \mathcal{O}(p^{-1}). \qquad (137)
\end{aligned}
$$

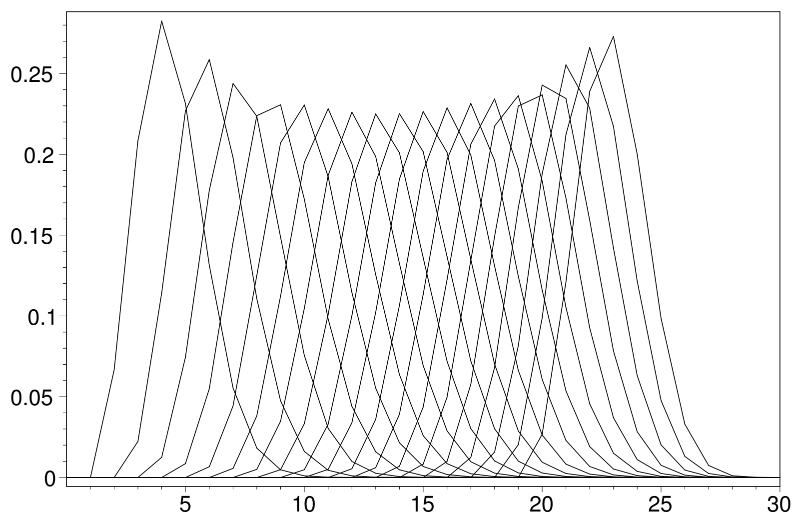


Figure 3.5: The probability distributions of the Steiner distance for $n = 30$, $p = 2, \ldots, 20$.

We have analysed heap ordered trees by using the differential equation satisfied by their generating function, a method which is described in the Preliminaries chapter. Although dealing with this equation proved to be challenging, its solution ultimately enabled us to completely characterize the properties of the parameters of interest.

4

Monotone Functions of Tree Structures

> The trees that are slow to grow bear the best fruit.
>
> Molière

4.1 Introduction

Let T be a rooted tree structure with n nodes $t_1, \ldots, t_n$. A function $f : \{t_1, \ldots, t_n\}$ into $\{1 < \ldots < k\}$ is monotone if whenever t_i is a descendant of t_j then $f(t_i) \geq f(t_j)$. Also, if $k = n$ and f is a bijection then it is called a monotone bijection. In their paper [38], Prodinger and Urbanek have determined the average number of such bijections for several classes of trees. The average height of the j-th leaf of monotonic trees with n leaves has been considered by Kirschenhofer in [20]. Further results on monotonic ordered trees appeared in [21].

The aim of this chapter is to consider the monotonic tree structures from [38] and analyse our regular parameters, namely the size of the ancestor tree and the Steiner distance. The results presented here appear in [25] and [27].

4.2 Monotonic binary trees

From Knuth [22] and Stanley [40, 41] we know that a binary tree is a finite set of nodes which either is empty or consists of a root and two binary trees called the left and right subtrees of the root.

The generating function for binary trees is

$$B(z) = 1 + zB^2(z)\,, \quad \text{with solution} \quad B(z) = \frac{1 - \sqrt{1 - 4z}}{2z}\,. \qquad (1)$$

4.1 Theorem [11] *Let B_n be the number of binary with n nodes. Then B_n is given by the*

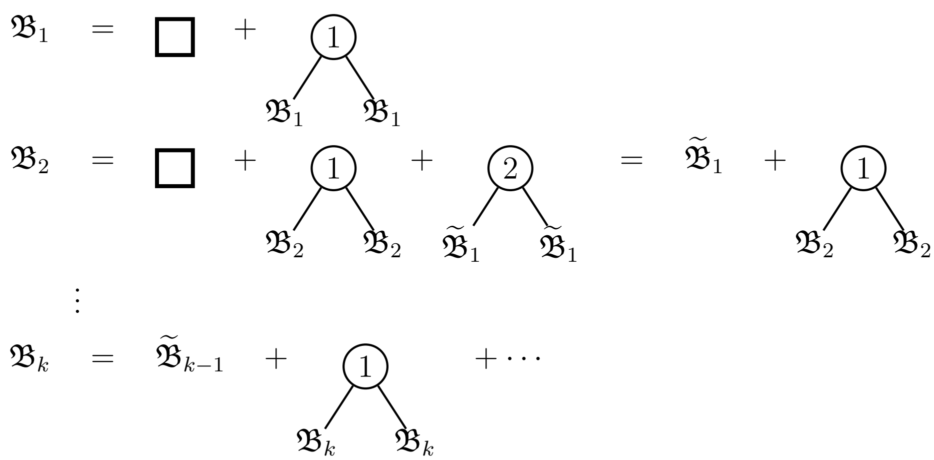
Catalan numbers

$$B_n = \frac{1}{n+1}\binom{2n}{n} = \frac{4^n}{\sqrt{\pi n^3}}\Big(1 + \mathcal{O}\Big(\frac{1}{n}\Big)\Big)\,. \tag{2}$$

We consider the binary model proposed in [38], i.e. the class, $\mathfrak{B}_k$, of binary trees whose nodes are monotonically labelled with $1, 2, \ldots, k$, with corresponding generating function

$$y_k(z) = \sum_{n\geq 0} b_n^{(k)}\, z^n\,,$$

where $b_n^{(k)}$ denotes the number of trees in $\mathfrak{B}_k$ with n nodes. Moreover, we let $\widetilde{\mathfrak{B}_k}$ be the class of binary trees whose nodes are monotonically labelled with $2, \ldots, k+1$. The generating functions of $\mathfrak{B}_k$ and $\widetilde{\mathfrak{B}}_k$ coincide. Using the terminology from Flajolet [7] and [38] we obtain that



4.2 Theorem [38] *The generating functions for monotonic binary trees are*

$$y_k(z) = y_{k-1}(z) + z y_k^2(z)\,, k \geq 1, y_0(z) = 1\,, \tag{3}$$

with solutions

$$y_k(z) = \frac{1 - \sqrt{\alpha_k(z)}}{2z}\,, \tag{4}$$

where $\alpha_1(z) = 1 - 4z$ *and* $\alpha_{k+1}(z) = -1 + 2\sqrt{\alpha_k(z)}$.

Other constants needed for our results are presented in the table below. Exact values are given for the first few cases of k.

constant	$k=1$	$k=2$	$k=3$
$c_k = 2(r_2 \cdots r_k)^{-1/4}$	2	$2\sqrt{2}$	$8/\sqrt{5}$
$p_k = c_k/(4\sqrt{\pi q_k})$	$1/\sqrt{\pi}$	$4/\sqrt{6\pi}$	$32/\sqrt{195\pi}$
$q_k = (1-r_k)/4$; $q_{k+1} = q_k(1-q_k)$	1/4	3/16	39/256
$r_k = 1 - 4q_k$	0	1/4	25/64

Since we are interested in the asymptotic behaviour of the coefficients of y_k, we have to determine the singularity q_k of y_k nearest the origin. This is given as the solution of the equation $\alpha_k(z) = 0$. As $y_k \to q_k$, from [38], we have

$$y_k = \frac{1}{2q_k} - \frac{c_k}{2q_k}(q_k - z)^{1/2} + \mathcal{O}(q_k - z)\,. \tag{5}$$

4.2.1 Size of the ancestor tree

It is perhaps easiest to begin the section with a result relating to the size of ancestor tree in binary trees and then use it to derive the corresponding one for monotonic binary trees.

4.3 Theorem [25] *The generating function for the size of the ancestor tree in binary trees satisfies*

$$B(z,u,v) = zv(1+u)B^2(z,u,v) - zvT^2(z) + T(z)\,, \tag{6}$$

where $T(z) = \frac{1-\sqrt{1-4z}}{2z}$ *is the generating function for binary trees.*

Proof: Let ϕ_p be the generating function of the size of the ancestor tree in a binary tree of size n with p nodes selected at random. Then ϕ_p satisfies the recurrence

$$\phi_p(z,v) = zv\sum_{i=0}^{p-1}\phi_i(z,v)\,\phi_{p-1-i}(z,v) + zv\sum_{i=0}^{p}\phi_i(z,v)\,\phi_{p-i}(z,v)\,, \quad p \ge 1\,, \tag{7}$$

where the first term comes from selecting the root and the second term arises when the root is not selected. Now we multiply both sides with u^p and sum over p to obtain

$$\sum_{p\ge1} u^p\,\phi_p(z,v) = zv\sum_{p\ge1}u^p\sum_{i=0}^{p-1}\phi_i(z,v)\,\phi_{p-1-i}(z,v) + zv\sum_{p\ge1}u^p\sum_{i=0}^{p}\phi_i(z,v)\,\phi_{p-i}(z,v)\,,\ p\ge1\,. \tag{8}$$

It follows that

$$\sum_{p\geq 0} u^p\,\phi_p(z,v) - \phi_0(z,v) = zv\sum_{p\geq 0} u^p \sum_{i=0}^{p-1}\phi_i(z,v)\,\phi_{p-1-i}(z,v) - zv\phi_0(z,v)\phi_0(z,v) + zv\sum_{p\geq 0} u^p \sum_{i=0}^{p}\phi_i(z,v)\,\phi_{p-i}(z,v)\,,\; p\geq 0\,. \tag{9}$$

We let $B(z,u,v) = \sum_{p\geq 0} u^p\,\phi_p(z,v)$ and $T(z) = \phi_0(z,v)$ and obtain

$$B(z,u,v) - \phi_0(z,v) = zvuB^2(z,u,v) + zvB^2(z,u,v) - zv\phi_0^2(z,v)\,. \tag{10}$$

■

We can easily modify (6) to obtain the equations of the size of the ancestor tree for the binary trees given in the Prodinger-Urbanek model. We replace $T(z)$ with y_k and $B(z,u,v)$ with $B_k(z,u,v)$.

4.4 Theorem [25] *The generating functions defining the size of the ancestor tree in monotonic binary trees are determined by*

$$B_k(z,u,v) = B_{k-1}(z,u,v) + zv(1+u)B_k^2(z,u,v) + z(1-v)y_k^2(z)\,, B_0(z,u,v) = 1\,. \tag{11}$$

The aim is to produce the expectation and variance for the size of the ancestor tree. First, we consider a few particular values for k.

Case $k = 1$

For notational convenience, the first derivative of B_1 with respect to v (evaluated at $v = 1$) is denoted by $\beta_1(z,u)$. Also, we represent $B_1(z,u,1) = y_1(z(1+u))$ by $\bar{y}_1(z)$. Then the following is obtained

$$\beta_1(z,u) = \frac{\bar{y}_1(z) - y_1(z)}{1 - 2z(1+u)\bar{y}_1(z)}\,, \tag{12}$$

where $z(1+u)y_1^2(z(1+u)) = y_1(z(1+u)) - 1$. Since we are interested in the asymptotic behaviour of the coefficients of u and z in β_1, it is appropriate to use the local expansion of y_1 (and of $\bar{y}_1$) as given in [38]

$$y_1(z) \sim \frac{1}{2q_1} - \frac{c_1}{2q_1}\,,\; z\to q_1\,, \tag{13}$$

which is substituted in the expression for β_1, with $q_1 = \frac{1}{4}$. We obtain

$$\beta_1(z,u) \sim \frac{2\sqrt{1-4z}}{\sqrt{1-4z(1+u)}}\,, \tag{14}$$

and this enables us to compute the required coefficients.

4.5 Lemma *As $z \to q_1$ the coefficients of u^p in β_1 are*

$$[u^p]\beta_1(z,u) \sim \frac{1}{2^{2p-1}}\binom{2p}{p}(1-4z)^{-p}\,. \tag{15}$$

Proof: The coefficient of u^p in (14) is computed as follows:

$$\begin{aligned}[u^p]\frac{2\sqrt{1-4z}}{\sqrt{1-4z(1+u)}} &= 2(1-4z)^{1/2}[u^p]\frac{1}{(1-4z)^{1/2}\left(1-\frac{4zu}{1-4z}\right)^{1/2}} \\ = 2(-4z)^p\binom{-\frac{1}{2}}{p}\left(1-\frac{z}{q_k}\right)^{-p} &\sim \frac{1}{2^{2p-1}}\binom{2p}{p}(1-4z)^{-p}\,,\ z \to q_1\,.\end{aligned} \tag{16}$$

■

4.6 Note *The results of the lemma above can be verified with MAPLE. From (14), one can produce a series expansion in u and extract the simplified coefficients of u. Since these now contain only powers of z, series expansions around $z = \frac{1}{4}$ (around the dominant singularity q_1) are computed for each one. The first few coefficients have the form*

$$\begin{aligned} w_2(z) &= \frac{3}{64}\left(\frac{1}{4}-z\right)^{-2} + \mathcal{O}\left(\left(\frac{1}{4}-z\right)^{-1}\right) \\ w_3(z) &= \frac{5}{512}\left(\frac{1}{4}-z\right)^{-3} + \mathcal{O}\left(\left(\frac{1}{4}-z\right)^{-2}\right) \\ w_4(z) &= \frac{35}{16384}\left(\frac{1}{4}-z\right)^{-4} + \mathcal{O}\left(\left(\frac{1}{4}-z\right)^{-3}\right) \\ w_5(z) &= \frac{63}{131072}\left(\frac{1}{4}-z\right)^{-5} + \mathcal{O}\left(\left(\frac{1}{4}-z\right)^{-4}\right). \end{aligned} \tag{17}$$

For $p = 2, \ldots, 5$, these coefficients fit perfectly the asymptotic formula given in Lemma 4.5.

The coefficients of z^n in β_1 are extracted, via Proposition 1.12 with $\alpha = -p$

$$[z^n u^p]\beta_1(z,u) \sim \frac{4^n}{2^{2p-1}}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}\,. \tag{18}$$

The expectation for the size of the ancestor tree is obtained by dividing the above with the normalising factor $p_1 q_1^{-n} n^{-3/2} \binom{n}{p} \sim \frac{1}{\sqrt{\pi}} \left(\frac{1}{4}\right)^{-n} n^{-3/2} \frac{n^p}{\Gamma(p+1)}$

$$E_{n,p}^{(1)} \sim \frac{p\sqrt{\pi}}{2^{2p-1}} \binom{2p}{p} \sqrt{n}\,. \tag{19}$$

In order to find the second moment, B_1 is differentiated twice with respect to v and then evaluated $v = 1$. If this derivative is denoted by $\Theta_1(z, u)$ then it follows that

$$\begin{aligned} \Theta_1(z,u) &= \frac{4z(1+u)\bar{y}_1(z)\beta_1(z,u) + 2z(1+u)\beta_1^2(z,u)}{1 - 2z(1+u)\bar{y}_1(z)} \\ &\sim \frac{8zu}{(1-4z(1+u))^{3/2}}\,, \quad z \to q_1\,. \end{aligned} \tag{20}$$

This asymptotic expansion will be used in analysing the behaviour of the coefficients of u^p and z^n in Θ_1.

4.7 Lemma *The coefficients of u^p in Θ_1 have the form*

$$[u^p]\Theta_1(z,u) \sim \frac{p}{2^{2p-2}} \binom{2p}{p} (1-4z)^{-p-1/2}\,, \quad z \to q_1\,. \tag{21}$$

Proof: This lemma can be easily proved by computing the coefficient of u^p from (20):

$$\begin{aligned} [u^p]\frac{8zu}{(1-4z(1+u))^{3/2}} &= 8z[u^{p-1}]\frac{1}{(1-4z)^{3/2}\left(1 - \frac{4zu}{1-4z}\right)^{3/2}} \\ = \frac{8pz(4z)^{p-1}}{2^{2p-1}} \binom{2p}{p} (1-4z)^{-p-1/2} &\sim \frac{p}{2^{2p-2}} \binom{2p}{p} (1-4z)^{-p-1/2}\,, \quad z \to q_1\,. \end{aligned} \tag{22}$$

■

4.8 Note *Once again MAPLE was employed to verify the coefficients of u^p in Θ_1. By first computing series expansions in u and then around $z = \frac{1}{4}$ from (20), we obtain*

$$\begin{aligned} e_2(z) &= \frac{3\sqrt{4}}{64}\left(\frac{1}{4} - z\right)^{-5/2} + \mathcal{O}\left(\left(\frac{1}{4} - z\right)^{-2}\right) \\ e_3(z) &= \frac{15\sqrt{4}}{1024}\left(\frac{1}{4} - z\right)^{-7/2} + \mathcal{O}\left(\left(\frac{1}{4} - z\right)^{-3}\right) \\ e_4(z) &= \frac{35\sqrt{4}}{8192}\left(\frac{1}{4} - z\right)^{-9/2} + \mathcal{O}\left(\left(\frac{1}{4} - z\right)^{-4}\right) \\ e_5(z) &= \frac{315\sqrt{4}}{262144}\left(\frac{1}{4} - z\right)^{-11/2} + \mathcal{O}\left(\left(\frac{1}{4} - z\right)^{-5}\right). \end{aligned} \tag{23}$$

Proposition 1.12 will be applied to the result in the lemma above. From this we get the coefficients of z^n

$$[z^n u^p]\Theta_1(z,u) \sim \frac{4^n p\sqrt{\pi}}{2^{2p-2}} \binom{2p}{p} \frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{24}$$

and we will normalise them to obtain the second moment

$$\frac{[z^n u^p]\Theta_1(z,u)}{\frac{1}{\sqrt{\pi}}(\frac{1}{4})^{-n} n^{-3/2} \frac{n^p}{\Gamma(p+1)}} \sim \frac{p\sqrt{\pi}}{2^{2p-2}} \frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})} n = 4pn\,. \tag{25}$$

Finally we can compute the variance of the size of the ancestor tree for the case $k = 1$

$$V^{(1)}_{n,p} = \frac{[z^n u^p]\Theta_1(z,u)}{\frac{1}{\sqrt{\pi}}(\frac{1}{4})^{-n} n^{-3/2} \frac{n^p}{\Gamma(p+1)}} + E^{(1)}_{n,p} - \left(E^{(1)}_{n,p}\right)^2 \sim \left(4\,p - \frac{\pi\, p^2}{2^{4p-2}} \binom{2p}{p}^2\right) n + \mathcal{O}(\sqrt{n})\,. \tag{26}$$

4.9 Note *As* $p \to \infty$ *we have* $4\,p - \frac{\pi\, p^2}{2^{4p-2}} \binom{2p}{p}^2 \to 0$ *in the above variance.*

Two more particular cases of k were analysed for B_k. Similar methods to the case $k = 1$ were employed to compute the usual statistics and the results are presented below.

Case $k = 2$

The derivative of B_2 with respect to v, evaluated at 1 is

$$\beta_2(z,u) \sim \frac{8(1 - \frac{16z}{3})^{1/2}}{3(1 - \frac{16z(1+u)}{3})^{1/2}}\,. \tag{27}$$

4.10 Lemma *The coefficients of* u^p *in* B_2 *are*

$$[u^p]\beta_2(z,u) \sim \frac{1}{3 \cdot 2^{2p-3}} \binom{2p}{p} \left(1 - \frac{16}{3} z\right)^{-p}, \quad z \to \frac{3}{16}\,. \tag{28}$$

Then the coefficients of z^n in β_2 have the form

$$[z^n u^p]\beta_2(z,u) \sim \frac{1}{3 \cdot 2^{2p-3}} \left(\frac{16}{3}\right)^n \binom{2p}{p} \frac{n^{p-1}}{\Gamma(p)}\,, \tag{29}$$

and upon normalisation by $\frac{4}{\sqrt{6\pi}}(\frac{16}{3})^n n^{-3/2} \frac{n^p}{\Gamma(p+1)}$, they lead to the expectation for the size of the ancestor tree

$$E^{(2)}_{n,p} \sim \frac{\sqrt{6\pi}\, p}{3 \cdot 2^{2p-1}} \binom{2p}{p} \sqrt{n}\,. \tag{30}$$

The second derivative of B_2 with respect to v, evaluated at $v = 1$ is

$$\Theta_2(z,u) \sim \frac{128\sqrt{6}zu}{27(1 - \frac{16z(1+u)}{3})^{3/2}}\,. \tag{31}$$

4.11 Lemma *The coefficients of u^p in Θ_2 have the form*

$$[u^p]\Theta_2(z,u) \sim \frac{p}{3\sqrt{3}\cdot 2^{2p-9/2}}\binom{2p}{p}\Big(1-\frac{16}{3}z\Big)^{-p-1/2}, \;\; z \to \frac{3}{16}\,. \tag{32}$$

Next, the coefficients of z^n in β_2 are

$$[z^n u^p]\Theta_2(z,u) \sim \frac{p}{3\sqrt{3}\cdot 2^{2p-9/2}}\Big(\frac{16}{3}\Big)^n\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{33}$$

and their normalisation yields the second moment

$$\frac{\sqrt{\pi}\,p}{3\cdot 2^{2p-3}}\binom{2p}{p}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})}\,n = \frac{8p}{3}\,n\,, \tag{34}$$

which leads to the variance for the size of the ancestor tree

$$V^{(2)}_{n,p} \sim \left(\frac{8p}{3} - \frac{\pi\,p^2}{3\cdot 2^{4p-3}}\binom{2p}{p}^2\right)n + \mathcal{O}(\sqrt{n}\,)\,, \tag{35}$$

where $\frac{8p}{3} - \frac{\pi\,p^2}{3\cdot 2^{4p-3}}\binom{2p}{p}^2 \to 0\,,\; p \to \infty$.

Case $k = 3$

The derivative of B_3 with respect to v, evaluated at 1 is

$$\beta_3(z,u) \sim \frac{128(1-\frac{256z}{39})^{1/2}}{39(1-\frac{256z(1+u)}{39})^{1/2}}\,. \tag{36}$$

4.12 Lemma *The coefficients of u^p in β_3 are*

$$[u^p]\beta_3(z,u) \sim \frac{1}{3\cdot 13\cdot 2^{2p-7}}\binom{2p}{p}\Big(1-\frac{256}{39}z\Big)^{-p}, \;\; z \to \frac{39}{256}\,. \tag{37}$$

Then the coefficients of z^n in β_3 have the form

$$[z^n u^p]\beta_3(z,u) \sim \frac{1}{3\cdot 13\cdot 2^{2p-7}}\Big(\frac{256}{39}\Big)^n\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}\,, \tag{38}$$

and after being normalised by $\frac{32}{\sqrt{195\pi}}(\frac{256}{39})^n n^{-3/2}\frac{n^p}{\Gamma(p+1)}$, they give the expectation for the size of the ancestor tree

$$E^{(3)}_{n,p} \sim \frac{\sqrt{195\pi}\,p}{3\cdot 13\cdot 2^{2p-2}}\binom{2p}{p}\sqrt{n}\,. \tag{39}$$

The second derivative of B_3 with respect to v (evaluated at 1) turns out to have the asymptotic expansion

$$\Theta_3(z,u) \sim \frac{65536\sqrt{5}zu}{39\sqrt{39}(1-\frac{256z(1+u)}{39})^{3/2}}\,. \tag{40}$$

4.13 Lemma *The coefficients of u^p in β_3 are*

$$[u^p]\Theta_3(z,u) \sim \frac{\sqrt{5}\cdot p}{3\sqrt{3}\cdot 13\sqrt{13}\cdot 2^{2p-9}}\binom{2p}{p}\Big(1-\frac{256}{39}z\Big)^{-p-1/2}, \; z \to \frac{39}{256}\,. \tag{41}$$

Finally, the coefficients of z^p in β_3 have the form

$$[z^n u^p]\Theta_3(z,u) \sim \frac{\sqrt{5}\cdot p}{3\sqrt{3}\cdot 13\sqrt{13}\cdot 2^{2p-9}}\Big(\frac{256}{39}\Big)^n\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{42}$$

and upon normalisation give the second moment

$$\frac{5\sqrt{\pi}\,p}{3\cdot 13\cdot 2^{2p-4}}\binom{2p}{p}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})}\,n = \frac{80p}{39}\,n\,. \tag{43}$$

This leads to the variance for the size of the ancestor tree

$$V_{n,p}^{(3)} \sim \Big(\frac{80p}{39} - \frac{5\pi\,p^2}{39\cdot 2^{4p-4}}\binom{2p}{p}^2\Big)n + \mathcal{O}(\sqrt{n}\,)\,, \tag{44}$$

where $\frac{80p}{39} - \frac{5\pi\,p^2}{39\cdot 2^{4p-4}}\binom{2p}{p}^2 \to 0\,,\; p \to \infty.$

We turn our attention to the general case. The aim is to produce the expectation and variance for the size of the ancestor tree and this is done via the usual method of differentiating the generating function, as outlined in the Preliminaries chapter. Firstly, we find the expectation by differentiating B_k with respect to v and then evaluating at $v = 1$. To simplify our notation we denote this first derivative by $\beta_k(z,u)$ and let $B_k(z,u,1) = y_k(z(1+u)) = \bar{y}_k(z)$ which yields

$$(1-2z(1+u)\bar{y}_k(z))\beta_k(z,u) = \beta_{k-1}(z,u) + \bar{y}_k(z) - \bar{y}_{k-1}(z) - y_k(z) + y_{k-1}(z)\,. \tag{45}$$

4.14 Lemma [25] *The solution of $\beta_k(z,u)$ is given by*

$$\beta_k(z,u) = \sum_{j=1}^{k}\Big[\prod_{i=j}^{k}(1-2z(1+u)\bar{y}_i(z))\Big]^{-1}\big(\bar{y}_j(z) - \bar{y}_{j-1}(z) - y_j(z) + y_{j-1}(z)\big)\,, \tag{46}$$

with $\beta_0(z) = 0$*. Hence the dominant term in this solution is*

$$\beta_k(z,u) \sim \frac{\bar{y}_k(z) - y_k(z)}{1 - 2z(1+u)\bar{y}_k(z)} \sim \frac{\sqrt{q_k - z} - \sqrt{q_k - z(1+u)}}{2q_k\sqrt{q_k - z(1+u)}} . \tag{47}$$

<u>Proof</u>: We multiply both sides of (45) with the product $\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))$ and obtain

$$\begin{aligned}
&\Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big](1-2z(1+u)\bar{y}_k(z))\beta_k(z,u) \\
&= \Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big]\beta_{k-1}(z,u) \\
&+ \Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big]\big(\bar{y}_k(z) - \bar{y}_{k-1}(z) - y_k(z) + y_{k-1}(z)\big) .
\end{aligned} \tag{48}$$

By rewriting $\Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big](1-2z(1+u)\bar{y}_k(z))\beta_k(z,u)$ as $T_k(z,u)$ we can easily solve this recursion as follows

$$\begin{aligned}
T_k(z,u) &= T_{k-1}(z,u) + \Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big]\big(\bar{y}_k(z) - \bar{y}_{k-1}(z) - y_k(z) + y_{k-1}(z)\big) , \\
T_{k-1}(z,u) &= T_{k-2}(z,u) + \Big[\prod_{i=1}^{k-2}(1-2z(1+u)\bar{y}_i(z))\Big]\big(\bar{y}_{k-1}(z) - \bar{y}_{k-2}(z) \\
&\quad - y_{k-1}(z) + y_{k-2}(z)\big) , \\
T_{k-2}(z,u) &= T_{k-3}(z,u) + \Big[\prod_{i=1}^{k-3}(1-2z(1+u)\bar{y}_i(z))\Big]\big(\bar{y}_{k-2}(z) - \bar{y}_{k-3}(z) \\
&\quad - y_{k-2}(z) + y_{k-3}(z)\big) ,
\end{aligned} \tag{49}$$

and so on. Now we can sum over the k's

$$T_k(z,u) = T_0(z,u) + \sum_{j=0}^{k-1}\Big[\prod_{i=1}^{j}(1-2z(1+u)\bar{y}_i(z))\Big]\big(\bar{y}_j(z) - \bar{y}_{j-1}(z) - y_j(z) + y_{j-1}(z)\big) , \tag{50}$$

which we write using the original $\beta_k(z,u)$

$$\Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big](1-2z(1+u)\bar{y}_k(z))\beta_k(z,u)$$
$$=\beta_0(z,u)+\sum_{j=0}^{k-1}\Big[\prod_{i=1}^{j}(1-2z(1+u)\bar{y}_i(z))\Big]\big(\bar{y}_j(z)-\bar{y}_{j-1}(z)-y_j(z)+y_{j-1}(z)\big)\,, \tag{51}$$

and dividing both sides of this equality by the product $\prod_{i=1}^{k}(1-2z(1+u)\bar{y}_i(z))$ we finally obtain our desired result

$$\beta_k(z,u)=\sum_{j=1}^{k}\Big[\prod_{i=j}^{k}(1-2z(1+u)\bar{y}_i(z))\Big]^{-1}\big(\bar{y}_j(z)-\bar{y}_{j-1}(z)-y_j(z)+y_{j-1}(z)\big)\,. \tag{52}$$

Now the dominant term occurs for $i=j=k$. Since the singularities of $y_{k-1}(z)$ and $\bar{y}_{k-1}(z)$ are further away from the origin than those of $y_k(z)$ and $\bar{y}_k(z)$ they will not contribute to the main term in $\beta_k(z,u)$ and we do not include them. Thus the dominant term in $\beta_k(z,u)$ is as given in the statement of the theorem. ■

4.15 Note *From (45) we also observe that*

$$\beta_k(z,u)=\beta_{k-1}(z,u)+\bar{y}_k(z)-\bar{y}_{k-1}(z)+2z(1+u)\bar{y}_k(z)\beta_k(z,u)-y_k(z)+y_{k-1}(z)\,, \tag{53}$$

and thus

$$\beta_k(z,u)=\frac{\beta_{k-1}(z,u)}{1-2z(1+u)\bar{y}_k(z)}+\frac{y_{k-1}(z)-\bar{y}_{k-1}(z)}{1-2z(1+u)\bar{y}_k(z)}+\frac{\bar{y}_k(z)-y_k(z)}{1-2z(1+u)\bar{y}_k(z)}\,. \tag{54}$$

4.16 Lemma *As $z\to q_k$, the coefficients of u^p in $\beta_k(z,u)$ have the form*

$$[u^p]\beta_k(z,u)\sim\frac{1}{2^{2p+1}\,q_k}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p}. \tag{55}$$

Proof: We look at the coefficient of u^p in (47)

$$[u^p]\frac{\sqrt{q_k-z}}{2q_k\sqrt{q_k-z(1+u)}}=\frac{\Big(1-\frac{z}{q_k}\Big)^{1/2}}{2q_k}[u^p]\frac{1}{\Big(1-\frac{z}{q_k}\Big)^{1/2}\Big(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\Big)^{1/2}}$$
$$=\frac{1}{2q_k}\Big(-\frac{z}{q_k}\Big)^p\binom{-\frac{1}{2}}{p}\Big(1-\frac{z}{q_k}\Big)^{-p}=\frac{z^p}{2^{2p+1}q_k^{p+1}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p}$$

$$\sim \frac{1}{2^{2p+1} q_k} \binom{2p}{p} \left(1 - \frac{z}{q_k}\right)^{-p}, \ z \to q_k \, . \tag{56}$$

■

In order to obtain the first moment the coefficients of z^n in $\beta_k(z,u)$ are also needed. This is done by using Proposition 1.12 with $\alpha = -p$, which yields

$$[z^n u^p]\beta_k(z,u) \sim \frac{1}{2^{2p+1}\, q_k^{n+1}} \binom{2p}{p} \frac{n^{p-1}}{\Gamma(p)} \, . \tag{57}$$

Finally, these coefficients will be normalised by $p_k q_k^{-n} n^{-3/2}\binom{n}{p} \sim p_k q_k^{-n} n^{-3/2}\frac{n^p}{\Gamma(p+1)}$.

4.17 Theorem [25] *The expectation for the size of the ancestor tree in monotonic binary trees is*

$$E_{n,p}^{(k)} \sim \frac{p}{2^{2p+1}\, p_k\, q_k} \binom{2p}{p} \sqrt{n} \, , n \to \infty \, , \ \text{fixed } k \, . \tag{58}$$

Higher order statistics can be computed in the same fashion, so for the variance of the size of the ancestor tree, B_k is differentiated twice with respect to v and evaluated at $v = 1$. For convenience, we denote this derivative by $\Theta_k(z,u)$, which gives

$$\big(1 - 2z(1+u)\bar{y}_k(z)\big)\Theta_k(z,u) = \Theta_{k-1}(z,u) + 4z(1+u)\bar{y}_k(z)\beta_k(z,u) + 2z(1+u)\beta_k^2(z,u) \, , \tag{59}$$

4.18 Lemma [27] *The solution to $\Theta_k(z,u)$ has the closed form*

$$\Theta_k(z,u) = \sum_{j=1}^{k} \Big[\prod_{i=j}^{k} (1 - 2z(1+u)\bar{y}_i(z))\Big]^{-1} \Big(4z(1+u)\bar{y}_j(z)\beta_j(z,u) + 2z(1+u)\beta_j^2(z,u)\Big) \, , \tag{60}$$

and the dominant term is asymptotically

$$\Theta_k \sim \frac{4z(1+u)\bar{y}_k\, \beta_k + 2z(1+u)\beta_k^2(z,u)}{1 - 2z(1+u)\bar{y}_k} \, . \tag{61}$$

Proof: First, multiply (59) with the product $\prod_{i=1}^{k-1} (1 - 2z(1+u)\bar{y}_i(z))$. We obtain

$$\begin{aligned}
&\Big[\prod_{i=1}^{k-1} (1 - 2z(1+u)\bar{y}_i(z))\Big] \big(1 - 2z(1+u)\bar{y}_k(z)\big)\Theta_k(z,u) \\
&= \Big[\prod_{i=1}^{k-1} (1 - 2z(1+u)\bar{y}_i(z))\Big] \Theta_{k-1}(z,u)
\end{aligned}$$

$$+ \Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big]\Big(4z(1+u)\bar{y}_k(z)\beta_k(z,u)+2z(1+u)\beta_k^2(z,u)\Big). \qquad (62)$$

By rewriting $\Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big]\big(1-2z(1+u)\bar{y}_k(z)\big)\Theta_k(z,u)$ as $S_k(z,u)$ we can easily solve this recursion as follows

$$\begin{aligned} S_k(z,u) &= S_{k-1}(z,u) + \Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big]\Big(4z(1+u)\bar{y}_k(z)\beta_k(z,u) \\ &\quad + 2z(1+u)\beta_k^2(z,u)\Big), \\ S_{k-1}(z,u) &= S_{k-2}(z,u) + \Big[\prod_{i=1}^{k-2}(1-2z(1+u)\bar{y}_i(z))\Big]\Big(4z(1+u)\bar{y}_{k-1}(z)\beta_{k-1}(z,u) \\ &\quad + 2z(1+u)\beta_{k-1}^2(z,u)\Big), \\ S_{k-2}(z,u) &= S_{k-3}(z,u) + \Big[\prod_{i=1}^{k-3}(1-2z(1+u)\bar{y}_i(z))\Big]\Big(4z(1+u)\bar{y}_{k-2}(z)\beta_{k-2}(z,u) \\ &\quad + 2z(1+u)\beta_{k-2}^2(z,u)\Big), \end{aligned} \qquad (63)$$

and so on. Now we can sum over the k's

$$\begin{aligned} S_k(z,u) =& S_0(z,u) + \sum_{j=0}^{k-1}\Big[\prod_{i=1}^{j}(1-2z(1+u)\bar{y}_i(z))\Big]\Big(4z(1+u)\bar{y}_j(z)\beta_j(z,u) \\ & + 2z(1+u)\beta_j^2(z,u)\Big), \end{aligned} \qquad (64)$$

which we write using the original $\Theta_k(z,u)$

$$\begin{aligned} &\Big[\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))\Big]\big(1-2z(1+u)\bar{y}_k(z)\big)\Theta_k(z,u) \\ &= \sum_{j=0}^{k-1}\Big[\prod_{i=1}^{j}(1-2z(1+u)\bar{y}_i(z))\Big]\Big(4z(1+u)\bar{y}_j(z)\beta_j(z,u)+2z(1+u)\beta_j^2(z,u)\Big), \end{aligned} \qquad (65)$$

since $S_0(z,u)=0$, $\Theta_0(z,u)=0$, and dividing both sides of this equality by the product $\prod_{i=1}^{k-1}(1-2z(1+u)\bar{y}_i(z))$ we obtain the desired result

$$\Theta_k(z,u) = \sum_{j=1}^{k}\Big[\prod_{i=j}^{k}(1-2z(1+u)\bar{y}_i(z))\Big]^{-1}\Big(4z(1+u)\bar{y}_j(z)\beta_j(z,u)+2z(1+u)\beta_j^2(z,u)\Big). \qquad (66)$$

By letting $i=j=k$ in this solution the dominant term arises. ■

Using the known local asymptotic expansions for $\bar{y}_k$ and y_k, we can rewrite (61) as

$$\begin{aligned}\Theta_k(z,u) &\sim \frac{4z(1+u)\bar{y}_k\left(\frac{\bar{y}_k-y_k}{1-2z(1+u)\bar{y}_k}\right)+2z(1+u)\left(\frac{\bar{y}_k-y_k}{1-2z(1+u)\bar{y}_k}\right)^2}{1-2z(1+u)\bar{y}_k}\\ &\sim \frac{4q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)}{1-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)}+\frac{2q_k\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)^2}{1-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)}\\ &= \frac{1}{q_k}-\frac{\sqrt{q_k-z}}{q_k\sqrt{q_k-z(1+u)}}+\frac{zu}{2c_kq_k(q_k-z(1+u))^{3/2}}\,,\end{aligned} \tag{67}$$

and this enables us to look at the asymptotic behaviour of the coefficients.

4.19 Lemma *The coefficients of u^p in $\Theta_k(z,u)$ have the form*

$$\begin{aligned}[u^p]\Theta_k &\sim \frac{p}{2^{2p}\,c_k\,q_k^{3/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2}\\ &= \frac{p}{2^{2p+2}\sqrt{\pi}\,p_k\,q_k^2}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2},\quad c_k=4p_k\sqrt{\pi\,q_k}\,,\ \ z\to q_k\,.\end{aligned} \tag{68}$$

Proof: The third term in (67) gives the main contribution. So the coefficient of u^p there is

$$\begin{aligned}&[u^p]\frac{zu}{2c_kq_k(q_k-z(1+u))^{3/2}} = \frac{z}{2c_kq_k^{5/2}}[u^{p-1}]\frac{1}{\left(1-\frac{z}{q_k}\right)^{3/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{3/2}}\\ &= \frac{pz^p}{2^{2p}c_kq_k^{p+3/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2} \sim \frac{p}{2^{2p}\,c_k\,q_k^{3/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2},\quad z\to q_k\,.\end{aligned} \tag{69}$$

■

Once again Proposition 1.12 with $\alpha=-p-1/2$ is applied to get the coefficients of z^n in Θ_k

$$[z^nu^p]\Theta_k \sim \frac{p}{2^{2p+2}\sqrt{\pi}\,p_k\,q_k^{n+2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{70}$$

and they are used in computing the second moment.

4.20 Theorem [25] *The variance for the size of the ancestor tree in monotonic binary trees is*

$$V_{n,p}^{(k)} = \frac{[z^nu^p]\Theta_k(z,u)}{p_kq_k^{-n}\frac{n}{\Gamma(p+1)}n^{-3/2}}+E_{n,p}^{(k)}-(E_{n,p}^{(k)})^2 = \frac{p}{4p_k^2q_k^2}\left(\frac{1}{\pi}-\frac{p}{2^{4p}}\binom{2p}{p}^2\right)n+\mathcal{O}(\sqrt{n})\,, \tag{71}$$

for $n\to\infty$ and fixed k.

One notes that $\frac{1}{\pi} - \frac{p}{2^{4p}}\binom{2p}{p}^2 \to 0$ as $p \to \infty$.

4.2.2 The limiting distribution for the size of the ancestor tree

Let $X_{n,p}$ be the random variable which counts the size of the ancestor tree of p randomly chosen nodes in a monotonic binary tree of size n. We study the probabilities

$$\mathbb{P}\{X_{n,p} = m\} = \frac{[z^n u^p v^m] B_k(z,u,v)}{b_n^{(k)} \binom{n}{p}}\,, \tag{72}$$

where $b_n^{(k)} \sim p_k q_k^{-n} n^{-3/2}$ is the number of monotonic binary trees which was computed in [38]. The first step in obtaining our limiting distribution is to find all the moments for B_k. If we differentiate B_k M times with respect to v then using Leibniz's rule we obtain

$$\begin{aligned}\frac{\partial^M B_k(z,u,v)}{\partial v^M} &= \frac{\partial^M B_{k-1}(z,u,v)}{\partial v^M} \\ &\quad + Mz(1+u)\sum_{i=0}^{M-1}\binom{M-1}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\frac{\partial^{M-i-1} B_k(z,u,v)}{\partial v^{M-i-1}} \\ &\quad + zv(1+u)\sum_{i=0}^{M}\binom{M}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\frac{\partial^{M-i} B_k(z,u,v)}{\partial v^{M-i}}\,,\ \ M \geq 2\,,\end{aligned} \tag{73}$$

and it follows that

$$\begin{aligned}&\frac{\partial^M B_k(z,u,v)}{\partial v^M}(1 - 2zv(1+u)B_k(z,u,v)) \\ &= \frac{\partial^M B_{k-1}(z,u,v)}{\partial v^M} + Mz(1+u)\sum_{i=0}^{M-1}\binom{M-1}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\frac{\partial^{M-i-1} B_k(z,u,v)}{\partial v^{M-i-1}} \\ &+ zv(1+u)\sum_{i=1}^{M-1}\binom{M}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\frac{\partial^{M-i} B_k(z,u,v)}{\partial v^{M-i}}\,,\ \ M \geq 2\,.\end{aligned} \tag{74}$$

This recursive relation is now used to solve for the M-th derivative of B_k and the latter is evaluated at a particular point.

4.21 Lemma [27] *The M-th derivative with respect to v in $B_k(z,u,v)$, evaluated at $v = 1$ has the form*

$$\begin{aligned}\left.\frac{\partial^M B_k(z,u,v)}{\partial v^M}\right|_{v=1} &= \sum_{j=1}^{k}\Big[\prod_{l=j}^{k}(1 - 2z(1+u)\bar{y}_l(z))\Big]^{-1} \times \\ &\qquad \Big[Mz(1+u)\sum_{i=0}^{M-1}\binom{M-1}{i}\left.\frac{\partial^i B_j(z,u,v)}{\partial v^i}\right|_{v=1}\left.\frac{\partial^{M-i-1} B_j(z,u,v)}{\partial v^{M-i-1}}\right|_{v=1}\end{aligned}$$

$$+ z(1+u)\sum_{i=1}^{M-1}\binom{M}{i}\frac{\partial^i B_j(z,u,v)}{\partial v^i}\Big|_{v=1}\frac{\partial^{M-i}B_j(z,u,v)}{\partial v^{M-i}}\Big|_{v=1}\Bigg],\ M\geq 2\,, \tag{75}$$

and its dominant term is

$$\begin{aligned}\frac{\partial^M B_k(z,u,v)}{\partial v^M}\Big|_{v=1} \sim \Big[(1-2z(1+u)\bar{y}_k(z))\Big]^{-1}\times \\ \Bigg[Mz(1+u)\sum_{i=0}^{M-1}\binom{M-1}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\Big|_{v=1}\frac{\partial^{M-i-1}B_k(z,u,v)}{\partial v^{M-i-1}}\Big|_{v=1} \\ + z(1+u)\sum_{i=1}^{M-1}\binom{M}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\Big|_{v=1}\frac{\partial^{M-i}B_k(z,u,v)}{\partial v^{M-i}}\Big|_{v=1}\Bigg].\end{aligned} \tag{76}$$

Proof: We multiply both sides of (74) with $\prod_{l=1}^{k-1}(1-2z(1+u)\bar{y}_l(z))$ and proceed with iterating the result. The method of proof is similar to the one for Lemma 4.18 so we do not give the details here. ■

The required moments are produced by analysing the behaviour of the coefficients in (76) near their dominant singularities q_k. For the coefficients of u^p we find that

$$[u^p]\frac{\partial^M B_k(z,u,v)}{\partial v^M}\Big|_{v=1} \sim \begin{cases}\dfrac{\prod_{i=0}^{M/2-1}(p+i)\prod_{i=1}^{M/2-1}(2p+2i-1)}{2^{2p-M/2+1}q_k^{(M+1)/2}c_k^{M-1}}\dbinom{2p}{p}\left(1-\dfrac{z}{q_k}\right)^{-p-(M-1)/2}, & M\text{ even}\,,\\[2ex] \dfrac{\prod_{i=0}^{(M-1)/2-1}(p+i)\prod_{i=1}^{(M-1)/2}(2p+2i-1)}{2^{2p-(M-1)/2+1}q_k^{(M+1)/2}c_k^{M-1}}\dbinom{2p}{p}\left(1-\dfrac{z}{q_k}\right)^{-p-(M-1)/2}, & M\text{ odd}\,,\end{cases} \tag{77}$$

and they simplify to the following

$$[u^p]\frac{\partial^M B_k(z,u,v)}{\partial v^M}\Big|_{v=1} \sim \frac{\Gamma(2p+M-1)}{p2^{2p}q_k^{(M+1)/2}c_k^{M-1}(\Gamma(p))^2}\left(1-\frac{z}{q_k}\right)^{-p-(M-1)/2}. \tag{78}$$

Moreover, the coefficients of z^n are

$$[z^n u^p]\frac{\partial^M B_k(z,u,v)}{\partial v^M}\Big|_{v=1} \sim \frac{2^{M-2}\Gamma(p+\frac{M}{2})}{p\sqrt{\pi}q_k^{(M+1)/2+n}c_k^{M-1}(\Gamma(p))^2}n^{p-(3-M)/2}\,, \tag{79}$$

which (after normalisation) lead to the next result.

4.22 Theorem [27] *The M-th moments, $\mathbb{E}(X_{n,p}^M)$, of the size of the ancestor tree of p randomly chosen nodes in a monotonic binary tree of size n are asymptotically given by*

$$\mathbb{E}(X_{n,p}^M) = \frac{[z^n u^p]\frac{\partial^M B_k(z,u,v)}{\partial v^M}\Big|_{v=1}}{p_k q_k^{-n}\frac{n^p}{\Gamma(p+1)}n^{-3/2}} = \frac{2^{M-2}}{\sqrt{\pi}c_k^{M-1}q_k^{(M+1)/2}p_k}\frac{\Gamma\left(p+\frac{M}{2}\right)}{\Gamma(p)}n^{M/2}\Big(1+\mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big)\Big), \tag{80}$$

for $M \geq 2$, fixed $p \geq 1$ and $n \to \infty$, where $c_k = 2\Big(\prod_{i=2}^{k}(1-4q_i)\Big)^{-1/4}$ and $p_k = \frac{c_k}{4\sqrt{\pi q_k}}$.

Proof: Induction on M is employed. Let $M = 2$, then the left hand side of (80) is

$$\begin{aligned}\frac{[z^n u^p]\frac{\partial^2 B_k(z,u,v)}{\partial v^2}\Big|_{v=1}}{p_k q_k^{-n}p_k\frac{n^p}{\Gamma(p+1)}n^{-3/2}} &= \frac{[z^n u^p]\frac{\left(4z(1+u)\bar{y}_k(z)\frac{\partial B_k(z,u,v)}{\partial v}\Big|_{v=1}+2z(1+u)\left(\frac{\partial B_k(z,u,v)}{\partial v}\Big|_{v=1}\right)^2\right)}{1-2z(1+u)\bar{y}_k(z)}}{p_k q_k^{-n}\frac{n^p}{\Gamma(p+1)}n^{-3/2}} \\ &= \frac{p}{4\pi p_k^2 q_k^2} + \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big),\end{aligned} \tag{81}$$

as computed earlier, which also equals the right hand side of (80). Now, assume (80) is true for M and prove it for $M+1$. We have

$$\begin{aligned}&\frac{1}{p_k q_k^{-n}\frac{n^p}{\Gamma(p+1)}n^{-3/2}}[z^n u^p]Mz(1+u)\sum_{i=0}^{M-1}\binom{M-1}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\Big|_{v=1}\frac{\partial^{M-i-1}B_k(z,u,v)}{\partial v^{M-i-1}}\Big|_{v=1} \\ &\sim \frac{2^{M-5}\Gamma(M+1)}{\pi c_k^{M-3}q_k^{(M-1)/2}p_k^2(\Gamma(p))^2}\sum_{i=0}^{M-1}\frac{\Gamma(p+\frac{i}{2})\Gamma(p+\frac{M-i-1}{2})}{\Gamma(i+1)\Gamma(M-i)},\end{aligned} \tag{82}$$

as well as

$$\begin{aligned}&\frac{1}{p_k q_k^{-n}\frac{n^p}{\Gamma(p+1)}n^{-3/2}}[z^n u^p]z(1+u)\sum_{i=1}^{M-1}\binom{M}{i}\frac{\partial^i B_k(z,u,v)}{\partial v^i}\Big|_{v=1}\frac{\partial^{M-i}B_k(z,u,v)}{\partial v^{M-i}}\Big|_{v=1} \\ &\sim \frac{2^{M-4}\Gamma(M+1)}{\pi c_k^{M-2}q_k^{M/2}p_k^2(\Gamma(p))^2}\sum_{i=1}^{M-1}\frac{\Gamma(p+\frac{i}{2})\Gamma(p+\frac{M-i}{2})}{\Gamma(i+1)\Gamma(M-i+1)}.\end{aligned} \tag{83}$$

■

It is well-known that the density function of the generalised Gamma distribution is

$$g(a,h,A;x) = \frac{|h|}{\Gamma(a)A}\Big(\frac{x}{A}\Big)^{ah-1}e^{-\left(\frac{x}{A}\right)^h}, \tag{84}$$

for $x > 0$ and it will be used for the derivation of our limiting distribution.

4.23 Theorem [27] *The probability, $\mathbb{P}\{X_{n,p} = m\}$, that the size of the ancestor tree of p randomly selected nodes in a monotonic binary tree of size n is equal to m is asymptotically*

$$\mathbb{P}\{X_{n,p} = m\} = m^{p-1} \frac{q_k^{p/2}}{n^p (p-1)!} e^{-q_k^{1/2} \frac{m}{n}} \left(1 + \mathcal{O}\Big(\frac{1}{\sqrt{n}}\Big) + \mathcal{O}\Big(\frac{n}{m}\Big) + \mathcal{O}\Big(\frac{1}{m}\Big)\right), \qquad (85)$$

for $p \geq 1$, $n \to \infty$ and $m \leq K\sqrt{n}$ ($K > 0$ arbitrary but fixed).

The above result leads to the following theorem which characterises the limiting distribution for the size of the ancestor trees in monotonic binary trees.

4.24 Theorem [27] *For fixed $p \geq 1$, $x > 0$, $m = x\sqrt{n}$ we have*

$$\sqrt{n}\, \mathbb{P}\{X_{n,p} = m\} \sim x^{p-1} \frac{q_k^{p/2}}{(p-1)!} e^{-q_k^{1/2} x} = g\Big(p, 1, \frac{1}{\sqrt{q_k}}; x\Big). \qquad (86)$$

Thus the limiting distribution of the normalised random variable $\frac{X_{n,p}}{\sqrt{n}}$ is asymptotically, for fixed $p \geq 1$ and $n \to \infty$, a generalised Gamma distribution with parameters $\Big(p, 1, \frac{1}{\sqrt{q_k}}\Big)$.

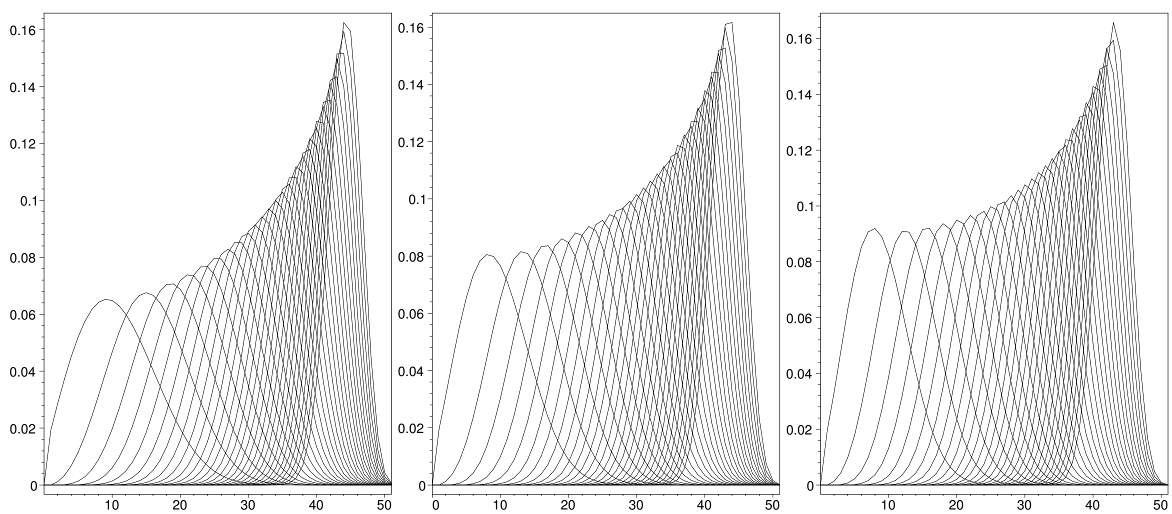


Figure 4.1: The probability distributions for the size of the ancestor tree in monotonic binary trees for $n = 51$, $p = 1 \ldots 30$ and $k = 1, 2, 3$.

4.2.3 The Steiner distance

Our second parameter of interest is now analysed. The generating function of the Steiner distance in ordinary binary trees is presented first and then modified suitably for the monotonic case.

4.25 Theorem [25] *The generating function for the Steiner distance in binary trees is*

$$C(z,u,v) = \frac{zv(1+u)B^2(z,u,v) - 2zv\,T(z)\,B(z,u,v) + zT^2(z)(v-2) + T(z)}{1-2z\,T(z)} \,. \qquad (87)$$

Proof: One starts with the recurrence

$$\psi_p(z,v) = zv\sum_{i=0}^{p-1}\phi_i(z,v)\,\phi_{p-1-i}(z,v) + zv\sum_{i=1}^{p-1}\phi_i(z,v)\,\phi_{p-i}(z,v) + 2z\,\psi_0(z,v)\,\psi_p(z,v)\,, p \geq 1\,, \qquad (88)$$

and then sum on p on both sides. ■

This theorem is adapted (by replacing $B(z,u,v)$ with $B_k(z,u,v)$ and $T(z)$ with $y_k(z)$) and gives the corresponding result for monotonic binary trees.

4.26 Theorem [25] *The generating function for the Steiner distance in monotonic binary trees satisfies the equation*

$$C_k(z,u,v) = \frac{zv(1+u)B_k^2(z,u,v) - 2zvy_k(z)B_k(z,u,v) + zy_k^2(z)(v-2) + y_k(z)}{1-2zy_k(z)} \,. \qquad (89)$$

First, we look at a particular case and compute the necessary statistics for the Steiner distance.

Case $k = 2$

Now, C_2 is differentiated with respect to v and evaluated $v = 1$. This derivative is denoted by $\Omega_2(z,u)$ and yields

$$\Omega_2(z,u) = \frac{z(1+u)\bar{y}_2^2(z) + \beta_2(z,u)[2z(1+u)\bar{y}_2(z) - 2zy_2(z)] - 2zy_2(z)\bar{y}_2(z) + zy_2^2(z)}{1-2zy_2(z)} \,, \qquad (90)$$

since $\bar{y}_2(z) = \bar{y}_1(z) + z(1+u)\bar{y}_2^2(z)$ (recall that $\frac{\partial B_2(z,u,v)}{\partial v}\Big|_{v=1} = \beta_2(z,u)$, as well as $B_2(z,u,1) = y_2(z(1+u)) = \bar{y}_2(z)$). We make use of the known asymptotic expansions to obtain an asymptotic expression for Ω_2

$$\Omega_2(z,u) \sim \frac{16\sqrt{3-16z}}{3\sqrt{3-16z(1+u)}} - \frac{128zu}{3\sqrt{3-16z}\sqrt{3-16z(1+u)}} \,, \qquad (91)$$

which is needed to analyse the behaviour of the coefficients.

4.27 Lemma *The coefficients of u^p in Ω_2 are*

$$[u^p]\Omega_2(z,u) \sim \frac{(p-1)}{3\,(2p-1)\,2^{2p-4}}\binom{2p}{p}\left(1-\frac{16}{3}z\right)^{-p}, \; z \to q_2 \,. \qquad (92)$$

Proof: The lemma will be proved by computing the coefficients of u^p from (91). The first term gives

$$[u^p]\frac{16\sqrt{3-16z}}{3\sqrt{3-16z(1+u)}} = 16\Big(1-\frac{16z}{3}\Big)^{1/2}[u^p]\frac{1}{\Big(1-\frac{16z}{3}\Big)^{1/2}\Big(1-\frac{\frac{16zu}{3}}{1-\frac{16z}{3}}\Big)^{1/2}}$$
$$= \frac{16}{3}\Big(-\frac{16}{3}\Big)^p\binom{-\frac{1}{2}}{p}\Big(1-\frac{16z}{3}\Big)^{-p} \sim \frac{1}{3\cdot 2^{2p-4}}\binom{2p}{p}\Big(1-\frac{16z}{3}\Big)^{-p},\ z\to\frac{3}{16}, \qquad (93)$$

and the second term yields

$$-[u^p]\frac{128zu}{3\sqrt{3-16z}\sqrt{3-16z(1+u)}} = -\frac{128z}{9\Big(1-\frac{16z}{3}\Big)^{1/2}}[u^{p-1}]\frac{1}{\Big(1-\frac{16z}{3}\Big)^{1/2}\Big(1-\frac{\frac{16zu}{3}}{1-\frac{16z}{3}}\Big)^{1/2}}$$
$$= -\frac{128z}{9}\Big(-\frac{16z}{3}\Big)^{p-1}\binom{-\frac{1}{2}}{p-1}\Big(1-\frac{16z}{3}\Big)^{-p}$$
$$\sim -\frac{p}{3\cdot 2^{2p-4}(2p-1)}\binom{2p}{p}\Big(1-\frac{16z}{3}\Big)^{-p},\ z\to\frac{3}{16}. \qquad (94)$$

The result in the lemma is obtained by adding these two coefficients. ■

Now Proposition 1.12 will be used in the computation of the coefficients of z^n from (91) as follows

$$[z^nu^p]\Omega_2(z,u) \sim \frac{(p-1)}{3\,(2p-1)\,2^{2p-4}}\binom{2p}{p}\Big(\frac{16}{3}\Big)^n\frac{n^{p-1}}{\Gamma(p)}, \qquad (95)$$

and we obtain the expectation of the Steiner distance for C_2 by dividing the above with the normalisation factor $\frac{4}{\sqrt{6\pi}}(\frac{16}{3})^n n^{-3/2}\frac{n^p}{\Gamma(p+1)}$

$$E^{(2)}_{n,p} \sim \frac{\sqrt{6\pi}\,(p-1)\,p}{3\,(2p-1)\,2^{2p-2}}\binom{2p}{p}\sqrt{n}. \qquad (96)$$

We move on to finding the second moment for the Steiner distance. The second derivative of C_2 with respect to v (evaluated at $v=1$) is denoted by $\Lambda_2(z,u)$ and gives

$$\Lambda_2(z,u) = \frac{1}{1-2z\,y_2(z)}\Big(4z(1+u)\,\bar{y}_2(z)\,\beta_2(z,u) + 2z(1+u)\,\beta_2^2(z,u)$$
$$+ 2z(1+u)\,\bar{y}_2(z)\,\Theta_2(z,u) - 4z\,y_2(z)\,\beta_2(z,u) - 2z\,y_2(z)\,\Theta_2(z,u)\Big)$$
$$\sim \frac{128\sqrt{2}zu}{(3-16z(1+u))^{3/2}} - \frac{16\sqrt{2}(3-16z)}{3(3-16z(1+u))^{3/2}}. \qquad (97)$$

This local expansion of Λ_2 gives us the possibility to look at the behaviour of its coefficients.

4.28 Lemma *For $z \to q_2$, the coefficients of u^p in Λ_2 are*

$$[u^p]\Lambda_2(z,u) \sim \frac{\sqrt{2}\,(p-1)}{3\sqrt{3}\cdot 2^{2p-4}}\binom{2p}{p}\left(1-\frac{16}{3}z\right)^{-p-1/2}. \tag{98}$$

Proof: As usual in this type of proof, the coefficients of u^p are computed from (97)

$$[u^p]\frac{128\sqrt{2}zu}{(3-16z(1+u))^{3/2}} = \frac{128\sqrt{2}z}{3\sqrt{3}}[u^{p-1}]\frac{1}{\left(1-\frac{16z}{3}\right)^{3/2}\left(1-\frac{\frac{16zu}{3}}{1-\frac{16z}{3}}\right)^{3/2}}$$

$$= \frac{128\sqrt{2}z}{3\sqrt{3}}\left(-\frac{16z}{3}\right)^{p-1}\binom{-\frac{3}{2}}{p-1}\left(1-\frac{16}{3}z\right)^{-p-1/2}$$

$$\sim \frac{\sqrt{2}p}{\sqrt{3}\cdot 2^{2p-4}}\binom{2p}{p}\left(1-\frac{16}{3}z\right)^{-p-1/2}, \quad z \to \frac{3}{16}, \tag{99}$$

as well as

$$-[u^p]\frac{16\sqrt{2}(3-16z)}{3(3-16z(1+u))^{3/2}} = -\frac{16\sqrt{2}\left(1-\frac{16z}{3}\right)}{3\sqrt{3}}[u^p]\frac{1}{\left(1-\frac{16z}{3}\right)^{3/2}\left(1-\frac{\frac{16zu}{3}}{1-\frac{16z}{3}}\right)^{3/2}}$$

$$= -\frac{16\sqrt{2}}{3\sqrt{3}}\left(-\frac{16z}{3}\right)^{p}\binom{-\frac{3}{2}}{p}\left(1-\frac{16}{3}z\right)^{-p-1/2}$$

$$\sim -\frac{\sqrt{2}(2p+1)}{3\sqrt{3}\cdot 2^{2p-4}}\binom{2p}{p}\left(1-\frac{16}{3}z\right)^{-p-1/2}, \quad z \to \frac{3}{16}. \tag{100}$$

Thus combining these coefficients gives the result in the lemma. ■

The lemma above enables us to find the coefficients of z^n easily

$$[z^n u^p]\Lambda_2(z,u) \sim \frac{\sqrt{2}\,(p-1)}{3\sqrt{3}\cdot 2^{2p-4}}\binom{2p}{p}\left(\frac{16}{3}\right)^n\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{101}$$

which (after normalisation) will give the second moment arising from C_2

$$\frac{[z^n u^p]\Lambda_2(z,u)}{\frac{4}{\sqrt{6\pi}}(\frac{16}{3})^n n^{-3/2}\frac{n^p}{\Gamma(p+1)}} \sim \frac{\sqrt{\pi}(p-1)}{3\cdot 2^{2p-3}}\binom{2p}{p}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})}n = \frac{8(p-1)}{3}n, \tag{102}$$

and now we are in a position to compute the variance of the Steiner distance for the case $k=2$

$$V_{n,p}^{(2)} \sim \left(\frac{8(p-1)}{3} - \frac{\pi(p-1)^2p^2}{3(2p-1)^2 2^{4p-5}}\binom{2p}{p}^2\right)n + \mathcal{O}(\sqrt{n}), \tag{103}$$

where $\frac{8(p-1)}{3} - \frac{\pi(p-1)^2p^2}{3(2p-1)^2 2^{4p-5}}\binom{2p}{p}^2 \to 0$ as $p \to \infty$.

The focus of the analysis will now be shifted to the general case k. First, the expectation of the Steiner distance will be computed. After denoting the derivative of C_k with respect to v (and evaluated at 1) by $\Omega_k(z,u)$ and making the usual substitutions we obtain

$$\begin{aligned}\Omega_k(z,u) = \frac{1}{1-2zy_k(z)}\Big(&z(1+u)\bar{y}_k^2(z) + 2z(1+u)\bar{y}_k(z)\beta_k(z,u)\\ &- 2zy_k(z)\bar{y}_k(z) - 2zy_k(z)\beta_k(z,u) + zy_k^2(z)\Big)\,. \end{aligned} \tag{104}$$

One can use the known asymptotic expansions and rewrite Ω_k as

$$\begin{aligned}\Omega_k(z,u) \sim \frac{1}{1-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z}}{2q_k}\right)}&\Bigg\{q_k\left(\frac{1}{1-2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)^2\\ &+2q_k\left(\frac{1}{1-2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)\\ &-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z}}{2q_k}\right)\left(\frac{1}{1-2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)\\ &-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z}}{2q_k}\right)\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)\\ &+q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z}}{2q_k}\right)^2\Bigg\}\\ &\sim \frac{\sqrt{q_k-z}}{q_k\sqrt{q_k-z(1+u)}}-\frac{zu}{2q_k\sqrt{q_k-z}\sqrt{q_k-z(1+u)}}\,. \end{aligned} \tag{105}$$

In order to proceed with our computations, we require the asymptotic behaviour of the coefficients in Ω_k.

4.29 Lemma *As $z \to q_k$, the coefficients of u^p in $\Omega_k(z,u)$ have the form*

$$[u^p]\Omega_k(z,u) \sim \frac{p-1}{2^{2p}\,(2p-1)\,q_k}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}. \tag{106}$$

Proof: The coefficient in the first term of (105) has been computed already

$$[u^p]\frac{\sqrt{q_k-z}}{q_k\sqrt{q_k-z(1+u)}} \sim \frac{1}{2^{2p}q_k}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}. \tag{107}$$

Then the coefficient in the second term is as follows

$$-[u^p]\frac{zu}{2q_k\sqrt{q_k-z}\sqrt{q_k-z(1+u)}} = -\frac{z}{2q_k^2}[u^{p-1}]\frac{1}{\left(1-\frac{z}{q_k}\right)^{1/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{1/2}}$$
$$= -\frac{z^p p}{2^{2p}q_k^{p+1}(2p-1)}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p} \sim -\frac{p}{2^{2p}q_k(2p-1)}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}, \quad z \to q_k\,. \tag{108}$$

Finally, adding the above coefficients gives the required result. ■

Next, the coefficients of z^n in $\Omega_k(z,u)$ turn out to be

$$[z^p u^p]\Omega_k(z,u) \sim \frac{p-1}{(2p-1)\,2^{2p}\,q_k^{n+1}}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}\,, \tag{109}$$

and their normalisation leads to the first moment.

4.30 Theorem [25] *The expectation for the Steiner distance in monotonic binary trees is*

$$E_{n,p}^{(k)} \sim \frac{(p-1)\,p}{(2p-1)\,2^{2p}\,p_k\,q_k}\binom{2p}{p}\sqrt{n}\,, \quad n\to\infty\,, \quad \textit{fixed } k. \tag{110}$$

We move on to finding the next moment for the Steiner distance. The second derivative of C_k with respect to v (evaluated at $v=1$), denoted by $\Lambda_k(z,u)$, is

$$\Lambda_k(z,u) = \frac{1}{1-2zy_k(z)}\Big(4z(1+u)\bar{y}_k(z)\beta_k(z,u) + 2z(1+u)\beta_k^2(z,u)$$
$$+ 2z(1+u)\bar{y}_k(z)\Theta_k(z,u) - 4zy_k(z)\beta_k(z,u) - 2zy_k(z)\Theta_k(z,u)\Big)\,. \tag{111}$$

Once again we make use of the known asymptotic expansions to rewrite Λ_k as follows

$$\Lambda_k(z,u) \sim \frac{1}{1-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z}}{2q_k}\right)}\Bigg\{2q_k\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)^2$$
$$+4q_k\left(\frac{1}{1-2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)$$
$$-4q_k\left(\frac{1}{1-2q_k}-\frac{c_k\sqrt{q_k-z}}{2q_k}\right)\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)$$
$$+\left[\frac{4q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)}{1-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)}\right.$$

$$+ \frac{2q_k\left(\frac{\sqrt{q_k-z}-\sqrt{q_k-z(1+u)}}{2q_k\sqrt{q_k-z(1+u)}}\right)^2}{1-2q_k\left(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\right)}\Bigg]\Bigg[2q_k\Big(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z(1+u)}}{2q_k}\Big)$$

$$- 2q_k\Big(\frac{1}{2q_k}-\frac{c_k\sqrt{q_k-z}}{2q_k}\Big)\Bigg]\Bigg\}$$

$$\sim \frac{3zu}{2c_kq_k(q_k-z(1+u))^{3/2}} - \frac{q_k-z}{c_kq_k(q_k-z(1+u))^{3/2}}\,. \tag{112}$$

This simplified form of the derivative is used for finding the behaviour of the coefficients of u^p and z^n.

4.31 Lemma *The coefficients of u^p in $\Lambda_k(z,u)$ are*

$$[u^p]\Lambda_k(z,u) \sim \frac{(p-1)}{2^{2p}\,c_k\,q_k^{3/2}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p-1/2}\,,\; z\to q_k\,. \tag{113}$$

Proof: The coefficients of u^p in the first term of (112) are

$$[u^p]\frac{3zu}{2c_kq_k(q_k-z(1+u))^{3/2}} = \frac{3z}{2c_kq_k^{5/2}}[u^{p-1}]\frac{1}{\left(1-\frac{z}{q_k}\right)^{3/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{3/2}}$$

$$= \frac{3pz^p}{2^{2p}c_kq_k^{p+3/2}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p-1/2} \sim \frac{3p}{2^{2p}c_kq_k^{3/2}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p-1/2}\,,\; z\to q_k\,, \tag{114}$$

and similarly

$$-[u^p]\frac{q_k-z}{c_kq_k(q_k-z(1+u))^{3/2}} \sim -\frac{2p+1}{2^{2p}c_kq_k^{3/2}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p-1/2}\,,\; z\to q_k\,. \tag{115}$$

Thus combining these coefficients gives the result in the lemma. ■

Next, we find that the coefficients of z^n are:

$$[z^nu^p]\,\Lambda_k(z,u) \sim \frac{p-1}{2^{2p+2}\sqrt{\pi}\,p_k\,q_k^{n+2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{116}$$

which lead to the following result.

4.32 Theorem [25] *The variance for the Steiner distance in monotonic binary trees is*

$$V_{n,p}^{(k)} = \frac{(p-1)}{p_k^2q_k^2}\left(\frac{1}{4\pi}-\frac{(p-1)\,p^2}{(2p-1)^2\,2^{4p}}\binom{2p}{p}^2\right)n + \mathcal{O}(\sqrt{n})\,, \tag{117}$$

as $n \to \infty$ and fixed k.

It is interesting to note that, by Stirling's formula, $\frac{1}{4\pi} - \frac{(p-1)\,p^2}{(2p-1)^2\,2^{4p}}\binom{2p}{p}^2$ goes to zero as $p \to \infty$.

4.2.4 The limiting distribution for the Steiner distance

Let $Y_{n,p}$ be the random variable which counts the size of the Steiner distance for p randomly chosen nodes in a monotonic binary tree of size n. So we are interested in the probabilities $\mathbb{P}\{Y_{n,p} = m\} = \frac{[z^n u^p v^m]C_k(z,u,v)}{b_n^{(k)}\binom{n}{p}}$.

4.33 Theorem [27] *The M-th moments, $\mathbb{E}(Y_{n,p}^M)$, of the Steiner distance for p randomly chosen nodes in a monotonically labelled binary tree of size n are asymptotically given by*

$$\mathbb{E}(Y_{n,p}^M) = \frac{[z^n u^p]\frac{\partial^M C_k(z,u,v)}{\partial v^M}\Big|_{v=1}}{p_k q_k^{-n}\frac{n^p}{\Gamma(p+1)}n^{-3/2}} = \frac{2^{M-2}}{\sqrt{\pi}c_k^{M-1}q_k^{(M+1)/2}p_k}\frac{\Gamma\left(p+\frac{M}{2}-1\right)}{\Gamma(p-1)}n^{M/2}\left(1+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right), \tag{118}$$

for fixed $p \geq 2$ and $n \to \infty$, where $c_k = 2\left(\prod_{i=2}^{k}(1-4q_i)\right)^{-1/4}$ and $p_k = \frac{c_k}{4\sqrt{\pi q_k}}$.

Once again the density function of the generalised Gamma distribution is used to derive the limiting distribution.

4.34 Theorem [27] *The probability, $\mathbb{P}\{Y_{n,p} = m\}$, that the Steiner distance for p randomly selected nodes in a monotonic binary tree of size n is equal to m is asymptotically*

$$\mathbb{P}\{Y_{n,p} = m\} \sim m^{p-2}\frac{q_k^{(p-1)/2}}{n^{p-1}(p-2)!}e^{-q_k^{1/2}\frac{m}{n}}\left(1+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)+\mathcal{O}\left(\frac{n}{m}\right)+\mathcal{O}\left(\frac{1}{m}\right)\right), \tag{119}$$

for $p \geq 2$, $n \to \infty$ and $m \leq K\sqrt{n}$ ($K > 0$ arbitrary but fixed).

Now we are in a position to find the limiting behaviour of the Steiner distance for our monotonic binary trees.

4.35 Theorem [27] *For fixed $p \geq 2$, $x > 0$, $m = x\sqrt{n}$ we have*

$$\sqrt{n}\,\mathbb{P}\{Y_{n,p} = m\} \sim x^{p-2}\frac{q_k^{(p-1)/2}}{(p-2)!}e^{-q_k^{1/2}x} = g\left(p-1, 1, \frac{1}{\sqrt{q_k}}; x\right). \tag{120}$$

Thus the limiting distribution of the normalised random variable $\frac{Y_{n,p}}{\sqrt{n}}$ is asymptotically, for fixed $p \geq 2$ and $n \to \infty$, a generalised Gamma distribution with parameters $\left(p-1, 1, \frac{1}{\sqrt{q_k}}\right)$.

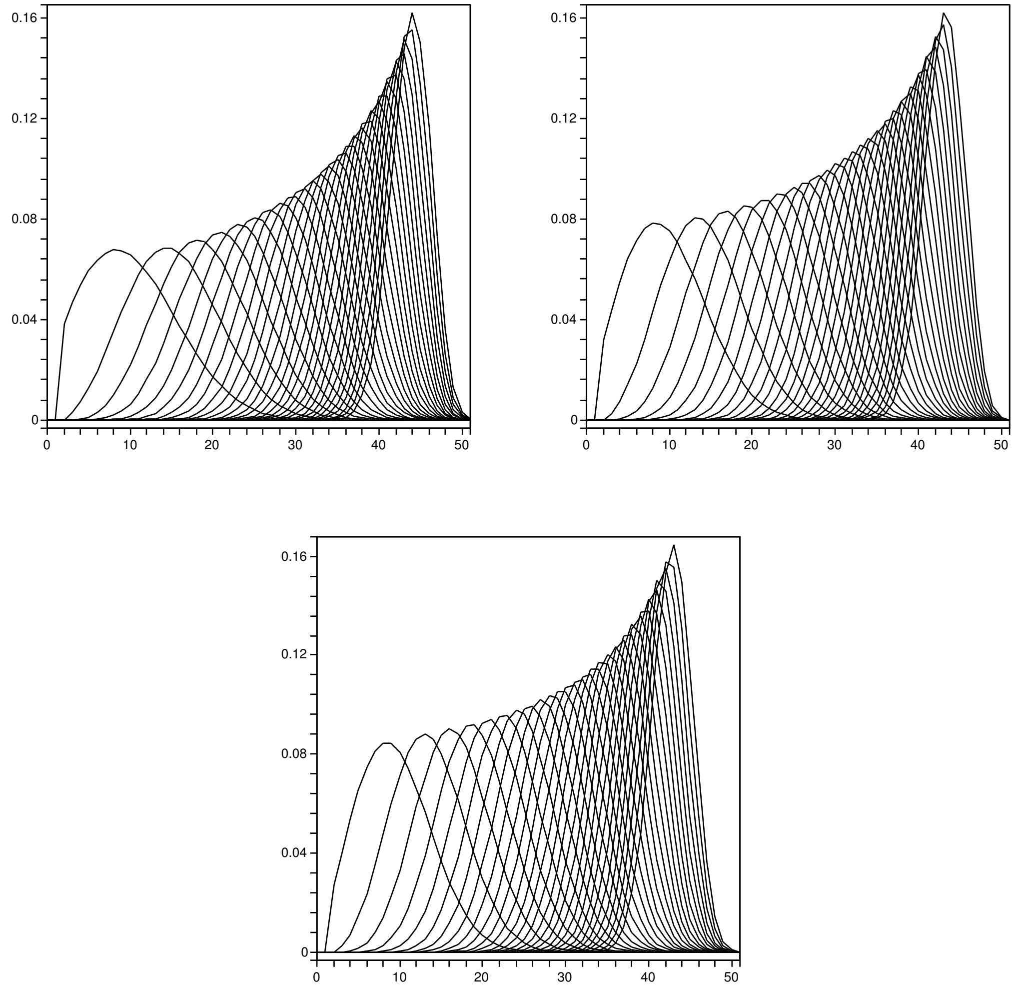


Figure 4.2: The probability distributions for the size of the Steiner distance in monotonic binary trees for $n = 51$, $p = 2 \ldots 30$ and $k = 1, 2, 3$.

4.3 Monotonic binary trees with two labels

Another interesting problem involving monotonic binary trees is their analysis when two labels are in place. Once again, the starting point is the binary model of Prodinger and Urbanek from [38] which can be extended to include more than one set of labels. We consider the class, $\mathfrak{B}_{k,l}$, of binary trees whose nodes are monotonically labelled with $\{1, 2, \ldots, k\}$ and $\{1, 2, \ldots, l\}$. Their generating function is

$$y_{k,l}(z) = \sum_{n \geq 0} b_n^{(k,l)} \, z^n \, ,$$

where $b_n^{(k,l)}$ represents the number of trees in $\mathfrak{B}_{k,l}$ with n nodes. We illustrate this with an example: consider monotonic binary trees with three nodes. Then there are five types of trees that arise, as shown in the figure below.

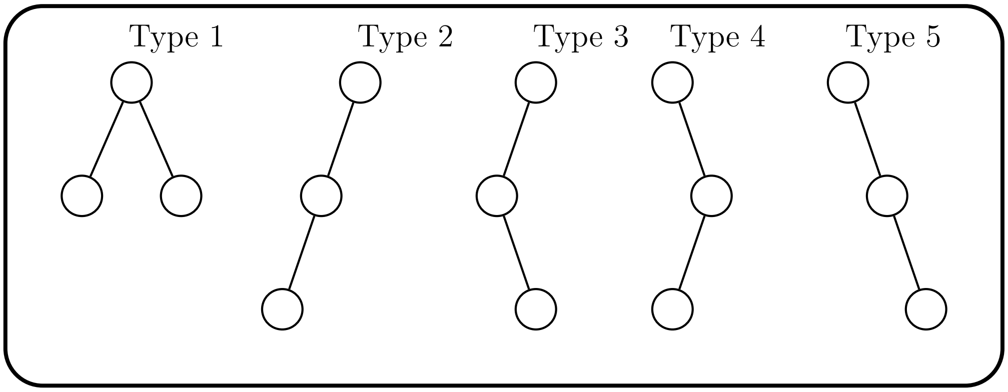


Figure 4.3: All 5 types of monotonic binary trees of size $n = 3$.

The next step is to attach a specific number of labels, p, to each node. For our example two labels from the set $\{1, 2\}$ will be considered.

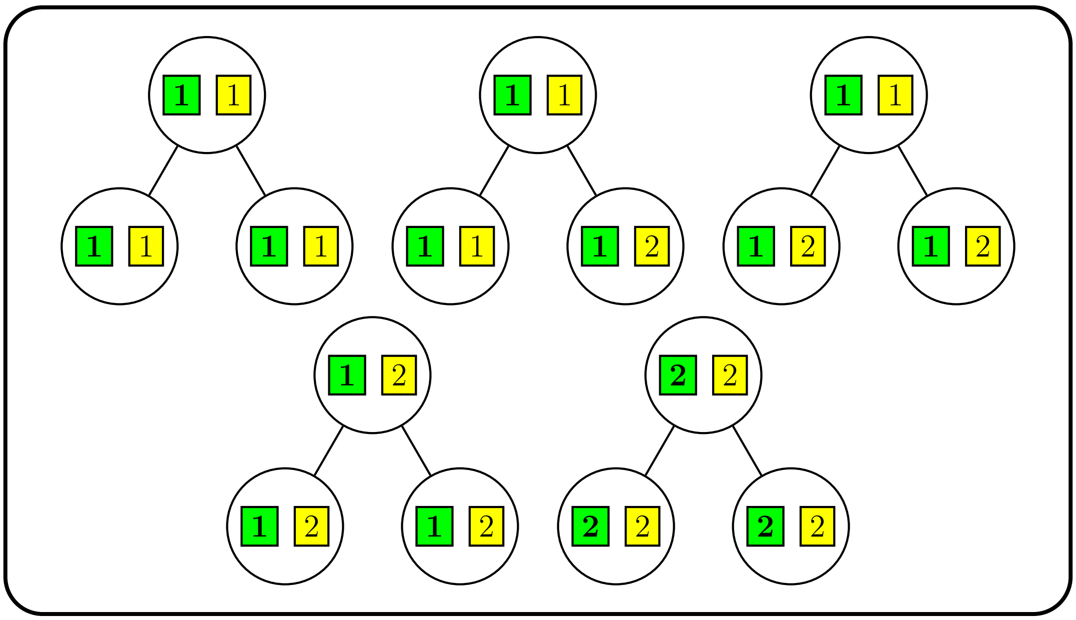


Figure 4.4: Type 1 of monotonic binary tree gives 5 possibilities for one label consisting of $\{1, 2\}$, thus giving 25 trees in total.

Using the generating function for the case of one label, which was done in the previous section on monotonic binary trees, one obtains the following result.

4.36 Theorem *The generating function for monotonic binary trees with two labels is*

$$y_{k,l}(z) = \sum_{i=1}^{k} \sum_{j=1}^{l} z y_{i,j}^2 + 1 \,, \tag{121}$$

where $y_{i,j} = y_{j,i}$.

Naturally, $y_{k,1}$ and $y_{1,l}$ constitute the old binary model. The singularities of the generating function need to be computed first. We begin by considering a few particular cases. First, we fix l say. If one lets $y_{k,1} = f_k$, then the following holds

$$f_k(z) = z\sum_{i=1}^{k} f_i^2(z) + 1 = z\sum_{i=1}^{k-1} f_i^2(z) + zf_k^2(z) + 1\,, \tag{122}$$

which has solution

$$f_k(z) = \frac{1-\sqrt{1-4zf_{k-1}(z)}}{2z}\,. \tag{123}$$

In order to find the singularities the equation $1-4zf_{k-1}(z) = 0$ needs to be solved. We are interested in the dominant one, $q_{k,1}$, which is the solution of $f_{k-1}(q_{k,1}) = \frac{1}{4q_{k,1}}$ as $z \longrightarrow q_{k,1}$. Next, if $y_{k,2} = g_k$ then

$$g_k(z) = z\sum_{i=1}^{k} f_i^2(z) + z\sum_{i=1}^{k} g_i^2(z) + 1 = f_k(z) + z\sum_{i=1}^{k-1} g_i^2(z) + zg_k^2(z)\,, \tag{124}$$

but $z\sum_{i=1}^{k-1} g_i^2(z) = g_{k-1}(z) - f_{k-1}(z)$ so the above equation can be written more conveniently as

$$zg_k^2(z) - g_k(z) + f_k(z) - f_{k-1}(z) = 0\,. \tag{125}$$

Its solution is

$$g_k(z) = \frac{1-\sqrt{1-4z(g_{k-1}(z)+f_k(z)-f_{k-1}(z))}}{2z}\,, \tag{126}$$

so it follows that the relevant singularity $q_{k,2}$ satisfies the equation

$$g_{k-1}(q_{k,2}) + f_k(q_{k,2}) - f_{k-1}(q_{k,2}) = \frac{1}{4q_{k,2}}\,,\ \ z \longrightarrow q_{k,2}\,. \tag{127}$$

Now, for $y_{k,3} = h_k$ we have the following

$$\begin{aligned} h_k(z) &= z\sum_{i=1}^{k} f_i^2(z) + z\sum_{i=1}^{k} g_i^2(z) + z\sum_{i=1}^{k} h_i^2(z) + 1 \\ &= f_k(z) - f_{k-1}(z) + g_k(z) - g_{k-1}(z) + h_{k-1}(z) + zh_k^2(z)\,, \end{aligned} \tag{128}$$

with solution

$$h_k(z) = \frac{1-\sqrt{1-4z(h_{k-1}(z)+g_k(z)-g_{k-1}(z)+f_k(z)-f_{k-1}(z))}}{2z}\,, \tag{129}$$

and its dominant singularity $q_{k,3}$ is the solution to the equation

$$h_{k-1}(q_{k,3}) + g_k(q_{k,3}) - g_{k-1}(q_{k,3}) + f_k(q_{k,3}) - f_{k-1}(q_{k,3}) = \frac{1}{4q_{k,3}}\,, \quad z \to q_{k,3}\,. \tag{130}$$

For the general setting, one can write (121) as

$$y_{k,l}(z) = z\sum_{i=1}^{k-1}\sum_{j=1}^{l-1} y_{i,j}^2(z) + zy_{k,l}^2(z) + 1\,, \tag{131}$$

which has solution

$$y_{k,l}(z) = \frac{1-\sqrt{1-4zy_{k-1,l-1}(z)}}{2z}\,, \tag{132}$$

and its singularity satisfies

$$y_{k-1,l-1}(q_{k,l}) = \frac{1}{4q_{k,l}}\,, \quad z \to q_{k,l}\,. \tag{133}$$

By employing similar methods to those used in the case of monotonic binary trees with one label, one can analyse these singularities in great detail. Various parameters such as the size of the ancestor tree and the Steiner distance could also be considered. However, this is beyond the scope of the thesis and we do not aim to present those results here.

4.4 Monotonic t-ary trees

We now look at t-ary trees which constitute a natural generalisation of binary trees.

4.37 Definition [11] A t-ary tree is either an external node or an internal node attached to an ordered sequence of t subtrees, all of which are t-ary trees.

4.38 Theorem [11] *The ordinary generating function that enumerates t-ary trees satisfies the functional equation*

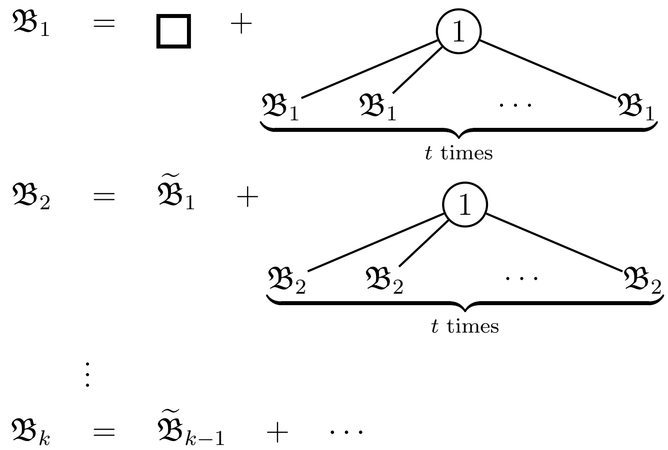
$$T(z) = z + T^t(z)\,. \tag{134}$$

The number of t-ary trees with n nodes is

$$\frac{1}{(t-1)(n+1)}\binom{tn}{n} \sim \frac{d_t(e_t)^n}{n^{3/2}}\,, \tag{135}$$

where $d_t = \frac{1}{\sqrt{\frac{2\pi(t-1)^3}{t}}}$ and $e_t = \frac{t^t}{(t-1)^{t-1}}$.

In this section we generalise the monotonic binary tree model and consider the class of monotonic t-ary trees, whose defining equations (as introduced in [38]) are given by



4.39 Theorem [38] *The generating functions for monotonic t-ary trees satisfy*

$$y_k(z) = y_{k-1}(z) + z y_k^t(z)\,, \quad \textit{with} \quad y_0(z) = 1\,. \tag{136}$$

From the paper of Prodinger and Urbanek, [38], we know that as z tends to the singularity q_k, the generating function has the following local expansion

$$\begin{aligned} y_k(z) &= y_k(q_k) - C_k(t)\sqrt{q_k - z} + \mathcal{O}(q_k - z) \\ &= \frac{1}{r_k} - C_k(t)\sqrt{q_k - z} + \mathcal{O}(q_k - z)\,, \end{aligned} \tag{137}$$

for $C_k(t)$ some constants which are defined below.

4.40 Lemma [27] *The constants $C_k(t)$ are given by*

$$\Big[2\Big(\sum_{i=1}^{k-1}\Big(\frac{(y_k(q_k) - y_{k-1}(q_k))}{z}\prod_{j=i}^{k-1}\frac{1}{1 - t\frac{(y_j(q_k) - y_{j-1}(q_k))}{y_j(q_k)}}\Big) + \frac{(y_k(q_k) - y_{k-1}(q_k))}{z}\Big)\frac{y_k(q_k)}{t-1}\Big]^{1/2}. \tag{138}$$

<u>Proof</u>: It was shown in [38] that

$$\frac{1}{2}C_k^2(t) = \lim_{z \to q_k} y_k'(z)(y_k(q_k) - y_k(z))\,, \tag{139}$$

as well as

$$y_k'(z) = \frac{y'_{k-1(z)} + y_k^t(z)}{1 - z t y_k^{t-1}(z)}\,. \tag{140}$$

Furthermore, we can expand the denominator around the singularity q_k as follows

$$\begin{aligned} 1 - zty_k^{t-1} &= 1 - \Big(\frac{y_k}{y_k(q_k)}\Big)^{t-1} \\ &= \Big(1 - \frac{y_k}{y_k(q_k)}\Big)\Big(1 + \frac{y_k}{y_k(q_k)} + \cdots + \Big(\frac{y_k}{y_k(q_k)}\Big)^{t-2}\Big). \end{aligned} \tag{141}$$

Now we need to determine y'_{k-1}. We use our recursive idea

$$\begin{aligned} y'_{k-1} &= \frac{y'_{k-2} + y^t_{k-1}}{1 - zty^{t-1}_{k-1}} \\ &= \Big((y'_{k-3} + y^t_{k-2})\Big(\frac{1}{1 - zty^{t-1}_{k-2}}\Big) + y^t_{k-1}\Big)\frac{1}{1 - zty^{t-1}_{k-1}} \\ &= y'_{k-3}\Big(\frac{1}{1 - zty^{t-1}_{k-2}}\Big)\Big(\frac{1}{1 - zty^{t-1}_{k-1}}\Big) + y^t_{k-2}\Big(\frac{1}{1 - zty^{t-1}_{k-2}}\Big)\Big(\frac{1}{1 - zty^{t-1}_{k-1}}\Big) \\ &+ y^t_{k-1}\Big(\frac{1}{1 - zty^{t-1}_{k-1}}\Big) \\ &= y_1\Big(\frac{1}{1 - zty^{t-1}_{1}}\Big)\Big(\frac{1}{1 - zty^{t-1}_{2}}\Big)\cdots\Big(\frac{1}{1 - zty^{t-1}_{k-1}}\Big) \\ &+ y_2\Big(\frac{1}{1 - zty^{t-1}_{2}}\Big)\Big(\frac{1}{1 - zty^{t-1}_{3}}\Big)\cdots\Big(\frac{1}{1 - zty^{t-1}_{k-1}}\Big) \\ &+ \cdots + y^t_{k-1}\Big(\frac{1}{1 - zty^{t-1}_{k-1}}\Big) \\ &= \sum_{i=1}^{k-1}\Big(y^t_i \prod_{j=i}^{k-1}\frac{1}{1 - zty^{t-1}_j}\Big). \end{aligned} \tag{142}$$

So we can substitute the above in (139) to obtain

$$\frac{1}{2}\, C^2_k(t) = \Big(\sum_{i=1}^{k-1}\Big(y^t_i(q_k)\prod_{j=i}^{k-1}\frac{1}{1 - zty^{t-1}_j(q_k)}\Big) + y^t_k(q_k)\Big)\frac{y_k(q_k)}{t-1}, \tag{143}$$

and moreover

$$\begin{aligned} C_k(t) &= \Big(2\Big(\sum_{i=1}^{k-1}\Big(y^t_i(q_k)\prod_{j=i}^{k-1}\frac{1}{1 - zty^{t-1}_j(q_k)}\Big) + y^t_k(q_k)\Big)\frac{y_k(q_k)}{t-1}\Big)^{1/2} \\ &= \Bigg[2\Big(\sum_{i=1}^{k-1}\Big(\frac{(y_k(q_k) - y_{k-1}(q_k))}{z}\prod_{j=i}^{k-1}\frac{1}{1 - t\frac{(y_j(q_k) - y_{j-1}(q_k))}{y_j(q_k)}}\Big) + \frac{(y_k(q_k) - y_{k-1}(q_k))}{z}\Big)\frac{y_k(q_k)}{t-1}\Bigg]^{1/2}, \end{aligned} \tag{144}$$

since $y^t_k(z) = \frac{y_k(q_k) - y_{k-1}(q_k)}{z}$ and $y^{t-1}_j(z) = \frac{y^t_j(z)}{y_j(z)}$. ■

To establish the validity of this formula we check it for $k = 1, 2$ against the binary case $t = 2$, for $z \longrightarrow q_k$,

$$C_1(2) = \Big(2\Big(\sum_{i=1}^{0}\Big(y_i^2(q_1)\prod_{j=i}^{0}\frac{1}{1-2zy_j(q_1)}\Big) + y_1^2(q_1)\Big)y_1(q_1)\Big)^{1/2} = \sqrt{2(0\cdot 1+4)2} = 4\,,$$

$$\begin{aligned} C_2(2) &= \Big(2\Big(\sum_{i=1}^{1}\Big(y_i^2(q_2)\prod_{j=i}^{1}\frac{1}{1-2zy_j(q_2)}\Big) + y_2^2(q_2)\Big)y_2(q_2)\Big)^{1/2} \\ &= \Big(2\Big(\frac{y_1^2(q_2)}{1-2q_2y_1(q_2)} + y_2^2(q_2)\Big)y_2(q_2)\Big)^{1/2} = \Big(2\Big(\frac{(\frac{4}{3})^2}{1-2\frac{3}{16}\frac{4}{3}} + \Big(\frac{8}{3}\Big)^2\Big)\frac{8}{3}\Big)^{1/2} = \frac{16\sqrt{2}}{3}\,. \end{aligned} \tag{145}$$

which are the constants we obtained in the binary tree section.

4.41 Note *We know that $y_i(q_k) = a_{i,k}(t) + \mathcal{O}(q_k - z)$ for $i < k$ and $a_{i,k}(t)$ some constant. The constant $a_{i,k}(t)$ has been computed in [20] and [38]*

$$y_i(q_k) = \Big(\frac{q_{k-i}}{q_k}\Big)^{1/(t-1)} = \frac{r_{k-i}}{r_k}\,. \tag{146}$$

For example consider $k = 1$ and $t = 2$. The solution to $y_1(z) = 1 + y_1^2(z)$ is $y_1(z) = \frac{1-\sqrt{1-4z}}{2z}$ which we can evaluate at $q_2 = \frac{3}{16}$, say. Thus we find that $y_1(q_2) = \frac{4}{3}$.

Other quantities needed for our results are presented in the table below.

constant	$k=1$	$k=2$	$k=3$
$d_k = \frac{1}{q_k}$	$\frac{t}{(\frac{t-1}{t})^{t-1}}$	$\frac{t}{(\frac{t-1-(\frac{t-1}{t})^t}{t})^{t-1}}$	$\frac{t}{(\frac{t-1-(\frac{t-1}{t})^t-(\frac{t-1-(\frac{t-1}{t})^t}{t})^t}{t})^{t-1}}$
$q_k = \frac{r_k^{t-1}}{t}$	$\frac{(\frac{t-1}{t})^{t-1}}{t}$	$\frac{(\frac{t-1-(\frac{t-1}{t})^t}{t})^{t-1}}{t}$	$\frac{(\frac{t-1-(\frac{t-1}{t})^t-(\frac{t-1-(\frac{t-1}{t})^t}{t})^t}{t})^{t-1}}{t}$
$r_{k+1} = r_k - \frac{r_k^t}{t}$; $r_0 = 1$	$\frac{t-1}{t}$	$\frac{t-1-(\frac{t-1}{t})^t}{t}$	$\frac{t-1-(\frac{t-1}{t})^t-(\frac{t-1-(\frac{t-1}{t})^t}{t})^t}{t}$
$y_k(q_k) = \frac{1}{r_k}$	$\frac{t}{t-1}$	$\frac{t}{t-1-(\frac{t-1}{t})^t}$	$\frac{t}{t-1-(\frac{t-1}{t})^t-(\frac{t-1-(\frac{t-1}{t})^t}{t})^t}$

4.4.1 Size of the ancestor tree

We begin this section by first presenting the generating function for t-ary trees (which is a generalisation of the result in the binary trees discussion).

4.42 Theorem [27] *The generating function for the size of the ancestor tree in t-ary trees is*

$$G(z,u,v) = zv(1+u)G^t(z,u,v) - zvT^t(z) + T(z)\,. \tag{147}$$

Proof: This can be done by generalising the proof from the monotonically labelled binary trees. ■

One can easily modify (147) to obtain the equations of the size of the ancestor tree for the recursive binary trees given in the Prodinger-Urbanek model. We replace T with y_k and G with G_k.

4.43 Theorem [25] *The generating functions definining the size of the ancestor tree in the monotonic t-ary trees are*

$$G_k(z,u,v) = G_{k-1}(z,u,v) + zv(1+u)G_k^t(z,u,v) + (1-v)zy_k^t(z)\,,\ G_0(z,u,v) = 1\,. \tag{148}$$

The aim is to produce the expectation and variance for our parameter using the generating function in the theorem above. It is convenient to start by considering the first two cases of k.

Case $k = 1$

As usual, G_1 is differentiated with respect to v and then evaluated at $v = 1$. If this derivative is denoted by $g_1(z,u)$ then we obtain

$$g_1(z,u) = \frac{z(1+u)G_1^t(z,u,1) - zy_1(z)}{1 - z(1+u)tG_1^{t-1}(z,u,1)} = \frac{\bar{y}_1(z) - y_1(z)}{1 - t\frac{(\bar{y}_1(z)-1)}{\bar{y}_1(z)}}\,, \tag{149}$$

where $G_1(z,u,1) = y_1(z(1+u)) = \bar{y}_1(z)$ and $\bar{y}_1^{t-1}(z) = \frac{\bar{y}_1^t(z)}{\bar{y}_1(z)}$. The asymptotic expansion

$$y_1(z) \sim \frac{1}{r_1} - C_1(t)\sqrt{q_1 - z}\,, \tag{150}$$

and the equivalent one for $\bar{y}_1(z)$ are now substituted in (149)

$$g_1(z,u) \sim \frac{t\sqrt{q_1 - z}}{(t-1)^2\sqrt{q_1 - z(1+u)}}\,, \tag{151}$$

which helps in determining the behaviour of the coefficients in g_1.

4.44 Lemma *The coefficients of u^p in $g_1(z,u)$ satisfy*

$$[u^p]g_1(z,u) \sim \frac{t}{2^{2p}(t-1)^2}\binom{2p}{p}\left(1-\frac{z}{q_1}\right)^{-p}, \quad z \to q_1 \,, \tag{152}$$

where $q_1 = \frac{(t-1)^{t-1}}{t^t}$.

Proof: In order to prove this lemma, we extract the coefficients of u^p from (149) as follows

$$\begin{aligned}[u^p]\frac{t\sqrt{q_1-z}}{(t-1)^2\sqrt{q_1-z(1+u)}} &= \frac{t(1-\frac{z}{q_1})^{1/2}}{(t-1)^2}[u^p]\frac{1}{(1-\frac{z}{q_k})^{1/2}\left(1-\frac{\frac{zu}{q_1}}{1-\frac{z}{q_1}}\right)^{1/2}} \\ = \frac{t}{(t-1)^2}\left(-\frac{z}{q_1}\right)^p\binom{-\frac{1}{2}}{p}\left(1-\frac{z}{q_1}\right)^{-p} &\sim \frac{t}{2^{2p}(t-1)^2}\binom{2p}{p}\left(1-\frac{z}{q_1}\right)^{-p}, \quad z \to q_1 \,.\end{aligned} \tag{153}$$

■

While the above lemma gives the behaviour of the coefficients of u^p for any natural number t, it is interesting to investigate them for particular values. In order to do this, MAPLE was used to obtain the coefficients for $t = 2, \ldots, 11$ and the results are presented in the table below.

t	singularity	$[u^p]g_1(z,u)$
2	$\frac{1}{4}$	$\binom{2p}{p}\frac{1}{2^{2p-1}}(1-4z)^{-p}$
3	$\frac{2^2}{3^3}$	$\binom{2p}{p}\frac{3}{2^{2p+2}}(1-\frac{3^3}{2^2}z)^{-p}$
4	$\frac{3^3}{2^8}$	$\binom{2p}{p}\frac{1}{3^2 \cdot 2^{2p-2}}(1-\frac{2^8}{3^3}z)^{-p}$
5	$\frac{2^8}{5^5}$	$\binom{2p}{p}\frac{5}{2^{2p+4}}(1-\frac{5^5}{2^8}z)^{-p}$
6	$\frac{5^5}{2^6 \cdot 3^6}$	$\binom{2p}{p}\frac{3}{5^2 \cdot 2^{2p-1}}(1-\frac{2^6 \cdot 3^6}{5^5}z)^{-p}$
7	$\frac{2^6 \cdot 3^6}{7^7}$	$\binom{2p}{p}\frac{7}{3^2 \cdot 2^{2p+2}}(1-\frac{7^7}{2^6 \cdot 3^6}z)^{-p}$
8	$\frac{7^7}{2^{24}}$	$\binom{2p}{p}\frac{1}{7^2 \cdot 2^{2p-3}}(1-\frac{2^{24}}{7^7}z)^{-p}$
9	$\frac{2^{24}}{3^{18}}$	$\binom{2p}{p}\frac{3^2}{7^2 \cdot 2^{2p+6}}(1-\frac{3^{18}}{2^{24}}z)^{-p}$
10	$\frac{3^{18}}{2^{10} \cdot 5^{10}}$	$\binom{2p}{p}\frac{5}{3^4 \cdot 2^{2p-1}}(1-\frac{2^{10} \cdot 5^{10}}{3^{18}}z)^{-p}$
11	$\frac{2^{10} \cdot 5^{10}}{11^{11}}$	$\binom{2p}{p}\frac{11}{5^2 \cdot 2^{2p+2}}(1-\frac{11^{11}}{2^{10} \cdot 5^{10}}z)^{-p}$

4.45 Note *For $t = 2$ we have $[u^p]g_1(z,u) \sim \frac{1}{2^{2p-1}}\binom{2p}{p}(1-4z)^{-p}$. As expected, these are the* *coefficients of u^p we found in the monotonic binary trees case (15).*

We use the result of Proposition 1.12, with $\alpha = -p$, and find that

$$[z^n u^p] g_1(z,u) \sim \frac{q_1^{-n} t}{2^{2p}(t-1)^2} \binom{2p}{p} \frac{n^{p-1}}{\Gamma(p)} \,. \tag{154}$$

It follows that the expectation for the size of the ancestor tree for $k = 1$ is obtained by dividing the above with the normalising factor $p_1 q_1^{-n} n^{-3/2} \frac{n^p}{\Gamma(p+1)}$

$$E_{n,p}^{(1)} \sim \frac{pt}{p_1 \cdot 2^{2p}(t-1)^2} \binom{2p}{p} \sqrt{n} = \frac{pt\sqrt{\pi}}{2^{2p}(t-1)^2} \binom{2p}{p} \sqrt{n}\,, \; p_1 = \frac{1}{\sqrt{\pi}} \,. \tag{155}$$

4.46 Note *By putting $t = 2$ in the above expectation one obtains the same result (19) as in the monotonic binary tree case.*

Next, we proceed with computing the second moment for the size of the ancestor tree for the particular value $k = 1$. The second derivative of G_1 with respect to v (evaluated at 1) is denoted by $h_1(z,u)$. After making the appropriate substitutions the following is obtained

$$\begin{aligned} h_1(z,u) &= \frac{1}{1 - \frac{t(\bar{y}_1(z)-1)}{\bar{y}_1(z)}} \Big(\frac{2t(\bar{y}_1(z)-1) g_1(z,u)}{\bar{y}_1(z)} \\ &\quad + \frac{t^2(\bar{y}_1(z)-1) g_1^2(z,u)}{\bar{y}_1^2(z)} - \frac{t(\bar{y}_1(z)-1) g_1^2(z,u)}{\bar{y}_1^2(z)} \Big) \\ &\sim \frac{t^2 z u}{C_1(t)(t-1)^4 (q_1 - z(1+u))^{3/2}} \,. \end{aligned} \tag{156}$$

4.47 Lemma *The coefficients of u^p in $h_1(z,u)$ have the form*

$$[u^p] h_1(z,u) \sim \frac{pt^2}{C_1(t) 2^{2p-1} q_1^{1/2} (t-1)^4} \binom{2p}{p} \Big(1 - \frac{z}{q_1}\Big)^{-p-1/2} \,, \; z \to q_1 \,, \tag{157}$$

where $C_1(t) = \Big(\frac{2(1-r_1)}{r_1^2 q_1 (t-1)}\Big)^{1/2}$.

Proof: The coefficients of u^p in h_1 are computed from (156) as follows

$$[u^p] \frac{t^2 z u}{C_1(t)(t-1)^4 (q_1 - z(1+u))^{3/2}} = \frac{t^2 z}{C_1(t) q_1^{3/2} (t-1)^4} [u^{p-1}] \frac{1}{(1-\frac{z}{q_1})^{3/2} \Big(1 - \frac{\frac{zu}{q_1}}{1-\frac{z}{q_1}}\Big)^{3/2}}$$

$$= \frac{t^2 z}{C_1(t) q_1^{3/2} (t-1)^4} \Big(-\frac{z}{q_1}\Big)^{p-1} \binom{-\frac{3}{2}}{p-1} \Big(1 - \frac{z}{q_1}\Big)^{-p-1/2}$$

$$\sim \frac{pt^2}{C_1(t)2^{2p-1}q_1^{1/2}(t-1)^4}\binom{2p}{p}\left(1-\frac{z}{q_1}\right)^{-p-1/2}, \quad z \to q_1\,. \tag{158}$$

■

The lemma above will be applied in MAPLE to compute some particular coefficients of u^p for $t = 2, \ldots, 7$. The results are summarised below.

t	singularity	$[u^p]h_1(z,u)$
2	$\frac{1}{4}$	$\binom{2p}{p}\frac{p}{C_1(t)\cdot 2^{2p-4}}(1-4z)^{-p}$
3	$\frac{2^2}{3^3}$	$\binom{2p}{p}\frac{3^{7/2}\cdot p}{C_1(t)\cdot 2^{2p+4}}(1-\frac{3^3}{2^2}z)^{-p}$
4	$\frac{3^3}{2^8}$	$\binom{2p}{p}\frac{p}{C_1(t)\cdot 3^{11/2}\cdot 2^{2p-9}}(1-\frac{2^8}{3^3}z)^{-p}$
5	$\frac{2^8}{5^5}$	$\binom{2p}{p}\frac{5^{9/2}\cdot p}{C_1(t)\cdot 2^{2p+11}}(1-\frac{5^5}{2^8}z)^{-p}$
6	$\frac{5^5}{2^6\cdot 3^6}$	$\binom{2p}{p}\frac{3^5\cdot p}{C_1(t)\cdot 5^{13/2}\cdot 2^{2p-6}}(1-\frac{2^6\cdot 3^6}{5^5}z)^{-p}$
7	$\frac{2^6\cdot 3^6}{7^7}$	$\binom{2p}{p}\frac{7^{11/2}\cdot p}{C_1(t)\cdot 3^7\cdot 2^{2p+6}}(1-\frac{7^7}{2^6\cdot 3^6}z)^{-p}$

4.48 Note *Once again for $t = 2$ we obtain exactly the same coefficient of u^p as in (21), where $C_1(2) = 4$ was computed earlier in this section.*

Now, Proposition 1.12 is employed with $\alpha = -p - \frac{1}{2}$ to extract the coefficients of z^n in h_1

$$[z^n u^p]h_1(z,u) \sim \frac{pt^2}{C_1(t)2^{2p-1}q_1^{n+1/2}(t-1)^4}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{159}$$

which will be normalised and give the second moment. Thus, the variance for the size of the ancestor tree for the case $k = 1$ is

$$\begin{aligned} V_{n,p}^{(1)} &\sim \frac{pt^2\sqrt{\pi}}{C_1(t)2^{2p-1}q_1^{1/2}(t-1)^4}\binom{2p}{p}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})}n - \frac{\pi p^2 t^2}{2^{4p}(t-1)^4}\binom{2p}{p}n + \mathcal{O}(\sqrt{n}) \\ &= \frac{pt^2}{(t-1)^4}\left(\frac{2\sqrt{\pi}}{C_1(t)q_1^{1/2}} - \frac{\pi p}{2^{4p}}\binom{2p}{p}\right)n + \mathcal{O}(\sqrt{n}), \; q_1 = \frac{(t-1)^{t-1}}{t^t}. \end{aligned} \tag{160}$$

4.49 Note *If we let $t = 2$ and $C_1(2) = 4$ in the above variance then we get the result from the section on monotonic binary trees (26).*

Case $k = 2$

We conduct a similar analysis on the size of the ancestor tree as in the previous case. Since the methods are the same as in the case $k = 1$, the results are presented without much detail or proofs. First, the expectation will be computed. It turns out that the derivative of G_2 with respect to v (evaluated at $v = 1$) and denoted by $g_2(z, u)$ is

$$\begin{aligned} g_2(z,u) &= \frac{g_1(z,u) + z(1+u)G_2^t(z,u,1) - zy_2^t(z)}{1 - zt(1+u)G_2^{t-1}(z,u,1)} \\ &= \frac{g_1(z,u) + (\bar{y}_2(z) - \bar{y}_1(z)) - (y_2(z) - y_1(z))}{1 - t\frac{(\bar{y}_2(z)-\bar{y}_1(z))}{\bar{y}_2(z)}} \end{aligned} \tag{161}$$

where we have used the usual substitutions and the fact that

$$G_2^{t-1}(z,u,1) = \bar{y}_2^{t-1}(z) = \frac{\bar{y}_2^t(z)}{\bar{y}_2(z)}\,. \tag{162}$$

It is convenient to substitute the known asymptotic expansions in (161) and we obtain

$$g_2(z,u) \sim \frac{t\sqrt{q_2 - z}}{(t-1)\big(t - 1 - (\frac{t-1}{t})^t\big)\sqrt{q_2 - z(1+u)}}\,, \tag{163}$$

since $y_1(q_2) = \frac{r_1}{r_2}$ (and similarly for $\bar{y}_1(q_2)$). This expansion will enable us to look at the behaviour of the coefficients in g_2. We note that g_1, which appears in (161), gives no contribution to the main term shown in the asymptotic form of g_2 (this was easily checked with MAPLE).

4.50 Lemma *The coefficients of u^p in $g_2(z, u)$ are given by*

$$[u^p]g_2(z,u) \sim \frac{t}{2^{2p}(t-1)\big(t-1-(\frac{t-1}{t})^t\big)}\binom{2p}{p}\Big(1 - \frac{z}{q_2}\Big)^{-p}, \; z \to q_2\,. \tag{164}$$

Moreover, the coefficients of z^n in g_2 have the form

$$[z^n u^p]g_2(z,u) \sim \frac{t}{2^{2p}(t-1)\big(t-1-(\frac{t-1}{t})^t\big)}\Big(\frac{1}{q_2}\Big)^n\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}\,, \tag{165}$$

which yield the expectation for the size of the ancestor tree upon normalisation with $p_2 q_2^{-n} n^{-3/2} \frac{n^p}{\Gamma(p+1)}$

$$E_{n,p}^{(2)} \sim \frac{pt}{2^{2p}p_2(t-1)\big(t-1-(\frac{t-1}{t})^t\big)}\binom{2p}{p}\sqrt{n}\,. \tag{166}$$

One can now proceed with finding the second moment for G_2. The second derivative of G_2 with respect to v and evaluated at $v = 1$ yields

$$h_2(z,u) = \frac{1}{1 - \frac{t(\bar{y}_2(z)-\bar{y}_1(z))}{\bar{y}_2(z)}}\Big(h_1(z,u) + \frac{2t(\bar{y}_2(z)-\bar{y}_1(z))g_2(z,u)}{\bar{y}_2(z)} + \frac{t^2(\bar{y}_2(z)-\bar{y}_1(z))g_2^2(z,u)}{\bar{y}_2^2(z)} - \frac{t(\bar{y}_2(z)-\bar{y}_1(z))g_2^2(z,u)}{\bar{y}_2^2(z)}\Big). \tag{167}$$

An asymptotic form of g_2 will be derived by replacing the terms in the above with their local expansions around q_2

$$h_2(z,u) \sim \frac{t^2 z u}{C_2(t)(t-1)^2\big(t-1-(\frac{t-1}{t})^t\big)^2(q_2 - z(1+u))^{3/2}}, \tag{168}$$

and this will make it easier for us to analyse the behaviour of the coefficients in g_2.

4.51 Lemma *The coefficients of u^p in $h_2(z,u)$ have the form*

$$[u^p]h_2(z,u) \sim \frac{pt^2}{C_2(t)2^{2p-1}q_2^{1/2}(t-1)^2\big(t-1-(\frac{t-1}{t})^t\big)^2}\binom{2p}{p}\Big(1-\frac{z}{q_2}\Big)^{-p-1/2}, \quad z \to q_2\,. \tag{169}$$

Finally, the coefficients of z^n in (168) are

$$[z^n u^p]h_2(z,u) \sim \frac{pt^2}{C_2(t)2^{2p-1}q_2^{n+1/2}(t-1)^2\big(t-1-(\frac{t-1}{t})^t\big)^2}\binom{2p}{p}\frac{n^{-p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{170}$$

which, upon normalisation, result in the second moment:

$$\frac{2pt^2}{\sqrt{\pi}C_2(t)p_2 q_2^{1/2}(t-1)^2\big(t-1-(\frac{t-1}{t})^t\big)^2}\,n\,. \tag{171}$$

The latter will be used in deriving the variance for the size of the ancestor tree for the particular case G_2 as follows

$$V_{n,p}^{(2)} \sim \frac{pt^2}{p_2(t-1)^2\big(t-1-(\frac{t-1}{t})^t\big)^2}\Big(\frac{2}{\sqrt{\pi}C_2(t)q_2^{1/2}} - \frac{p}{2^{4p}p_2}\binom{2p}{p}^2\Big)n + \mathcal{O}(\sqrt{n})\,. \tag{172}$$

We are now in a position to produce the usual statistics for the general case. The first step in getting the expectation for the size of the ancestor tree is to differentiate G_k with respect to v and evaluate it at $v = 1$. Then we denote it by $g_k(z,u)$ and set $G_k(z,u,1) = \bar{y}_k(z)$

$$(1 - tz(1+u)\bar{y}_k^{t-1}(z))g_k(z,u) = g_{k-1}(z,u) + \bar{y}_k(z) - \bar{y}_{k-1}(z) - y_k(z) + y_{k-1}(z)\,, \tag{173}$$

since $zy_k^t(z) = y_k(z) - y_{k-1}(z)$ and $z(1+u)y_k^t(z(1+u)) = \bar{y}_k(z) - \bar{y}_{k-1}(z)$.

4.52 Lemma [27] *The solution to the recursion in (173) is*

$$g_k(z,u) = \sum_{j=1}^{k} \Big[\prod_{i=j}^{k}(1 - tz(1+u)\bar{y}_i^{t-1}(z))\Big]^{-1} \big(\bar{y}_j(z) - \bar{y}_{j-1}(z) - y_j(z) + y_{j-1}(z)\big)\,, \quad (174)$$

and its dominant term has the form

$$g_k(z,u) \sim \frac{\bar{y}_k(z) - y_k(z)}{1 - t\frac{(\bar{y}_k(z) - \bar{y}_{k-1}(z))}{\bar{y}_k(z)}}\,. \quad (175)$$

Proof: Both sides of (173) are multiplied with the product $\prod_{i=1}^{k-1}(1 - tz(1+u)\bar{y}_i^{t-1}(z))$. Then we sum over the k's and obtain

$$\begin{aligned}&\Big[\prod_{i=1}^{k-1}(1 - tz(1+u)\bar{y}_i^{t-1}(z))\Big](1 - tz(1+u)\bar{y}_k^{t-1}(z))g_k(z,u)\\ &= \sum_{j=0}^{k-1}\Big[\prod_{i=1}^{j}(1 - tz(1+u)\bar{y}_i^{t-1}(z))\Big]\big(\bar{y}_j(z) - \bar{y}_{j-1}(z) - y_j(z) + y_{j-1}(z)\big)\,,\end{aligned} \quad (176)$$

which yields the required result upon division with $\prod_{i=1}^{k}(1 - tz(1+u)\bar{y}_i^{t-1}(z))$. The dominant term in this occurs when $i = j = k$. Since the singularities of $y_{k-1}(z)$ and $\bar{y}_{k-1}(z)$ are further away from the origin than those of $y_k(z)$ and $\bar{y}_k(z)$ they will not contribute to the main term in $g_k(z,u)$ and we do not include them. Thus the dominant term in $g_k(z,u)$ is as given in the statement of the theorem. ■

Now we substitute the known asymptotic expansions in (175) and obtain

$$g_k(z,u) \sim \frac{\sqrt{q_k - z}}{(t-1)(tq_k)^{1/(t-1)}\sqrt{q_k - z(1+u)}}\,. \quad (177)$$

which enables us to look at the behaviour of the coefficients in g_k.

4.53 Lemma *The coefficients of u^p in $g_k(z,u)$ have the form*

$$[u^p]g_k(z,u) \sim \frac{1}{2^{2p}q_k^{1/(t-1)}(t-1)t^{1/(t-1)}}\binom{2p}{p}\Big(1 - \frac{z}{q_k}\Big)^{-p}, \; z \to q_k\,. \quad (178)$$

Proof: The coefficients of u^p in (177) are computed, as follows

$$[u^p]\frac{\sqrt{q_k - z}}{(t-1)(tq_k)^{1/(t-1)}\sqrt{q_k - z(1+u)}} = \frac{(1-\frac{z}{q_k})^{1/2}}{(t-1)(tq_k)^{1/(t-1)}}[u^p]\frac{1}{(1-\frac{z}{q_k})^{1/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{1/2}}$$
$$= \frac{z^p}{2^{2p}(t-1)(tq_k)^{1/(t-1)}q_k^p}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p} \sim \frac{1}{2^{2p}(t-1)(tq_k)^{1/(t-1)}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}, \tag{179}$$

as z tends to the dominant singularity q_k. ■

Finally, we use Proposition 1.12, with $\alpha = -p$, to obtain the asymptotic expansion for the coefficients of z^n

$$[z^n u^p]g_k(z,u) \sim \frac{1}{2^{2p}q_k^{n/(t-1)}(t-1)t^{1/(t-1)}}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}. \tag{180}$$

4.54 Theorem [25] *The expectation for the size of the ancestor tree in monotonic t-ary trees is*

$$E_{n,p}^{(k)} \sim \frac{p}{2^{2p}q_k^{1/(t-1)}p_k(t-1)t^{1/(t-1)}}\binom{2p}{p}\sqrt{n}, \ n \to \infty, \ \textit{fixed } k. \tag{181}$$

One can proceed with finding the variance for the size of the ancestor tree. The second derivative of G_k with respect to v (evaluated at 1) is denoted by $h_k(z,u)$ and gives

$$\begin{aligned}(1 - tz(1+u)\bar{y}_k^{t-1}(z))h_k(z,u) = h_{k-1}(z,u) + 2tz(1+u)\bar{y}_k^{t-1}(z)g_k(z,u) \\ + tz(1+u)\bar{y}_k^{t-2}(z)g_k^2(z,u)(t-1).\end{aligned} \tag{182}$$

4.55 Lemma [27] *The solution to the recursion in (182) has the closed form*

$$\begin{aligned}h_k(z,u) = \sum_{j=1}^{k}\Big\{\Big(\prod_{i=j}^{k}(1 - tz(1+u)\bar{y}_i^{t-1}(z))\Big)^{-1}\big[2tz(1+u)\bar{y}_j^{t-1}(z)g_j(z,u) \\ + tz(1+u)\bar{y}_j^{t-2}(z)g_j^2(z,u)(t-1)\big]\Big\},\end{aligned} \tag{183}$$

and the dominant term is

$$h_k(z,u) \sim \frac{2tz(1+u)\bar{y}_k^{t-1}(z)g_k(z,u) + tz(1+u)\bar{y}_k^{t-2}(z)g_k^2(z,u)(t-1)}{1 - tz(1+u)\bar{y}_k^{t-1}(z))}. \tag{184}$$

Proof: We solve the recursion in (182) by first multiplying both sides with the product $\prod_{i=1}^{k-1}(1-tz(1+u)\bar{y}_i^{t-1}(z))$ and then summing up all the $h_k(z,u)$

$$\begin{aligned}&\prod_{i=1}^{k-1}(1-tz(1+u)\bar{y}_i^{t-1}(z))(1-tz(1+u)\bar{y}_k^{t-1}(z))h_k(z,u)\\&=\sum_{j=0}^{k-1}\Big\{\prod_{i=1}^{j}(1-tz(1+u)\bar{y}_i^{t-1}(z))\big[2tz(1+u)\bar{y}_j^{t-1}(z)g_j(z,u)\\&+tz(1+u)\bar{y}_j^{t-2}(z)g_j^2(z,u)(t-1)\big]\Big\},\end{aligned} \tag{185}$$

where $h_0(z,u)=0$. Next, we solve for $h_k(z,u)$

$$\begin{aligned}h_k(z,u)=&\sum_{j=1}^{k}\Big\{\Big(\prod_{i=j}^{k}(1-tz(1+u)\bar{y}_i^{t-1}(z))\Big)^{-1}\big[2tz(1+u)\bar{y}_j^{t-1}(z)g_j(z,u)\\&+tz(1+u)\bar{y}_j^{t-2}(z)g_j^2(z,u)(t-1)\big]\Big\}.\end{aligned} \tag{186}$$

The dominant term here occurs when $i=j=k$ and this completes the proof. ■

Furthermore, if we substitute $\bar{y}_{k-1}(z)$ and $\bar{y}_k(z)$ with their known asymptotic expressions then $h_k(z,u)$ simplifies to

$$h_k(z,u)\sim\frac{zu}{C_k(t)(t-1)^2(q_k-z(1+u))^{3/2}(tq_k)^{2/(t-1)}}\,. \tag{187}$$

Our analysis requires us to find the asymptotic behaviour of the coefficients of u^p and z^n in the simplified version of h_k.

4.56 Lemma *The coefficients of u^p in $h_k(z,u)$ have the asymptotic expansion*

$$[u^p]h_k(z,u)\sim\frac{pq_k^{-(3+t)/(2(t-1))}}{2^{2p-1}C_k(t)(t-1)^2t^{2/(t-1)}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p-1/2},\ \ z\to q_k\,. \tag{188}$$

Proof: The coefficient of u^p in (187) gives the result of the lemma

$$\begin{aligned}&[u^p]\frac{zu}{C_k(t)(t-1)^2(q_k-z(1+u))^{3/2}(tq_k)^{2/(t-1)}}\\&=\frac{z}{C_k(t)(t-1)^2q_k^{3/2}(tq_k)^{2/(t-1)}}[u^{p-1}]\frac{1}{\Big(1-\frac{z}{q_k}\Big)^{3/2}\Big(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\Big)^{3/2}}\\&=\frac{pz^p}{2^{2p-1}C_k(t)(t-1)^2q_k^{p+1/2}(tq_k)^{2/(t-1)}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p-1/2}\end{aligned}$$

$$\sim \frac{p}{2^{2p-1} C_k(t)(t-1)^2 q_k^{1/2} (tq_k)^{2/(t-1)}} \binom{2p}{p} \left(1 - \frac{z}{q_k}\right)^{-p-1/2}, \quad z \to q_k \,. \tag{189}$$

■

Proposition 1.12, with $\alpha = -p - \frac{1}{2}$, is then used to find the asymptotic expansion for the coefficients of z^n

$$[z^n u^p] h_k(z,u) \sim \frac{p q_k^{-(3+t)/(2(t-1))-n}}{2^{2p-1} C_k(t)(t-1)^2 t^{2/(t-1)}} \binom{2p}{p} \frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{190}$$

which will be normalised and give the second moment

$$\begin{aligned} &\frac{1}{p_k q_k^{-n} \frac{n^p}{\Gamma(p+1)} n^{-3/2}} [z^n u^p]\, h_k(z,u) \\ &\sim \frac{p q_k^{-(3+t)/(2(t-1))}}{2^{2p-1} C_k(t)(t-1)^2 t^{2/(t-1)}} \binom{2p}{p} \frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})} n = \frac{2p q_k^{-(3+t)/(2(t-1))}}{\sqrt{\pi} p_k C_k(t)(t-1)^2 t^{2/(t-1)}} n \,. \end{aligned} \tag{191}$$

4.57 Theorem [27] *The variance for the size of the ancestor tree in monotonic t-ary trees is*

$$V_{n,p}^{(k)} = \frac{p}{p_k (t-1)^2 t^{2/(t-1)}} \left(\frac{2}{\sqrt{\pi} C_k(t) q_k^{(3+t)/(2(t-1))}} - \frac{p}{2^{4p} q_k^{2/(t-1)} p_k} \binom{2p}{p}^2 \right) n + \mathcal{O}(\sqrt{n}\,), \tag{192}$$

for $n \to \infty$ and k fixed.

4.4.2 The Steiner distance

We begin by presenting the generating function of ordinary t-ary trees which gives us a starting point for developing the corresponding result for monotonic t-ary trees.

4.58 Theorem [27] *The generating function for the Steiner distance in t-ary trees is*

$$F(z,u,v) = \frac{zv(1+u)G^t(z,u,v) - tzvT^{t-1}(z)G(z,u,v) + zT^t(z)(v-t) + T(z)}{1 - tzT^{t-1}(z)} \,. \tag{193}$$

This result is modified, by replacing $T(z)$ with $y_k(z)$, so that it gives the Steiner distance for the t-ary tree model in [38].

4.59 Theorem [27] *The Steiner distance for monotonic t-ary trees satisfies*

$$F_k(z,u,v) = \frac{zv(1+u)G_k^t(z,u,v) - tzvy_k^{t-1}(z)G_k(z,u,v) + zy_k^t(z)(v-t) + y_k(z)}{1 - tzy_k^{t-1}(z)} \,. \tag{194}$$

It is useful to look at this parameter for some particular values of k. We consider $k = 1, 2$ and briefly state the statistics obtained for F_1 and F_2. The usual method of computing the moments is employed.

Case $k = 1$

The first derivative of F_1 with respect to v, where $v = 1$, is

$$\frac{\partial F_1(z,u,v)}{\partial v}\bigg|_{v=1} = \frac{1}{1 - tzy_1^{t-1}(z)}\Big(z(1+u)\bar{y}_1^t(z) + tz(1+u)\bar{y}_1^{t-1}(z)g_1(z,u) - tzy_1^{t-1}(z)\bar{y}_1(z,u) - tzy_1^{t-1}(z)g_1(z,u) + zy_1^t(z)\Big) \sim \frac{2t\sqrt{q_1 - z}}{(t-1)^2\sqrt{q_1 - z(1+u)}} - \frac{tzu}{(t-1)^2\sqrt{q_1 - z}\sqrt{q_1 - z(1+u)}} \tag{195}$$

4.60 Lemma *The coefficients of u^p in $\frac{\partial F_1(z,u,v)}{\partial v}\big|_{v=1}$ are*

$$[u^p]\frac{\partial F_1(z,u,v)}{\partial v}\bigg|_{v=1} \sim \frac{(p-1)t}{2^{2p-1}(2p-1)(t-1)^2}\binom{2p}{p}\left(1 - \frac{z}{q_1}\right)^{-p}, \ z \to q_1 \,. \tag{196}$$

The coefficients of z^n in the first derivative of F_1 have the form

$$[z^n u^p]\frac{\partial F_1(z,u,v)}{\partial v}\bigg|_{v=1} \sim \frac{(p-1)t}{2^{2p-1}(2p-1)(t-1)^2 q_1^n}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}, \tag{197}$$

and their normalisation leads to the expectation of the Steiner distance for $k = 1$

$$E_{n,p}^{(1)} \sim \frac{p(p-1)t}{2^{2p-1}(2p-1)(t-1)^2 p_1}\binom{2p}{p}\sqrt{n}\,. \tag{198}$$

The second derivative of F_1 with respect to v, evaluated at $v = 1$, is

$$\frac{\partial^2 F_1(z,u,v)}{\partial v^2}\bigg|_{v=1} = \frac{1}{1 - tzy_1^{t-1}(z)}\Big(2tz(1+u)\bar{y}_1^{t-1}(z)g_1(z,u) + tz(1+u)\bar{y}_1^{t-2}(z)g_1^2(z,u)(t-1) + tz(1+u)\bar{y}_1^{t-1}(z)h_1(z,u) - 2tzy_1^{t-1}(z)g_1(z,u) - tzy_1^{t-1}(z)h_1(z,u)\Big) \sim \frac{t^2 zu}{C_1(t)(t-1)^4(q_1 - z(1+u))^{3/2}} - \frac{2t^2}{C_1(t)(t-1)^4\sqrt{q_1 - z(1+u)}} \tag{199}$$

4.61 Lemma *The coefficients of u^p in $\frac{\partial^2 F_1(z,u,v)}{\partial v^2}\big|_{v=1}$ have the form*

$$[u^p]\frac{\partial^2 F_1(z,u,v)}{\partial v^2}\bigg|_{v=1} \sim \frac{(p-1)t^2}{C_1(t)2^{2p-1}(t-1)^4 q_2^{1/2}}\binom{2p}{p}\left(1 - \frac{z}{q_1}\right)^{-p-1/2}, \ z \to q_1 \,. \tag{200}$$

We use Proposition 1.12 to compute the coefficients of z^n in the second derivative of F_1 as follows

$$[z^n u^p]\frac{\partial^2 F_1(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{(p-1)t^2}{C_1(t)2^{2p-1}(t-1)^4 q_2^{n+1/2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{201}$$

which, upon normalisation, give us the second moment and this will be used to compute the variance of the Steiner distance when $k = 1$

$$V_{n,p}^{(1)} \sim \frac{(p-1)t^2}{p_1(t-1)^4}\Big(\frac{2}{\sqrt{\pi}C_1(t)q_1^{1/2}} - \frac{p^2(p-1)}{2^{4p-2}(2p-1)^2 p_1}\binom{2p}{p}^2\Big)n. \tag{202}$$

Case $k = 2$

The first derivative of F_2 with respect to v, where $v = 1$, is

$$\frac{\partial F_2(z,u,v)}{\partial v}\Big|_{v=1} = \frac{1}{1-tzy_2^{t-1}(z)}\Big(z(1+u)\bar{y}_2^t(z) + tz(1+u)\bar{y}_2^{t-1}(z)g_2(z,u)$$
$$- tzy_2^{t-1}(z)\bar{y}_2(z,u) - tzy_2^{t-1}(z)g_2(z,u) + zy_2^t(z)\Big)$$
$$\sim \frac{2t\sqrt{q_2-z}}{(t-1)\big(t-1-(\frac{t-1}{t})^t\big)\sqrt{q_2-z(1+u)}} - \frac{tzu}{(t-1)\big(t-1-(\frac{t-1}{t})^t\big)\sqrt{q_2-z(1+u)}} \tag{203}$$

4.62 Lemma *The coefficients of u^p in $\frac{\partial F_2(z,u,v)}{\partial v}\big|_{v=1}$ are*

$$[u^p]\frac{\partial F_2(z,u,v)}{\partial v}\Big|_{v=1} \sim \frac{(p-1)t^{t+1}}{2^{2p-1}(2p-1)(t-1)^2(t^t-(t-1)^{t-1})}\binom{2p}{p}\Big(1-\frac{z}{q_2}\Big)^{-p}, \quad z \to q_2. \tag{204}$$

The coefficients of z^n in the first derivative of F_2 have the form

$$[z^n u^p]\frac{\partial F_2(z,u,v)}{\partial v}\Big|_{v=1} \sim \frac{(p-1)t^{t+1}}{2^{2p-1}(2p-1)(t-1)^2(t^t-(t-1)^{t-1})q_2^n}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}, \tag{205}$$

and their normalisation leads to the expectation of the Steiner distance for $k = 2$

$$E_{n,p}^{(2)} \sim \frac{p(p-1)t^{t+1}}{2^{2p-1}(2p-1)(t-1)^2(t^t-(t-1)^{t-1})p_2}\binom{2p}{p}\sqrt{n}. \tag{206}$$

The second derivative of F_2 with respect to v, evaluated at $v = 1$, is

$$\begin{aligned}\frac{\partial^2 F_2(z,u,v)}{\partial v^2}\Big|_{v=1} &= \frac{1}{1-tzy_2^{t-1}(z)}\Big(2tz(1+u)\bar{y}_2^{t-1}(z)g_2(z,u) + tz(1+u)\bar{y}_2^{t-2}(z)g_2^2(z,u)(t-1) \\ &\quad + tz(1+u)\bar{y}_2^{t-1}(z)h_2(z,u) - 2tzy_2^{t-1}(z)g_2(z,u) - tzy_2^{t-1}(z)h_2(z,u)\Big) \\ &\sim \frac{t^2zu}{C_2(t)(t-1)^2\big(t-1-(\frac{t-1}{t})^t\big)^2(q_2-z(1+u))^{3/2}} \\ &\quad - \frac{2t^2}{C_2(t)(t-1)^2\big(t-1-(\frac{t-1}{t})^t\big)^2\sqrt{q_2-z(1+u)}}\,.\end{aligned} \tag{207}$$

4.63 Lemma *As $z \to q_k$, the coefficients of u^p in $\frac{\partial^2 F_2(z,u,v)}{\partial v^2}\big|_{v=1}$ have the form*

$$[u^p]\frac{\partial^2 F_2(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{(p-1)t^{2t+2}}{C_2(t)2^{2p-1}(t-1)^4(t^t-(t-1)^{t-1})^2 q_2^{1/2}}\binom{2p}{p}\Big(1-\frac{z}{q_2}\Big)^{-p-1/2}. \tag{208}$$

Proposition 1.12 is used to find the coefficients of z^n in the second derivative of F_2 as follows

$$[z^n u^p]\frac{\partial^2 F_2(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{(p-1)t^{2t+2}}{C_2(t)2^{2p-1}(t-1)^4(t^t-(t-1)^{t-1})^2 q_2^{n+1/2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{209}$$

and they will be normalised in order to compute the second moment. The latter is employed in the derivation of the variance of the Steiner distance for $k = 2$

$$V_{n,p}^{(2)} \sim \frac{(p-1)t^{2t+2}}{(t-1)^4(t^t-(t-1)^{t-1})^2 p_2}\left(\frac{2}{\sqrt{\pi}C_2(t)q_2^{1/2}} - \frac{p^2(p-1)}{2^{4p-2}(2p-1)^2 p_2}\binom{2p}{p}^2\right)n\,. \tag{210}$$

The focus of the analysis shifts to the general case now. The first moment is computed by differentiating F_k with respect to v and evaluating at $v = 1$. Then the known asymptotic formulas are substituted in to simplify the derivative

$$\begin{aligned}\frac{\partial F_k(z,u,v)}{\partial v}\Big|_{v=1} &= \frac{1}{1-tzy_k^{t-1}(z)}\Big(z(1+u)\bar{y}_k^t(z) + tz(1+u)\bar{y}_k^{t-1}(z)g_k(z,u) \\ &\quad - tzy_k^{t-1}(z)\bar{y}_k(z,u) - tzy_k^{t-1}(z)g_k(z,u) + zy_k^t(z)\Big) \\ &\sim \frac{2\sqrt{q_k-z}}{(t-1)(tq_k)^{1/(t-1)}\sqrt{q_k-z(1+u)}} \\ &\quad - \frac{zu}{(t-1)(tq_k)^{1/(t-1)}\sqrt{q_k-z}\sqrt{q_k-z(1+u)}}\,,\end{aligned} \tag{211}$$

and this enables us to consider the behaviour of the coefficients of u^p around their dominant singularity q_k.

4.64 Lemma *For $z \to q_k$, the coefficients of u^p in $\frac{\partial F_k(z,u,v)}{\partial v}\big|_{v=1}$ are*

$$[u^p]\frac{\partial F_k(z,u,v)}{\partial v}\Big|_{v=1} \sim \frac{p-1}{2^{2p-1}(2p-1)(t-1)(tq_k)^{1/(t-1)}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}. \tag{212}$$

Proof: The coefficient of u^p from the first term in (211) is

$$[u^p]\frac{2\sqrt{q_k - z}}{(t-1)(tq_k)^{1/(t-1)}\sqrt{q_k - z(1+u)}} = \frac{2\left(1-\frac{z}{q_k}\right)^{1/2}}{(t-1)(tq_k)^{1/(t-1)}}[u^p]\frac{1}{\left(1-\frac{z}{q_k}\right)^{1/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{1/2}}$$

$$= \frac{z^p}{2^{2p-1}(t-1)(tq_k)^{1/(t-1)}q_k^p}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p} \sim \frac{1}{2^{2p-1}(t-1)(tq_k)^{1/(t-1)}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}, \tag{213}$$

for $z \to q_k$. Moreover the second term yields

$$\begin{aligned} &-[u^p]\frac{zu}{(t-1)(tq_k)^{1/(t-1)}\sqrt{q_k - z}\sqrt{q_k - z(1+u)}} \\ &\sim -\frac{p}{2^{2p-1}(2p-1)(t-1)(tq_k)^{1/(t-1)}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}, \end{aligned} \tag{214}$$

around the dominant singularity q_k. Adding these two coefficients gives the result of the lemma. ■

Next, the coefficients of z^n are computed as follows

$$[z^n u^p]\frac{\partial F_k(z,u,v)}{\partial v}\Big|_{v=1} \sim \frac{(p-1)}{2^{2p-1}(2p-1)(t-1)(tq_k)^{1/(t-1)}q_k^n}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}, \tag{215}$$

which will be normalised and give the first moment.

4.65 Theorem [27] *The expectation of the Steiner distance in monotonic t-ary trees is*

$$E_{n,p}^{(k)} \sim \frac{p(p-1)}{2^{2p-1}(2p-1)(t-1)(tq_k)^{1/(t-1)}p_k}\binom{2p}{p}\sqrt{n}\,,\ n \to \infty\,,\ \textit{fixed } k\,. \tag{216}$$

We proceed by computing the next moment for the Steiner distance. After differentiating F_k twice with respect to v (and evaluating at 1) we get

$$\frac{\partial^2 F_k(z,u,v)}{\partial v^2}\Big|_{v=1} = \frac{1}{1 - tzy_k^{t-1}(z)}\Big(2tz(1+u)\bar{y}_k^{t-1}(z)g_k(z,u) + tz(1+u)\bar{y}_k^{t-2}(z)g_k^2(z,u)(t-1)$$

$$+ tz(1+u)\bar{y}_k^{t-1}(z)h_k(z,u) - 2tzy_k^{t-1}(z)g_k(z,u) - tzy_k^{t-1}(z)h_k(z,u)\Big)$$
$$\sim \frac{zu}{C_k(t)(t-1)^2(tq_k)^{2/(t-1)}(q_k - z(1+u))^{3/2}} - \frac{2}{C_k(t)(t-1)^2(tq_k)^{2/(t-1)}\sqrt{q_k - z(1+u)}} . \tag{217}$$

This asymptotic expression makes it easy for us to analyse the behaviour of its coefficients.

4.66 Lemma *The coefficients of u^p in $\frac{\partial^2 F_k(z,u,v)}{\partial v^2}\Big|_{v=1}$ are*

$$[u^p]\frac{\partial^2 F_k(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{p-1}{2^{2p-1}C_k(t)(t-1)^2 t^{2/(t-1)} q_k^{(3+t)/(2(t-1))}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2}, \; z \to q_k . \tag{218}$$

Proof: The coefficients of u^p from the terms in (217) are extracted as follows

$$[u^p]\frac{zu}{C_k(t)(t-1)^2(tq_k)^{2/(t-1)}(q_k - z(1+u))^{3/2}}$$
$$= \frac{z}{C_k(t)(t-1)^2(tq_k)^{2/(t-1)}q_k^{3/2}}[u^{p-1}]\frac{1}{\left(1-\frac{z}{q_k}\right)^{3/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{3/2}}$$
$$= \frac{pz^p}{2^{2p-1}C_k(t)(t-1)^2(tq_k)^{2/(t-1)}q_k^{p+1/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2}$$
$$\sim \frac{p}{2^{2p-1}C_k(t)(t-1)^2(tq_k)^{2/(t-1)}q_k^{1/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2}, \; z \to q_k , \tag{219}$$

and similarly

$$-[u^p]\frac{2}{C_k(t)(t-1)^2(tq_k)^{2/(t-1)}\sqrt{q_k - z(1+u)}}$$
$$\sim \frac{1}{2^{2p-1}C_k(t)(t-1)^2(tq_k)^{2/(t-1)}q_k^{1/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2}, \; z \to q_k . \tag{220}$$

By adding these two coefficients we obtain the result of the lemma. ■

The last required coefficients are:

$$[z^n u^p]\frac{\partial^2 F_k(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{p-1}{2^{2p-1}C_k(t)(t-1)^2 t^{2/(t-1)} q_k^{(3+t)/(2(t-1))} q_k^n}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})},$$

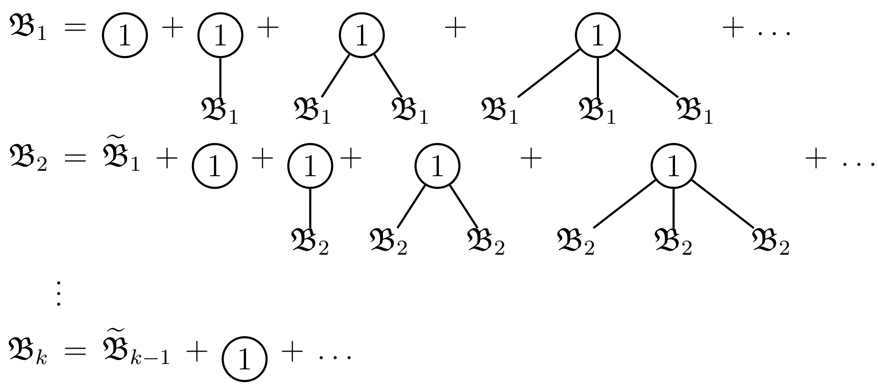
and after normalisation they give rise to the second moment.

4.67 Theorem [27] *The variance for the Steiner distance in monotonic t-ary trees is*

$$V_{n,p}^{(k)} = \frac{(p-1)}{p_k(t-1)^2 t^{2/(t-1)}}\Big(\frac{2}{\sqrt{\pi}C_k(t)q_k^{(3+t)/(2(t-1))}p_k} - \frac{p^2(p-1)}{2^{4p-2}(2p-1)^2 q_k^{2/(t-1)}p_k}\binom{2p}{p}^2\Big)n + \mathcal{O}(\sqrt{n}), \ n \to \infty, \ \textit{fixed } k. \tag{221}$$

4.5 Monotonic ordered trees

We now consider the class, $\mathfrak{B}_k$, of ordered trees whose nodes are monotonically labelled with $1, 2, \ldots, k$ and we let $\widetilde{\mathfrak{B}_k}$ be the class of ordered trees whose nodes are monotonically labelled with $2, \ldots, k+1$. The defining equations for the classes $\mathfrak{B}_k$ are given in [38] by



4.68 Theorem [38] *The generating functions for monotonic ordered trees are*

$$\begin{aligned} y_1(z) &= \frac{z}{1-y_1(z)}, \\ y_2(z) &= y_1(z) + \frac{z}{1-y_2(z)}, \\ &\cdots \\ y_k(z) &= y_{k-1}(z) + \frac{z}{1-y_k(z)}. \end{aligned} \tag{222}$$

The results in this section contain various constants depending on k. They are illustrated in the table below and specific values are given for the first few cases of k.

constant	$k=1$	$k=2$	$k=3$
$q_k = \frac{1}{r_k^2}$	$1/4$	$4/25$	$100/841$
$r_{k+1} = r_k + \frac{1}{r_k}$; $r_0 = 1$	2	$5/2$	$29/10$
$y_k(q_k) = 1 - \frac{1}{r_k}$	$1/2$	$3/5$	$19/29$

4.5.1 Size of the ancestor tree

In chapter one, we have presented the generating function for ordered trees. We can easily adapt that result and obtain the functional equation for monotonic ordered trees.

4.69 Theorem [25] *The generating functions for the size of the ancestor tree in monotonic ordered trees are*

$$P_k(z,u,v) = P_{k-1}(z,u,v) + \frac{zv(1+u)}{1-P_k(z,u,v)} + \frac{z(1-v)}{1-y_k(z)}\,,\ P_0(z,u,v) = 0\,. \tag{223}$$

As usual, statistics for this parameter will be computed. We begin by looking at two particular values of k.

Case $k = 1$

After differentiating P_1 with respect to v and then setting $v = 1$, we get

$$\begin{aligned} A_1(z,u) &= \left(1 - \frac{z(1+u)}{(1-\bar{y}_1(z))^2}\right)^{-1}\left(\frac{z(1+u)}{1-\bar{y}_1(z)} - \frac{z}{1-y_1(z)}\right) \\ &= \left(1 - \frac{\bar{y}_1(z)}{1-\bar{y}_1(z)}\right)^{-1}(\bar{y}_1(z) - y_1(z))\,, \end{aligned} \tag{224}$$

where $\frac{\partial P_1(z,u,v)}{\partial v}\Big|_{v=1} = A_1(z,u)$ and $P_1(z,u,1) = y_1(z(1+u)) = \bar{y}_1(z)$. An asymptotic expression for the first derivative of P_1 around the dominant singularity $q_1 = \frac{1}{4}$ will be obtained by substituting the known expansions in the above

$$A_1(z,u) \sim \frac{\sqrt{q_1 - z}}{4\sqrt{q_1 - z(1+u)}}\,, \tag{225}$$

which enables us to look at the behaviour of its coefficients.

4.70 Lemma *The coefficients of u^p in $A_1(z,u)$ are*

$$[u^p]A_1(z,u) \sim \frac{1}{2^{2p+2}}\binom{2p}{p}(1-4z)^{-p}\,,\ z \to \frac{1}{4}\,. \tag{226}$$

Next, the coefficients of z^n in A_1 will be computed using Proposition 1.12

$$[z^n u^p]A_1(z,u) \sim \frac{4^n}{2^{2p+2}}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}, \tag{227}$$

and upon normalisation by $p_1 q_1^{-n} n^{-3/2} \frac{n^p}{\Gamma(p+1)}$ (where $p_1 = \frac{1}{4\sqrt{\pi}}$) they yield the expectation of the ancestor tree for $k = 1$

$$E_{n,p}^{(1)} \sim \frac{p\sqrt{\pi}}{2^{2p}}\binom{2p}{p}\sqrt{n}\,. \tag{228}$$

The second moment can be computed now. By denoting the second derivative of P_1 with respect to v (evaluated at $v = 1$) with $B_1(z,u)$ one gets

$$\begin{aligned} B_1(z,u) &= \Big(1 - \frac{z(1+u)}{(1-\bar{y}_1(z))^2}\Big)^{-1}\Big(\frac{2z(1+u)A_1(z,u)}{(1-\bar{y}_1(z))^2} + \frac{2z(1+u)A_1^2(z,u)}{(1-\bar{y}_1(z))^3}\Big) \\ &= \Big(1 - \frac{\bar{y}_1(z)}{1-\bar{y}_1(z)}\Big)^{-1}\Big(\frac{2\bar{y}_1(z)A_1(z,u)}{1-\bar{y}_1(z)} + \frac{2\bar{y}_1(z)A_1^2(z,u)}{(1-\bar{y}_1(z))^2}\Big) \\ &\sim \frac{zu}{16d_1(q_1 - z(1+u))^{3/2}}\,, \end{aligned} \tag{229}$$

around the singularity q_1, where $d_1 = 1$. This local expansion helps in determining the behaviour of the coefficients of u^p and z^n in B_1.

4.71 Lemma *The coefficients of u^p in $B_1(z,u)$ have the form*

$$[u^p]B_1(z,u) \sim \frac{p}{2^{2p+2}}\binom{2p}{p}\Big(1-4z\Big)^{-p-1/2}, \quad z \to \frac{1}{4}\,. \tag{230}$$

Furthermore, the coefficients of z^n are

$$[z^n u^p]B_1(z,u) \sim \frac{4^n p}{2^{2p+2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{231}$$

which will be normalised in order to compute the second moment

$$\frac{p\sqrt{\pi}}{2^{2p}}\binom{2p}{p}\frac{\Gamma(p+1)}{\Gamma(p+\frac{1}{2})}\,n\,. \tag{232}$$

Then, the variance for the size of the ancestor tree is

$$V_{n,p}^{(1)} \sim p\Big(1 - \frac{p\pi}{2^{4p}}\binom{2p}{p}^2\Big)n\,, \tag{233}$$

where $1 - \frac{p\pi}{2^{4p}}\binom{2p}{p}^2 \to 0$ as $p \to \infty$.

4.72 Note *The asymptotic values of* $E^{(1)}_{n,p}$ *and* $V^{(1)}_{n,p}$ *obtained above coincide with those of the ordered trees considered in their chapter.*

Case $k = 2$

The derivative of P_2 with respect to v, evaluated at $v = 1$, which is denoted by $A_2(z,u)$ gives the following

$$\begin{aligned} A_2(z,u) &= \Big(1 - \frac{z(1+u)}{(1-\bar{y}_2(z))^2}\Big)^{-1}\Big(A_1(z,u) + \frac{z(1+u)}{1-\bar{y}_2(z)} - \frac{z}{1-y_2(z)}\Big) \\ &= \Big(1 - \frac{(\bar{y}_2(z)-\bar{y}_1(z))}{1-\bar{y}_2(z)}\Big)^{-1}(A_1(z,u) + \bar{y}_2(z) - \bar{y}_1(z) - y_2(z) + y_1(z)) \\ &\sim \frac{\sqrt{q_2 - z}}{5\sqrt{q_2 - z(1+u)}}, \end{aligned} \tag{234}$$

and this expansion makes it possible to look at the behaviour of its coefficients.

4.73 Lemma *The coefficients of* u^p *in* $A_2(z,u)$ *are*

$$[u^p]A_2(z,u) \sim \frac{1}{5\cdot 2^{2p}}\binom{2p}{p}\Big(1 - \frac{25z}{4}\Big)^{-p-1/2}, \ z \to \frac{4}{25}. \tag{235}$$

Then the coefficients of z^n in A_2 have the form

$$[z^n u^p]A_2(z,u) \sim \frac{1}{5\cdot 2^{2p}}\Big(\frac{25}{4}\Big)^n \binom{2p}{p}\frac{n^p}{\Gamma(p+1)}, \tag{236}$$

so their normalisation leads to the expectation for the size of the ancestor tree

$$E^{(2)}_{n,p} \sim \frac{p}{5\cdot 2^{2p} p_2}\binom{2p}{p}\sqrt{n}. \tag{237}$$

One can proceed with computing the next moment. Let $B_2(z,u)$ be the second derivative of P_2 with respect to v (evaluated at 1). Then we obtain

$$\begin{aligned} B_2(z,u) &= \Big(1 - \frac{z(1+u)}{(1-\bar{y}_2(z))^2}\Big)^{-1}\Big(B_1(z,u) + \frac{2z(1+u)A_2(z,u)}{(1-\bar{y}_2(z))^2} + \frac{2z(1+u)A_2^2(z,u)}{(1-\bar{y}_2(z))^3}\Big) \\ &= \Big(1 - \frac{(\bar{y}_2(z)-\bar{y}_1(z))}{1-\bar{y}_2(z)}\Big)^{-1}\Big(B_1(z,u) + \frac{2(\bar{y}_2(z)-y_1(z))A_2(z,u)}{1-\bar{y}_2(z)} + \frac{2(\bar{y}_2(z)-y_1(z))A_2^2(z,u)}{(1-\bar{y}_2(z))^2}\Big) \\ &\sim \frac{zu}{25 d_2 (q_2 - z(1+u))^{3/2}}, \end{aligned} \tag{238}$$

and the behaviour of the coefficients of u^p around q_2 is presented in the next lemma.

4.74 Lemma *The coefficients of u^p in $B_2(z,u)$ are*

$$[u^p]B_2(z,u) \sim \frac{p}{25 \cdot 2^{2p-1} d_2} \binom{2p}{p} \Big(1 - \frac{25z}{4}\Big)^{-p-1/2}, \quad z \to \frac{4}{25}. \tag{239}$$

The other required coefficients in the second derivative of P_2 are given by

$$[z^n u^p]B_2(z,u) \sim \frac{p}{25 \cdot 2^{2p-1} d_2} \Big(\frac{25}{4}\Big)^n \binom{2p}{p} \frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{240}$$

which will be normalised in order to obtain the second moment. Finally, the variance for the size of the ancestor tree in the case $k = 2$ is

$$V_{n,p}^{(2)} \sim \frac{p}{25 p_2} \Big(\frac{2}{\sqrt{\pi} d_2} - \frac{p}{2^{4p} p_2} \binom{2p}{p}^2 \Big) n. \tag{241}$$

The results presented above will be generalised now. The expectation for the size of the ancestor tree is the first quantity of interest. We denote the derivative of P_k with respect to v (evaluated at 1) by $A_k(z,u)$ and let $P_k(z,u,1) = y_k(z(1+u)) = \bar{y}_k(z)$. This gives

$$\begin{aligned} A_k(z,u) &= \Big(1 - \frac{z(1+u)}{(1-\bar{y}_k(z))^2}\Big)^{-1} \Big(A_{k-1}(z,u) + \frac{z(1+u)}{1-\bar{y}_k(z)} - \frac{z}{1-y_k(z)}\Big) \\ &= \Big(1 - \frac{(\bar{y}_k(z) - \bar{y}_{k-1}(z))}{1-\bar{y}_k(z)}\Big)^{-1} \Big(A_{k-1}(z,u) + \bar{y}_k(z) - \bar{y}_{k-1}(z) - y_k(z) + y_{k-1}(z)\Big). \end{aligned} \tag{242}$$

4.75 Lemma [27] *The solution to the recursion in (242) is*

$$A_k(z,u) = \sum_{j=1}^{k} \Big[\prod_{i=1}^{j} \Big(1 - \frac{z(1+u)}{(1-\bar{y}_i(z))^2}\Big)^{-1} \Big(\frac{z(1+u)}{1-\bar{y}_j(z)} - \frac{z}{1-y_j(z)}\Big)\Big], \tag{243}$$

with $A_0(z,u) = 0$, hence the dominant term has the asymptotic form

$$A_k(z,u) \sim \Big(1 - \frac{z(1+u)}{(1-\bar{y}_k(z))^2}\Big)^{-1} \Big(\frac{z(1+u)}{1-\bar{y}_k(z)} - \frac{z}{1-y_k(z)}\Big). \tag{244}$$

Proof: We multiply both sides of (242) with the product $\prod_{i=0}^{k-1}\left(1-\frac{z(1+u)}{(1-\bar{y}_i(z))^2}\right)$ then sum over the k's and obtain

$$\begin{aligned}&\prod_{i=0}^{k-1}\Big(1-\frac{z(1+u)}{(1-\bar{y}_i(z))^2}\Big)\Big(1-\frac{z(1+u)}{(1-\bar{y}_k(z))^2}\Big)A_k(z,u)\\ &=\sum_{j=0}^{k-1}\Big[\prod_{i=0}^{j}\Big(1-\frac{z(1+u)}{(1-\bar{y}_i(z))^2}\Big)\Big(\frac{z(1+u)}{1-\bar{y}_j(z)}-\frac{z}{1-y_j(z)}\Big)\Big]\,.\end{aligned} \tag{245}$$

The result of the lemma is obtained by dividing this with $\prod_{i=0}^{k}\left(1-\frac{z\,(1+u)}{(1-\bar{y}_i(z))^2}\right)$ and the dominant term occurs when $i=j=k$. ∎

Now from [38] one knows the following: the local expansion of the generating function

$$\begin{aligned}y_k(z)&=y_k(q_k)-d_k\sqrt{q_k-z}+\mathcal{O}(q_k-z)\\ &=1-\sqrt{q_k}-d_k\sqrt{q_k-z}+\mathcal{O}(q_k-z)\,,\end{aligned} \tag{246}$$

as $z\to q_k$. The recursion of the y_k's around the dominant singularities is

$$y_{k-l}(q_k)=1-r_l w\,, \tag{247}$$

for $w=\frac{1}{r_l}=\sqrt{q_l}$ and one also has

$$y_{k-1}(q_k)=1-r_1 w=1-2\sqrt{q_k}\,. \tag{248}$$

We observe that $\frac{z}{1-y_k(z)}=y_k(z)-y_{k-1}(z)$ (and similarly for $\bar{y}_k(z)$), so these are substituted into (244)

$$A_k(z,u)\sim\frac{\sqrt{q_k}\sqrt{q_k-z}}{2\sqrt{q_k-z(1+u)}}\,. \tag{249}$$

This expression for the first derivative makes it possible for us to find the asymptotic behaviour of the coefficients of u^p and z^n around their dominant singularities q_k.

4.76 Lemma *The coefficients of u^p in $A_k(z,u)$ are*

$$[u^p]A_k(z,u)\sim\frac{q_k^{1/2}}{2^{2p+1}}\binom{2p}{p}\Big(1-\frac{z}{q_k}\Big)^{-p}\,,\ z\to q_k\,. \tag{250}$$

Proof: We proceed by finding the coefficient of u^p from (249), which proves the result

$$[u^p]\frac{\sqrt{q_k}\sqrt{q_k-z}}{2\sqrt{q_k-z(1+u)}} = \frac{q_k^{1/2}\left(1-\frac{z}{q_k}\right)^{1/2}}{2}[u^p]\frac{1}{\left(1-\frac{z}{q_k}\right)^{1/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{1/2}} \tag{251}$$
$$= \frac{q_k^{1/2-p}z^p}{2^{2p+1}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p} \sim \frac{q_k^{1/2}}{2^{2p+1}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p}, \; z \to q_k \,.$$

■

Next, the coefficients of z^n are computed

$$[z^n u^p]A_k(z,u) \sim \frac{q_k^{1/2-n}}{2^{2p+1}}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}\,, \tag{252}$$

which will be normalised in order to find the first moment.

4.77 Theorem [27] *The expectation for the size of the ancestor tree in monotonic ordered trees is asymptotically given by*

$$E_{n,p}^{(k)} \sim \frac{p\sqrt{q_k}}{2^{2p+1}\,p_k}\binom{2p}{p}\sqrt{n}\,, \; n \to \infty\,, \; \textit{fixed } k. \tag{253}$$

Now we move on and compute the variance. For convenience, the second derivative of P_k with respect to v, evaluated at $v=1$, is denoted by $B_k(z,u)$, and it follows that

$$\left(1-\frac{z(1+u)}{(1-\bar{y}_k(z))^2}\right)B_k(z,u) = B_{k-1}(z,u) + \frac{2z(1+u)A_k(z,u)}{(1-\bar{y}_k(z))^2} + \frac{2z(1+u)A_k^2(z,u)}{(1-\bar{y}_k(z))^3}\,. \tag{254}$$

4.78 Lemma [27] *The solution of the recursion in (254) is*

$$B_k(z,u) = \sum_{j=1}^{k}\left[\prod_{i=1}^{j}\left(1-\frac{z(1+u)}{(1-\bar{y}_i(z))^2}\right)^{-1}\left(\frac{2z(1+u)A_j(z,u)}{(1-\bar{y}_j(z))^2} + \frac{2z(1+u)A_j^2(z,u)}{(1-\bar{y}_j(z))^3}\right)\right], \tag{255}$$

with dominant term

$$B_k(z,u) \sim \left(1-\frac{z(1+u)}{(1-\bar{y}_k(z))^2}\right)^{-1}\left(\frac{2z(1+u)A_k(z,u)}{(1-\bar{y}_k(z))^2} + \frac{2z(1+u)A_k^2(z,u)}{(1-\bar{y}_k(z))^3}\right). \tag{256}$$

Proof: By multiplying both sides of (254) with the product $\prod_{i=0}^{k-1}\left(1-\frac{z(1+u)}{(1-\bar{y}_i(z))^2}\right)$ and then summing all the k's, the recursion becomes

$$\prod_{i=0}^{k-1}\left(1-\frac{z(1+u)}{(1-\bar{y}_i(z))^2}\right)\left(1-\frac{z(1+u)}{(1-\bar{y}_k(z))^2}\right)B_k(z,u)$$
$$=\sum_{j=0}^{k-1}\left[\prod_{i=0}^{j}\left(1-\frac{z\,(1+u)}{(1-\bar{y}_i(z))^2}\right)\left(\frac{2z\,(1+u)\,A_j(z,u)}{(1-\bar{y}_j(z))^2}+\frac{2z\,(1+u)\,A_j^2(z,u)}{(1-\bar{y}_j(z))^3}\right)\right], \tag{257}$$

or simply

$$B_k(z,u)=\sum_{j=1}^{k}\left[\prod_{i=1}^{j}\left(1-\frac{z(1+u)}{(1-\bar{y}_i(z))^2}\right)^{-1}\left(\frac{2z(1+u)A_j(z,u)}{(1-\bar{y}_j(z))^2}+\frac{2z(1+u)A_j^2(z,u)}{(1-\bar{y}_j(z))^3}\right)\right], \tag{258}$$

and from this the dominant term arises for $i=j=k$. ■

After substituting in the known asymptotic expansions, (256) simplifies to the following

$$B_k(z,u)\sim\frac{zuq_k}{4d_k(q_k-z(1+u))^{3/2}}\,, \tag{259}$$

and this expression is used in analysing the behaviour of its coefficients.

4.79 Lemma *The coefficients of u^p in $B_k(z,u)$ are*

$$[u^p]B_k(z,u)\sim\frac{p\,q_k^{1/2}}{d_k\,2^{2p+1}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2},\ z\to q_k\,. \tag{260}$$

Proof: The extraction of the coefficient of u^p in (259) gives the following

$$[u^p]\frac{zuq_k}{4d_k(q_k-z(1+u))^{3/2}}=\frac{z}{4d_kq_k^{1/2}}[u^{p-1}]\frac{1}{\left(1-\frac{z}{q_k}\right)^{3/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{3/2}}$$
$$=\frac{pz^p}{2^{2p+1}d_kq_k^{p-1/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2}\sim\frac{pq_k^{1/2}}{2^{2p+1}d_k}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2},\ z\to q_k\,. \tag{261}$$

■

Moreover, the coefficient of z^n

$$[z^nu^p]B_k(z,u)\sim\frac{p\,q_k^{1/2-n}}{d_k\,2^{2p+1}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}=\frac{p\,q_k^{1-n}}{\sqrt{\pi}p_k2^{2p+2}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}\,, \tag{262}$$

and it yields the second moment upon normalisation.

4.80 Theorem [27] *The variance for the size of the ancestor tree in monotonic ordered trees is asymptotically*

$$V_{n,p}^{(k)} = \frac{pq_k}{p_k^2}\Big(\frac{1}{4\pi} - \frac{p}{2^{4p+2}}\binom{2p}{p}^2\Big)n + \mathcal{O}(\sqrt{n}\,)\,,\ n \to \infty\,,\ \textit{fixed } k\,, \tag{263}$$

and $\frac{1}{4\pi} - \frac{p}{2^{4p+2}}\binom{2p}{p}^2 \to 0$ *as* $p \to \infty$.

4.5.2 The Steiner distance

Once again, the derivation of the generating function for this parameter is most convenient when we use the defining equation of the Steiner distance in ordered trees as a starting point. This was presented already in the previous chapter on ordered trees, but for convenience we give it below

$$S(z,u,v) = \frac{P(z,u,v)(1 - \frac{zv}{(1-N(z))^2}) - \frac{z(1-v)N(z)}{(1-N(z))^2}}{1 - \frac{z}{(1-N(z))^2}}\,. \tag{264}$$

Then, one can modify the above to produce an equivalent result for monotonic ordered trees.

4.81 Theorem [25] *The generating function for the Steiner distance in monotonic ordered trees is given by*

$$S_k(z,u,v) = \frac{P_k(z,u,v)\Big(1 - \frac{zv}{(1-y_k(z))^2}\Big) - \frac{z(1-v)y_k(z)}{(1-y_k(z))^2}}{1 - \frac{z}{(1-y_k(z))^2}}\,. \tag{265}$$

As usual we first settle particular cases of k before looking at the general analysis. For $k=1$ this has already been done in the chapter on ordered trees. Therefore we consider the case $k=2$ only.

Case $k = 2$

In order to find the expectation S_2 is differentiated with respect to v (and evaluated at $v=1$)

$$\frac{\partial S_2(z,u,v)}{\partial v}\Big|_{v=1} = \frac{A_2(z,u)\Big(1 - \frac{z}{(1-y_2(z))^2}\Big) - \frac{z\bar{y}_2(z)}{(1-y_2(z))^2} + \frac{zy_2(z)}{(1-y_2(z))^2}}{1 - \frac{z}{(1-y_2(z))^2}}$$
$$= \frac{A_2(z,u)\Big(1 - \frac{(y_2(z)-y_1(z))}{1-y_2(z)}\Big) - \frac{\bar{y}_2(z)(y_2(z)-y_1(z))}{1-y_2(z)} + \frac{y_2(z)(y_2(z)-y_1(z))}{1-y_2(z)}}{1 - \frac{(y_2(z)-y_1(z))}{1-y_2(z)}}$$

$$\sim \frac{2\sqrt{q_2 - z}}{5\sqrt{q_2 - z(1+u)}} - \frac{zu}{5\sqrt{q_2 - z}\sqrt{q_2 - z(1+u)}}, \tag{266}$$

where A_2 is the derivative of P_2 as defined in the section for the size of the ancestor tree. Then the behaviour of the coefficients around q_2 will be investigated below.

4.82 Lemma *The coefficients of u^p in $\frac{\partial S_2(z,u,v)}{\partial v}\Big|_{v=1}$ are*

$$[u^p]\frac{\partial S_2(z,u,v)}{\partial v}\Big|_{v=1} \sim \frac{p-1}{5(2p-1)2^{2p-1}}\binom{2p}{p}\left(1 - \frac{25z}{4}\right)^{-p}, \quad z \to \frac{4}{25}. \tag{267}$$

The coefficients of z^n in the first derivative of S_2 are

$$[z^n u^p]\frac{\partial S_2(z,u,v)}{\partial v}\Big|_{v=1} \sim \frac{p-1}{5(2p-1)2^{2p-1}q_2^n}\binom{2p}{p}\frac{n^{p-1}}{\Gamma(p)}, \tag{268}$$

and after normalisation they lead to the expectation for the Steiner distance

$$E^{(2)}_{n,p} \sim \frac{p-1}{5(2p-1)2^{2p-1}p_2}\binom{2p}{p}\sqrt{n}. \tag{269}$$

Now the second moment will be computed. The second derivative of S_2 with respect to v, at $v = 1$ is

$$\begin{aligned} \frac{\partial^2 S_2(z,u,v)}{\partial v^2}\Big|_{v=1} &= \frac{B_2(z,u)\left(1 - \frac{(y_2(z)-y_1(z))}{1-y_2(z)}\right) - \frac{2A_2(z,u)(y_2(z)-y_1(z))}{1-y_2(z)}}{1 - \frac{(y_2(z)-y_1(z))}{1-y_2(z)}} \\ &\sim \frac{3zu}{25d_2(q_2 - z(1+u))^{3/2}} - \frac{2(q_2 - z)}{25d_2(q_2 - z(1+u))^{3/2}}, \end{aligned} \tag{270}$$

where B_2 is the second derivative of P_2 with respect to v, as introduced in the previous section. This asymptotic expansion enables one to look at the behaviour of the coefficients in the above derivative.

4.83 Lemma *The coefficients of u^p in $\frac{\partial^2 S_2(z,u,v)}{\partial v^2}\Big|_{v=1}$ have the form*

$$[u^p]\frac{\partial^2 S_2(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{p-1}{5 \cdot 2^{2p} d_2}\binom{2p}{p}\left(1 - \frac{25z}{4}\right)^{-p-1/2}, \quad z \to \frac{4}{25}. \tag{271}$$

Next, the coefficients of z^n will be computed using Proposition 1.12

$$[z^n u^p]\frac{\partial^2 S_2(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{p-1}{5 \cdot 2^{2p} d_2 q_2^n}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{272}$$

and their normalisation will give the second moment. Thus the variance of the Steiner distance is

$$V_{n,p}^{(2)} \sim \frac{p-1}{5p_2}\left(\frac{1}{\sqrt{\pi}d_2} - \frac{p-1}{5(2p-1)^2 2^{4p-2} p_2}\binom{2p}{p}^2\right) n\,. \tag{273}$$

Our focus shifts to the general case now. To compute the expectation, we differentiate S_k with respect to v and evaluate at 1. Then we substitute in the known asymptotic expressions and get

$$\begin{aligned}
\left.\frac{\partial S_k(z,u,v)}{\partial v}\right|_{v=1} &= \frac{A_k(z,u)\Big(1-\frac{z}{(1-y_k(z))^2}\Big) - \frac{z\bar{y}_k(z)}{(1-y_k(z))^2} + \frac{zy_k(z)}{(1-y_k(z))^2}}{1-\frac{z}{(1-y_k(z))^2}} \\
&= \frac{A_k(z,u)\Big(1-\frac{(y_k(z)-y_{k-1}(z))}{1-y_k(z)}\Big) - \frac{\bar{y}_k(z)(y_k(z)-y_{k-1}(z))}{1-y_k(z)} + \frac{y_k(z)(y_k(z)-y_{k-1}(z))}{1-y_k(z)}}{1-\frac{(y_k(z)-y_{k-1}(z))}{1-y_k(z)}} \\
&\sim \frac{\sqrt{q_k}\sqrt{q_k-z}}{\sqrt{q_k-z(1+u)}} - \frac{zu\sqrt{q_k}}{2\sqrt{q_k-z}\sqrt{q_k-z(1+u)}}\,.
\end{aligned} \tag{274}$$

This simplified expansion helps us with finding the asymptotic behaviour of the coefficients in the first derivative of S_k.

4.84 Lemma *The coefficients of u^p in $\left.\frac{\partial S_k(z,u,v)}{\partial v}\right|_{v=1}$ have the form*

$$[u^p]\left.\frac{\partial S_k(z,u,v)}{\partial v}\right|_{v=1} \sim \frac{q_k^{1/2}(p-1)}{(2p-1)2^{2p}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p},\ z\to q_k\,. \tag{275}$$

Proof: We compute the coefficients of u^p in (274). The first term gives the following

$$\begin{aligned}
[u^p]\frac{\sqrt{q_k}\sqrt{q_k-z}}{\sqrt{q_k-z(1+u)}} &= q_k^{1/2}\left(1-\frac{z}{q_k}\right)^{1/2}[u^p]\frac{1}{\left(1-\frac{z}{q_k}\right)^{1/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{1/2}} \\
&= \frac{z^p}{2^{2p}q_k^{p-1/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p} \sim \frac{q_k^{1/2}}{2^{2p}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p},\ z\to q_k\,.
\end{aligned} \tag{276}$$

Finally, for the second term in (274) we find

$$-[u^p]\frac{zu\sqrt{q_k}}{2\sqrt{q_k-z}\sqrt{q_k-z(1+u)}} \sim -\frac{pq_k^{1/2}}{2^{2p}(2p-1)}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p},\ z\to q_k\,. \tag{277}$$

The result of the lemma is obtained by adding these two coefficients. ■

Now it is easy to find the other coefficients:

$$[z^n u^p]\frac{\partial S_k(z,u,v)}{\partial v}\Big|_{v=1} \sim \frac{q_k^{1/2}\,(p-1)}{(2p-1)\,2^{2p}}\binom{2p}{p} q_k^{-n}\frac{n^{p-1}}{\Gamma(p)}\,, \tag{278}$$

and they are used to find the first moment.

4.85 Theorem [25] *The expectation for the Steiner distance in monotonic ordered trees is*

$$E_{n,p}^{(k)} \sim \frac{q_k^{1/2}p(p-1)}{p_k(2p-1)2^{2p}}\binom{2p}{p}\sqrt{n}\,,\ n\to\infty\,,\ \textit{fixed } k. \tag{279}$$

Finally, the variance of the Steiner distance can be analysed. The second derivative of S_k with respect to v (evaluated at $v=1$) yields

$$\begin{aligned}\frac{\partial^2 S_k(z,u,v)}{\partial v^2}\Big|_{v=1} &= \frac{B_k(z,u)\left(1-\frac{(y_k(z)-y_{k-1}(z))}{1-y_k(z)}\right)-\frac{2A_k(z,u)(y_k(z)-y_{k-1}(z))}{1-y_k(z)}}{1-\frac{(y_k(z)-y_{k-1}(z))}{1-y_k(z)}}\\ &\sim \frac{3zuq_k}{4d_k(q_k-z(1+u))^{3/2}}-\frac{q_k(q_k-z)}{2d_k(q_k-z(1+u))^{3/2}}\,.\end{aligned} \tag{280}$$

This asymptotic expression is used to find the behaviour of the coefficients of u^p and z^n in the above derivative of S_k.

4.86 Lemma *The coefficients of u^p in $\frac{\partial^2 S_k(z,u,v)}{\partial v^2}\Big|_{v=1}$ are*

$$[u^p]\frac{\partial^2 S_k(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{q_k^{1/2}\,(p-1)}{d_k\,2^{2p+1}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2},\ z\to q_k\,. \tag{281}$$

Proof: One proceeds by computing the coefficients of u^p in (280). For the first term we find

$$\begin{aligned}&[u^p]\frac{3zuq_k}{4d_k(q_k-z(1+u))^{3/2}} = \frac{3z}{4d_k q_k^{1/2}}[u^{p-1}]\frac{1}{\left(1-\frac{z}{q_k}\right)^{3/2}\left(1-\frac{\frac{zu}{q_k}}{1-\frac{z}{q_k}}\right)^{3/2}}\\ &= \frac{3pz^p}{2^{2p+1}q_k^{p-1/2}}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2} \sim \frac{3pq_k^{1/2}}{2^{2p+1}d_k}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2},\ z\to q_k\,.\end{aligned} \tag{282}$$

For the second term in (280) we obtain

$$-[u^p]\frac{q_k(q_k-z)}{2d_k(q_k-z(1+u))^{3/2}} \sim -\frac{(2p+1)q_k^{1/2}}{2^{2p+1}d_k}\binom{2p}{p}\left(1-\frac{z}{q_k}\right)^{-p-1/2},\ z\to q_k\,. \tag{283}$$

By adding these two coefficients we get the result of the lemma. ■

Then one finds that the coefficients of z^n are:

$$[z^n u^p]\frac{\partial^2 S_k(z,u,v)}{\partial v^2}\Big|_{v=1} \sim \frac{q_k^{1/2-n}(p-1)}{d_k 2^{2p+1}}\binom{2p}{p}\frac{n^{p-1/2}}{\Gamma(p+\frac{1}{2})}, \tag{284}$$

which will be normalised to produce the second moment.

4.87 Theorem [27] *The variance for the Steiner distance in monotonic ordered trees is*

$$V_{n,p}^{(k)} \sim \frac{(p-1)q_k^{1/2}}{p_k}\Big(\frac{1}{2\sqrt{\pi}d_k} - \frac{p^2(p-1)q_k^{1/2}}{2^{4p}(2p-1)^2 p_k}\binom{2p}{p}^2\Big)n + \mathcal{O}(\sqrt{n}\,),\ n\to\infty,\ \textit{fixed } k, \tag{285}$$

where $\frac{1}{2\sqrt{\pi}d_k} - \frac{p^2(p-1)q_k^{1/2}}{2^{4p}(2p-1)^2 p_k}\binom{2p}{p}^2 \to 0$ *as* $p\to\infty$.

4.6 Monotonic non-crossing trees

4.6.1 Introduction

This section is based on non-crossing geometric configurations built on vertices of a convex polygon in the plane as described in [8]. Let $P_n = \{v_1, v_2, \dots, v_n\}$ be a set of points conventionally ordered counter-clockwise that are vertices of a convex n-gon. A *non-crossing graph* is defined as a graph with vertex set P_n whose edges are straight line segments that do not cross. We recall that a graph is connected if any two vertices can be joined by a path. A tree is a connected acyclic graph and the number of edges in a tree is one less than the number of vertices.

4.88 Definition [8] Let d be the degree of a vertex v_1 in a *non-crossing* tree t. Then t is a sequence attached to v_1 of d ordered pairs of trees sharing a common vertex. Moreover, a *butterfly* is an ordered pair of non-crossing trees with a common vertex.

The name aims to convey the idea that the pair of trees looks like the two wings of a butterfly. If v_1 has degree d, then the non-crossing tree t can be identified with a sequence of d butterflies pending from v_1. Or more simply: to the root vertex we attach an ordered collection of vertices, each of which has an end-node that is the common root of two non-crossing trees, one on the left of the edge and the other on the right of the edge.

4.89 Theorem [8] *The number of non-crossing trees with n vertices, T_n, is*

$$T_n = \frac{1}{2n-1}\binom{3n-3}{n-1}. \tag{286}$$

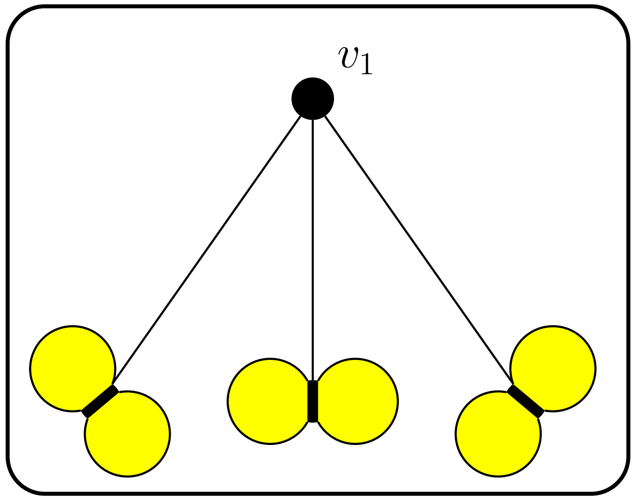


Figure 4.5: Butterflies pending from vertex v_1

If $T(z)$ and $B(z)$ denote the generating functions for non-crossing trees and butterflies respectively, then the following equations hold

$$T(z) = \frac{z}{1 - B(z)}, \quad B(z) = \frac{T^2(z)}{z}, \tag{287}$$

where the division by z in the second equation occurs because one identifies two root vertices to form a butterfly.

Bijections between ternary trees and non-crossing trees have been constructed in [31]. It turns out that the number T_n of non-crossing trees of size n satisfies

$$T_{n+1} = t_n \quad \text{with} \quad t_n = \frac{1}{2n+1}\binom{3n}{n}, \tag{288}$$

where t_n enumerates the ternary trees of size n. Thus a useful relationship between these classes of trees can be established.

4.90 Theorem [31] *The number of ternary trees of size n, t_n, satisfies the recurrence*

$$3(3n-1)(3n-2)t_{n-1} = 2n(2n+1)t_n\,. \tag{289}$$

In this section we are concerned with the analysis of the monotonically labelled non-crossing tree model.

4.91 Theorem *The generating functions for monotonic non-crossing trees is*

$$y_k(z) = y_{k-1}(z) + \frac{z}{1 - \frac{y_k^2(z)}{z}}, \quad y_0(z) = 0\,. \tag{290}$$

4.6.2 Singularities

We have seen in previous sections that tree generating functions such as the one for non-crossing trees above are best dealt with using singularity analysis. One needs to manipulate (290) in order to get it into a more convenient form. First, we use the transformation $z = Z^2$

$$y_k(Z^2) = y_{k-1}(Z^2) + \frac{Z^2}{1 - \frac{y_k^2(Z^2)}{Z^2}}, \quad y_0(z) = 0\,, \tag{291}$$

and then put $T_k(Z) = \frac{y_k(Z^2)}{Z}$ which gives

$$ZT_k(Z) = ZT_{k-1}(Z) + \frac{Z^2}{1 - T_k^2(Z)}\,, \tag{292}$$

or

$$T_k(Z) = T_{k-1}(Z) + \frac{Z}{1 - T_k^2(Z)}\,. \tag{293}$$

Upon differentiating with respect to $T_k(Z)$ we obtain

$$1 = \frac{2\tau_k \rho_k}{(1 - \tau_k^2)^2} \tag{294}$$

for $Z \to \rho_k$ and $T_k(Z) \to \tau_k$, so we have an expression for ρ_k in terms of τ_k

$$\rho_k = \frac{(1 - \tau_k^2)^2}{2\tau_k}\,. \tag{295}$$

Then (293) becomes

$$T_{j-1}(\rho_k) = T_j(\rho_k) - \frac{(1 - \tau_k^2)^2}{2\tau_k} \frac{1}{1 - T_j^2(\rho_k)}\,, \tag{296}$$

moreover

$$T_{k-l-1}(\rho_k) = T_{k-l}(\rho_k) - \frac{(1 - \tau_k^2)^2}{2\tau_k} \frac{1}{1 - T_{k-l}^2(\rho_k)}\,. \tag{297}$$

For convenience, we let $r_l = T_{k-l}(\rho_k)$.

$$r_{l+1} = r_l - \frac{(1 - \tau_k^2)^2}{2\tau_k} \frac{1}{1 - r_l^2} \tag{298}$$

We know that $r_0 = \tau_k$ and $r_k = 0$. So we must solve equation (298) which produces the τ_k. The solutions which tend to one are relevant, since $\rho_k \to 0$. We consider τ_k to be the special values of t which solve the equation $r_k(t) = 0$ for $t \sim 1$. Then we must solve the recursion

$$r_{l+1} = r_l - \frac{(1-t^2)^2}{2t}\frac{1}{1-r_l^2}\,. \tag{299}$$

Since $r_0 = t$, the first term in (299) is

$$r_1 = t - \frac{(1-t^2)}{2t} \tag{300}$$

and it has the following series expansion around $t = 1$

$$r_1 = 1 - 2(1-t) + \mathcal{O}((1-t)^2)\,. \tag{301}$$

We substitute this expression for r_1 into the recursion for r_2

$$\begin{aligned} r_2 &= r_1 - \frac{(1-t^2)^2}{2t}\frac{1}{(1-r_1^2)} \\ &\sim 1 - 2(1-t) - \frac{(1-t^2)^2}{2t}\frac{1}{(1-(1-2(1-t))^2)}\,, \end{aligned} \tag{302}$$

and the series expansion of r_2 around $t = 1$ is

$$r_2 = 1 - \frac{5}{2}(1-t) + \mathcal{O}((1-t)^2)\,. \tag{303}$$

Similarly,

$$r_3 = 1 - \frac{29}{10}(1-t) + \mathcal{O}((1-t)^2)\,. \tag{304}$$

Proceeding in this way we find the series expansion around $t = 1$ for the k-th term in the recursion. First, we let $r_{k-1} = 1 - B_{k-1}(1-t) + \mathcal{O}((1-t)^2)$. Then

$$\begin{aligned} r_k &= r_{k-1} - \frac{(1-t^2)^2}{2t}\frac{1}{1-r_{k-1}^2} \\ &\sim 1 - B_{k-1}(1-t) - \frac{(1-t^2)^2}{2t}\frac{1}{1-(1-B_{k-1}(1-t))^2} \end{aligned} \tag{305}$$

and r_k has the following series expansion around $t = 1$

$$\begin{aligned} r_k &= 1 - \left(B_{k-1} + \frac{1}{B_{k-1}}\right)(1-t) + \mathcal{O}((1-t)^2) \\ &= 1 - B_k(1-t) + \mathcal{O}((1-t)^2)\,. \end{aligned} \tag{306}$$

Thus we have obtained an asymptotic recursion for the r_k's

$$r_k \sim 1 - B_k(1-t) \tag{307}$$

and a recursion for the B_k's

$$B_k = B_{k-1} + \frac{1}{B_{k-1}}, \quad B_0 = 1\,. \tag{308}$$

The solution to (308), which can be handled similarly to De Bruijn, [3], is

$$B_k = k + \log k + \mathcal{O}\left(\frac{\log k}{k}\right)\,. \tag{309}$$

Next, we solve the equation $r_k(t) = 0$, that is $1 - B_k(1-t) = 0$, and we get

$$B_k \sim \frac{1}{1-\tau_k}\,, \quad \tau_k \sim 1 - \frac{1}{B_k}\,. \tag{310}$$

Since τ_k is "known" approximately for $k \to \infty$, we have an expression for ρ_k

$$\rho_k \sim \frac{(1-(1-\frac{1}{B_k})^2)^2}{2(1-\frac{1}{B_k})} = \frac{(2B_k-1)^2}{2B_k^3(B_k-1)} \sim \frac{2}{k^2}\,, \quad k \to \infty\,. \tag{311}$$

The local expansion of T_k about ρ_k starts with

$$T_k(Z) \sim \lambda_k - \mu_k\sqrt{\rho_k - Z}\,, \quad Z \to \rho_k \tag{312}$$

and from $T_k(Z) = \frac{y_k(Z^2)}{Z}$ we obtain

$$\begin{aligned} y_k(Z^2) &\sim \lambda_k Z - \mu_k Z\sqrt{\rho_k - Z} \\ &\sim \lambda_k\rho_k - \mu_k\rho_k\sqrt{\rho_k - Z}\,. \end{aligned} \tag{313}$$

Now we need to go back to z in the above local expansion. We observe that

$$\rho_k^2 - Z^2 = (\rho_k - Z)(\rho_k + Z) \sim 2\rho_k(\rho_k - Z)\,, \quad Z \to \rho_k\,. \tag{314}$$

Thus $\rho_k - Z \sim \frac{1}{2\rho_k}(\rho_k^2 - Z^2)$ and since $z = Z^2$ it follows that $Z \to \rho_k$ and $z \to \sqrt{\rho_k}$, so $\rho_k - \sqrt{z} \sim \frac{1}{2\rho_k}(\rho_k^2 - z)$. The local expansion for y_k becomes

$$y_k(z) \sim \lambda_k\rho_k - \mu_k\frac{\rho_k}{\sqrt{2\rho_k}}\sqrt{\rho_k^2 - z}\,, \quad z \to \rho_k^2\,. \tag{315}$$

The constants λ_k and ρ_k are known for $k \to \infty$ so it remains to produce an asymptotic form for the μ_k's.

4.92 Lemma *The constants μ_k have the asymptotic form*

$$\mu_k \sim \frac{\sqrt{2}}{2}\, k^{1/2}\,,\ \ k \to \infty\,. \tag{316}$$

Proof: From Harary and Palmer, [17] and the Preliminaries chapter, we know that the constants μ_k can be computed by means of the equation

$$\frac{1}{2}\left(\mu_k \frac{\rho_k}{\sqrt{2\rho_k}}\right)^2 = \lim_{z\to\rho_k^2} y_k'(z)(\lambda_k\rho_k - y_k(z))\,. \tag{317}$$

We differentiate (290) with respect to z

$$y_k'(z) = y_{k-1}'(z) + \frac{1}{1-\frac{y_k^2(z)}{z}} - \frac{z\left(\frac{-2y_k(z)y_k'(z)}{z} + \frac{y_k^2(z)}{z^2}\right)}{\left(1-\frac{y_k^2(z)}{z}\right)^2}\,, \tag{318}$$

which simplifies to

$$\left(1 - \frac{2y_k(z)}{\left(1-\frac{y_k^2(z)}{z}\right)^2}\right)y_k'(z) = y_{k-1}'(z) + \frac{1}{1-\frac{y_k^2(z)}{z}} - \frac{y_k^2(z)}{z\left(1-\frac{y_k^2(z)}{z}\right)^2}\,. \tag{319}$$

This first order recursion can be solved as follows. We multiply (319) with the product $\prod_{i=0}^{k-1}\left(1 - \frac{2y_i(z)}{\left(1-\frac{y_i^2(z)}{z}\right)^2}\right)$. Then, a more convenient notation is introduced

$$a_k(z) = \prod_{i=0}^{k-1}\left(1 - \frac{2y_i(z)}{\left(1-\frac{y_i^2(z)}{z}\right)^2}\right)\left(1 - \frac{2y_k(z)}{\left(1-\frac{y_k^2(z)}{z}\right)^2}\right)y_k'(z)\,, \tag{320}$$

which gives the new recursion

$$a_k(z) = a_{k-1}(z) + \prod_{i=0}^{k-1}\left(1 - \frac{2y_i(z)}{\left(1-\frac{y_i^2(z)}{z}\right)^2}\right)\left(\frac{1}{1-\frac{y_k^2(z)}{z}} - \frac{y_k^2(z)}{z\left(1-\frac{y_k^2(z)}{z}\right)^2}\right). \tag{321}$$

By iterating the above k times one gets

$$a_k(z) = a_0(z) + \sum_{j=0}^{k}\left[\prod_{i=0}^{j}\left(1 - \frac{2y_i(z)}{\left(1-\frac{y_i^2(z)}{z}\right)^2}\right)\right]\left(\frac{1}{1-\frac{y_j^2(z)}{z}} - \frac{y_j^2(z)}{z\left(1-\frac{y_j^2(z)}{z}\right)^2}\right), \tag{322}$$

and, after dividing through by $\prod_{i=0}^{k-1}\Big(1-\frac{2y_i(z)}{\big(1-\frac{y_i^2(z)}{z}\big)^2}\Big)$, the solution to (319) is obtained

$$\begin{aligned} y_k'(z) &= \sum_{j=0}^{k}\Big[\prod_{i=j+1}^{k}\Big(1-\frac{2y_i(z)}{(1-\frac{y_i^2(z)}{z})^2}\Big)\Big]^{-1}\Big(\frac{1}{1-\frac{y_j^2(z)}{z}}-\frac{y_j^2(z)}{z\big(1-\frac{y_j^2(z)}{z}\big)^2}\Big) \\ &= \sum_{j=0}^{k}\Big[\prod_{i=j+1}^{k}\Big(1-\frac{2y_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)\Big]^{-1}\Big(\frac{y_j(z)-y_{j-1}(z)}{z} \\ &\qquad -\Big(1-\frac{z}{y_j(z)-y_{j-1}(z)}\Big)\Big(\frac{y_j(z)-y_{j-1}(z)}{z}\Big)^2\Big) \\ &= \sum_{j=0}^{k}\Big[\prod_{i=j+1}^{k}\Big(1-\frac{2y_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)\Big]^{-1}\Big(\frac{(y_j(z)-y_{j-1}(z))(2z-y_j(z)+y_{j-1}(z))}{z^2}\Big), \end{aligned} \tag{323}$$

where $1-\frac{y_k^2(z)}{z}=\frac{z}{y_k(z)-y_{k-1}(z)}$. The dominant term here occurs when $i=j=k$ and since the singularities of $y_{k-1}(z)$ are further away from the origin than those of $y_k(z)$ they will not contribute to the dominant term in the first derivative of y_k which is asymptotically

$$y_k'(z)\sim\frac{y_k(z)(2z-y_k(z))}{z^2-2y_k^3(z)}\,. \tag{324}$$

As $z\to\rho_k^2$ we have $y_k(z)\sim\lambda_k\rho_k$ and $y_k'(z)\to\infty$, so the limit in (317) is of indeterminate form $\infty\cdot 0$. Then (317) is written to make use of l'Hôpital's rule

$$\frac{1}{2}\Big(\mu_k\frac{\rho_k}{\sqrt{2\rho_k}}\Big)^2=\lim_{z\to\rho_k^2}\frac{\lambda_k\rho_k-y_k(z)}{\frac{1}{y_k'(z)}}=\lim_{z\to\rho_k^2}\frac{-y_k'(z)}{\frac{-y_k''(z)}{(y_k'(z))^2}}=\lim_{z\to\rho_k^2}\frac{(y_k'(z))^3}{y_k''(z)}\,. \tag{325}$$

Now (290) is differentiated twice with respect to z

$$\begin{aligned} y_k''(z) &= y_{k-1}''(z)+(2(y_k'(z))^2+2y_k(z)y_k''(z))\Big(1-\frac{y_k^2(z)}{z}\Big)^{-2} \\ &\quad +2z\Big(-\frac{2y_k(z)y_k'(z)}{z}+\frac{y_k^2(z)}{z^2}\Big)^2\Big(1-\frac{y_k^2(z)}{z}\Big)^{-3}, \end{aligned} \tag{326}$$

and simplification yields

$$\begin{aligned} \Big(1-\frac{2y_k(z)}{(1-\frac{y_k^2(z)}{z})^2}\Big)y_k''(z) &= y_k''(z)+2(y_k'(z))^2\Big(1-\frac{y_k^2(z)}{z}\Big)^{-2} \\ &\quad +2z\Big(-\frac{2y_k(z)y_k'(z)}{z}+\frac{y_k^2(z)}{z^2}\Big)^2\Big(1-\frac{y_k^2(z)}{z}\Big)^{-3}. \end{aligned} \tag{327}$$

In order to solve this recursion, we first multiply (327) with the product $\prod_{i=0}^{k-1}\Big(1-\frac{2y_i(z)}{\big(1-\frac{y_i^2(z)}{z}\big)^2}\Big)$

and rewrite the terms in a simpler notation

$$b_k(z) = \prod_{i=0}^{k-1}\Big(1 - \frac{2y_i(z)}{(1-\frac{y_i^2(z)}{z})^2}\Big)\Big(1 - \frac{2y_k(z)}{(1-\frac{y_k^2(z)}{z})^2}\Big)y_k''(z)\,. \tag{328}$$

The new recursion

$$b_k(z) = b_{k-1}(z) + \prod_{i=0}^{k-1}\Big(1-\frac{2y_i(z)}{(1-\frac{y_i^2(z)}{z})^2}\Big)\Big(\frac{2(y_k'(z))^2}{(1-\frac{y_k^2(z)}{z})^2} + \frac{2z\big(-\frac{2y_k(z)y_k'(z)}{z}+\frac{y_k^2(z)}{z^2}\big)^2}{(1-\frac{y_k^2(z)}{z})^3}\Big), \tag{329}$$

is iterated k times which gives

$$b_k(z) = b_0(z) + \sum_{j=0}^{k}\Big[\prod_{i=0}^{j}\Big(1-\frac{2y_i(z)}{(1-\frac{y_i^2(z)}{z})^2}\Big)\Big]\Big(\frac{2(y_j'(z))^2}{(1-\frac{y_j^2(z)}{z})^2} + \frac{2z\big(-\frac{2y_j(z)y_j'(z)}{z}+\frac{y_j^2(z)}{z^2}\big)^2}{(1-\frac{y_j^2(z)}{z})^3}\Big). \tag{330}$$

Upon division by $\prod_{i=0}^{k-1}\Big(1-\frac{2y_i(z)}{(1-\frac{y_i^2(z)}{z})^2}\Big)$, the solution to (327) arises

$$\begin{aligned} y_k''(z) &= \sum_{j=0}^{k}\Big[\prod_{i=j+1}^{k}\Big(1-\frac{2y_i(z)}{(1-\frac{y_i^2(z)}{z})^2}\Big)\Big]^{-1}\Big(\frac{2(y_j'(z))^2}{(1-\frac{y_j^2(z)}{z})^2} + \frac{2z\big(-\frac{2y_j(z)y_j'(z)}{z}+\frac{y_j^2(z)}{z^2}\big)^2}{(1-\frac{y_j^2(z)}{z})^3}\Big) \\ &= \sum_{j=0}^{k}\Big[\prod_{i=j+1}^{k}\Big(1-\frac{2y_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)\Big]^{-1}\Big(\frac{2(y_j'(z))^2(y_j(z)-y_{j-1}(z))^2}{z^2} \\ &\quad + 2z\Big(-\frac{2y_j(z)y_j'(z)}{z}+\frac{1}{z}\Big(1-\frac{z}{y_j(z)-y_{j-1}(z)}\Big)\Big)^2\Big(\frac{z}{y_j(z)-y_{j-1}(z)}\Big)^{-3} \\ &= \sum_{j=0}^{k}\Big[\prod_{i=j+1}^{k}\Big(1-\frac{2y_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)\Big]^{-1}\Big(\frac{2(y_j'(z))^2(y_j(z)-y_{j-1}(z))^2}{z^2} \\ &\quad + \frac{2(y_j(z)-y_{j-1}(z))\big\{(y_j(z)-y_{j-1}(z))(1-2y_j(z)y_j'(z))-z\big\}^2}{z^4}\Big), \end{aligned} \tag{331}$$

where the substitution $1-\frac{y_k^2(z)}{z} = \frac{z}{y_k(z)-y_{k-1}(z)}$ was used once again. It follows that an asymptotic expansion for this second derivative of y_k can be derived by considering the dominant term in the above and disregarding y_{i-1} and y_{j-1} (since they will not contribute to the main term)

$$\begin{aligned} y_k''(z) &\sim \Big(1-\frac{2y_k^3(z)}{z^2}\Big)^{-1}\Big(\frac{2(y_k'(z))^2y_k^2(z)}{z^2} + \frac{2y_k(z)\big\{y_k(z)(1-2y_k(z)y_k'(z))-z\big\}^2}{z^4}\Big) \\ &= \frac{2y_k(z)\big(y_k^6(z)(2z-y_k(z))^2+z^3(zy_k(z)-z^2-2y_k^3(z))^2\big)}{z^3(z^2-2y_k^3(z))^3}. \end{aligned} \tag{332}$$

At this stage one can use the solutions of $y_k'(z)$ and $y_k''(z)$ to proceed with the computation of μ_k from (325)

$$\begin{aligned}
\mu_k &= \frac{2}{\sqrt{\rho_k}}\sqrt{\frac{(y_k'(\rho_k^2))^3}{y_k''(\rho_k^2)}} \\
&= \frac{2}{\sqrt{\rho_k}}\left(\frac{\left(\sum_{j=0}^{k}\left[\prod_{i=j+1}^{k}\left(1-\frac{2y_i(\rho_k^2)}{\left(1-\frac{y_i^2(\rho_k^2)}{\rho_k^2}\right)^2}\right)\right]^{-1}\left(\frac{\rho_k^2-2y_j^2(\rho_k^2)}{\rho_k^2\left(1-\frac{y_j^2(\rho_k^2)}{\rho_k^2}\right)^2}\right)\right)^3}{\sum_{j=0}^{k}\left[\prod_{i=j+1}^{k}\left(1-\frac{2y_i(\rho_k^2)}{\left(1-\frac{y_i^2(\rho_k^2)}{\rho_k^2}\right)^2}\right)\right]^{-1}\left(\frac{2(y_j'(\rho_k^2))^2}{\left(1-\frac{y_j^2(\rho_k^2)}{\rho_k^2}\right)^2}+\frac{2\rho_k^2\left(-\frac{2y_j(\rho_k^2)y_j'(\rho_k^2)}{\rho_k^2}+\frac{y_j^2(\rho_k^2)}{\rho_k^4}\right)^2}{\left(1-\frac{y_j^2(\rho_k^2)}{\rho_k^2}\right)^3}\right)}\right)^{1/2} \\
&\sim \frac{2}{\sqrt{\rho_k}}\left(\frac{\frac{y_k^3(\rho_k^2)(2\rho_k^2-y_k(\rho_k^2))^3}{(\rho_k^4-2y_k^3(\rho_k^2))^3}}{\frac{2y_k(\rho_k^2)\left(y_k^6(\rho_k^2)(2\rho_k^2-y_k(\rho_k^2))^2+\rho_k^6(\rho_k^2y_k(\rho_k^2)-\rho_k^4-2y_k^3(\rho_k^2))^2\right)}{\rho_k^6(\rho_k^4-2y_k^3(\rho_k^2))^3}}\right)^{1/2} \\
&= \frac{2}{\sqrt{\rho_k}}\left(\frac{\rho_k^6y_k^2(\rho_k^2)(2\rho_k^2-y_k(\rho_k^2))^3}{2\left(y_k^6(\rho_k^2)(2\rho_k^2-y_k(\rho_k^2))^2+\rho_k^6(\rho_k^2y_k(\rho_k^2)-\rho_k^4-2y_k^3(\rho_k^2))^2\right)}\right)^{1/2}.
\end{aligned} \tag{333}$$

But as $z \to \rho_k^2$ it was established that $y_k(\rho_k^2) \sim \lambda_k\rho_k$ and $\lambda_k \sim 1-\frac{1}{k}$ as well as $\rho_k \sim \frac{2}{k^2}$ for $k \to \infty$. Therefore, we obtain the following

$$\mu_k \sim \frac{\sqrt{2}}{2}\,k^{1/2}\,,\ k \to \infty\,. \tag{334}$$

■

In their paper [14], Gittenberger and Panholzer have produced general results regarding the singularities of monotonically labelled simply generated trees. One can also apply those results to our analysis in order to describe the limiting distributions for the parameters of interest.

Let a sequence of non-negative numbers $(\varphi_k)_{k\geq 0}$ define the weight $w(T)$ of any non-crossing tree, T by

$$w(T) = \prod_{v}\varphi_{d(v)}\,, \tag{335}$$

where v ranges over all vertices of T and $d(v)$ is the out-degree of v. Moreover, let

$$\varphi(t) = \sum_{k\geq 0}\varphi_k t^k\,, \tag{336}$$

be the degree-weight generating function $\varphi(t)$.

4.93 Proposition [14] *The degree-weight generating function $\varphi(t)$ satisfies the following:*
(i) $\varphi(t)$ is aperiodic, that is $\gcd\{k : \varphi_k > 0\} = 1$,
(ii) $\varphi(t)$ has a positive radius of convergence $R > 0$,

(iii) for all $k \geq 1$ there exists a minimal positive solution $\tau_k < R$ of the equation

$$t = \frac{\varphi(t)}{\varphi'(t)} + T_{k-1}\Big(\frac{1}{\varphi'(t)}\Big), \tag{337}$$

From the above result and the methods from [6] and [24] it follows that the dominant singularity ρ_k of $T_k(z)$ is given by

$$\rho_k = \frac{1}{\varphi'(\tau_k)}, \tag{338}$$

which agrees with the results we obtained in the beginning of this subsection. The local expansion of $T_k(z)$ around the dominant singularity $z = \rho_k$ follows also directly from [6]

$$\begin{aligned} T_k(z) &= g_k(z) - h_k(z)\sqrt{1 - \frac{z}{\rho_k}} \\ &= \tau_k - \sqrt{\frac{2(\varphi(\tau_k) + T'_{k-1}(\rho_k))}{\varphi''(\tau_k)}}\sqrt{1 - \frac{z}{\rho_k}} + \mathcal{O}\Big(1 - \frac{z}{\rho_k}\Big), \end{aligned} \tag{339}$$

where $g_k(z)$ and $h_k(z)$ are analytic functions in a neighbourhood of $z = \rho_k$.

4.6.3 Size of the ancestor tree

We begin the analysis of the first parameter of interest for the monotonic non-crossing trees. As usual one requires statistics such as the expectation and variance. The method followed in the beginning is similar to that used in the previous sections. However, we will apply the general results from [14] to derive the limiting distribution.

4.94 Theorem *The equation for the size of the ancestor tree in monotonic non-crossing trees is*

$$A_k(z,u,v) = A_{k-1}(z,u,v) + \frac{zv(1+u)}{1 - \frac{A_k^2(z,u,v)}{z}} + \frac{z(1-v)}{1 - \frac{y_k^2(z)}{z}}. \tag{340}$$

We differentiate A_k with respect to v, let $v = 1$ and use the substitutions

$$\frac{\partial A_k(z,u,v)}{\partial v}\bigg|_{v=1} = \alpha_k(z,u), \ \ A_k(z,u,1) = y_k(z(1+u)) = \bar{y}_k(z), \tag{341}$$

which give the following recursion

$$\Big(1 - \frac{2(1+u)\bar{y}_k(z)}{(1 - \frac{\bar{y}_k^2(z)}{z})^2}\Big)\alpha_k(z,u) = \alpha_{k-1}(z,u) + \frac{z(1+u)}{1 - \frac{\bar{y}_k^2(z)}{z}} - \frac{z}{1 - \frac{y_k^2(z)}{z}}. \tag{342}$$

Using the substitution $1-\frac{y_k^2(z)}{z}=\frac{z}{y_k(z)-y_{k-1}(z)}$ (respectively for $\bar{y}_k$), the above can be more conveniently written as

$$\Big(1-\frac{2(1+u)\bar{y}_k(z)(y_k(z)-y_{k-1}(z))^2}{z^2}\Big)\alpha_k(z,u)=\alpha_{k-1}(z,u)+(1+u)(\bar{y}_k(z)-\bar{y}_{k-1}(z)) \\ -y_k(z)+y_{k-1}(z)\,. \tag{343}$$

This first order recursion will be easily solved by multiplying through with the product $\prod_{i=1}^{k-1}\Big(1-\frac{2(1+u)\bar{y}_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)$ and then summing over all k, which leads to the following

$$\prod_{i=1}^{k-1}\Big(1-\frac{2(1+u)\bar{y}_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)\Big(1-\frac{2(1+u)\bar{y}_k(z)(y_k(z)-y_{k-1}(z))^2}{z^2}\Big)\alpha_k(z,u) \\ =\sum_{j=0}^{k-1}\Big[\prod_{i=1}^{j}\Big(1-\frac{2(1+u)\bar{y}_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)\Big] \\ \times\Big[(1+u)(\bar{y}_j(z)-\bar{y}_{j-1}(z))-y_j(z)+y_{j-1}(z)\Big]\,. \tag{344}$$

Finally, we divide by $\prod_{i=1}^{k}\Big(1-\frac{2(1+u)\bar{y}_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)$ and obtain

$$\alpha_k(z,u)=\sum_{j=1}^{k}\Big[\prod_{i=j}^{k}\Big(1-\frac{2(1+u)\bar{y}_i(z)(y_i(z)-y_{i-1}(z))^2}{z^2}\Big)\Big]^{-1}\Big[(1+u)(\bar{y}_j(z)-\bar{y}_{j-1}(z)) \\ -y_j(z)+y_{j-1}(z)\Big]\,, \tag{345}$$

since $\alpha_0(z,u)=0$. However, the dominant term occurs for $i=j=k$ and y_{k-1}, $\bar{y}_{k-1}$ will not contribute to this term so it follows that

$$\alpha_k(z,u)\sim\frac{z^2((1+u)\bar{y}_k(z)-y_k(z))}{z^2-2(1+u)\bar{y}_k^3(z)}\,. \tag{346}$$

In order to obtain the expectation from the above, one can substitute in the local expansions for y_k and $\bar{y}_{k-1}$ and compute the coefficients $[z^n u^p]\alpha_k(z,u)$. For computing the second moment, we differentiate A_k twice with respect to v and then evaluate at $v=1$.

After denoting this second derivative by $\beta_k(z,u)$, the following recursion arises

$$\Big(1-\frac{2(1+u)\bar{y}_k(z)}{\big(1-\frac{\bar{y}_k^2(z)}{z}\big)^2}\Big)\beta_k(z,u)=\beta_{k-1}(z,u)+\frac{2(1+u)\alpha_k(z,u)\big(\alpha_k(z,u)+2\bar{y}_k(z)\big)}{\big(1-\frac{\bar{y}_k^2(z)}{z}\big)^2} + \frac{8(1+u)\bar{y}_k^2(z)\alpha_k^2(z,u)}{z\big(1-\frac{\bar{y}_k^2(z)}{z}\big)^3}\,. \tag{347}$$

Its solution will be obtained in a similar manner to the that of the recursion for α_k. Then the dominant term is

$$\beta_k(z,u)\sim\frac{2(1+u)\bar{y}_k^2(z)\alpha_k(z,u)\big[z^2(\alpha_k(z,u)+2\bar{y}_k(z))+4\bar{y}_k^3(z)\alpha_k^2(z,u)\big]}{z^2(z^2-2(1+u)\bar{y}_k^3(z))}\,, \tag{348}$$

which enables one to look at the behaviour of the coefficients $[z^nu^p]\beta_k(z,u)$ around the singularity ρ_k^2 and it ultimately leads to the variance.

Naturally, one is interested in computing all the moments for the size of the ancestor tree and finding a suitable function for them. A very useful technique developed by Panholzer in [30] can be employed: differentiate the generating function for the size of the ancestor tree, $A(z,u,v)$, p-times with respect to u and evaluate at $u=0$. Then the following representation can be used

$$N_uD_u^pA_k(z,u,v)=p![u^p]A_k(z,u,v)\,, \tag{349}$$

where D_u denotes the differential operator with respect to u and N_u is the evaluation operator at $u=0$. The next result was adjusted accordingly for our monotonic non-crossing trees.

4.95 Lemma [14] *For $p\geq 1$, $m=\mathcal{O}(\sqrt{n}\,)$ and $n\to\infty$ we have the following*

$$N_uD_u^pA_k(z,u,v)\sim\frac{(p-1)!}{\varphi''(T_k(z))2^{p-1}}\binom{2(p-1)}{p-1}\big(C_k(z)\varphi''(T_k(z))\big)^p\frac{(zv)^{2p-1}}{(1-zv\varphi'(T_k(z)))^{2p-1}}\,, \tag{350}$$

with

$$C_k(z)=\sum_{l=1}^{k}\frac{(\varphi'(T_k(z)))^{k-l}\varphi(T_l(z))}{\prod_{s=l}^{k-l}\big(\varphi'(T_k(z))-\varphi'(T_s(z))\big)}\,. \tag{351}$$

Extracting the coefficients of v^m from (350) gives

$$[v^m]N_uD_u^pA_k(z,u,v)\sim\frac{m^{2p-2}}{(p-1)!}\frac{(C_k(z)\varphi''(T_k(z)))^p}{\varphi''(T_k(z))2^{p-1}(\varphi'(T_k(z)))^{2p-1}}(z\varphi'(T_k(z)))^m\,. \tag{352}$$

Let $X_{n,p}$ be the random variable which counts the size of the ancestor tree of p randomly selected nodes in monotonic non-crossing trees of size n. We wish to compute the probabilities

$$\mathbb{P}\{X_{n,p}=m\}=\frac{[z^n u^p v^m]A_k(z,u,v)}{\binom{n}{p}T_n}\,, \tag{353}$$

so the coefficients of z^n need to be extracted from (352). This can be done in the same fashion as for simply generated trees in [30] and together with the evaluation

$$C_k(\rho_k)=\varphi(\tau_k)+T'_{k-1}(\rho_k)\,, \tag{354}$$

which follows by induction, one obtains the following result (slightly adjusted to depict the class of monotonically labelled non-crossing trees).

4.96 Lemma [14] *The probability that the size of the ancestor tree of p randomly chosen nodes in a monotonic non-crossing tree of size n is equal to m is asymptotically*

$$\begin{aligned}\mathbb{P}\{X_{n,p}=m\}&=\frac{[z^n v^m]N_u D_p^u A_k(z,u,v)}{p!\binom{n}{p}[z^n]T_k(z)}\\ &\sim\frac{2m^{2p-1}}{n^p(p-1)!}\left(\frac{\sigma_k}{\sqrt{2}}\right)^{2p}e^{-\frac{\sigma_k^2 m^2}{2n}}\,,\end{aligned} \tag{355}$$

where $\sigma_k=\sqrt{\rho_k^2\varphi''(\tau_k)(\varphi(\tau_k)+T'_{k-1}(\rho_k))}$.

By making an appropriate substitution for m, one obtains the following result which characterises the limiting distribution for our parameter.

4.97 Theorem [14] *For $m=x\sqrt{n}+o(\sqrt{n})$ and fixed $p\geq 1$ we have*

$$\begin{aligned}\sqrt{n}\,\mathbb{P}\{X_{n,p}=m\}&\sim\frac{2}{(p-1)!}\left(\frac{\sigma_k}{\sqrt{2}}\right)^{2p}x^{2p-1}e^{-\frac{\sigma_k^2x^2}{2}}\\ &=g(p,2,\frac{\sqrt{2}}{\sigma_k};x)\,,\end{aligned} \tag{356}$$

where $g(b,h,B;x)=\frac{|h|}{\Gamma(b)B}\left(\frac{x}{B}\right)^{bh-1}e^{-(\frac{x}{B})^h}$ is the density function of the generalised Gamma distribution. Thus the limiting distribution of the normalised random variable $\frac{X_{n,p}}{\sqrt{n}}$ is asymptotically for fixed $p\geq 1$ and $n\to\infty$ a generalised Gamma distribution with parameters $(p,2,\frac{\sqrt{2}}{\sigma_k};x)$.

4.6.4 The Steiner distance

We have already seen that the Steiner distance is closely related to the size of the ancestor tree and this shows yet again in the following result.

4.98 Theorem *The Steiner distance for monotonic non-crossing trees is*

$$S_k(z,u,v) = \frac{A_k(z,u,v)\left(1 - \frac{2vy_k(z)}{\left(1-\frac{y_k^2(z)}{z}\right)^2}\right) - \frac{2(1-v)y_k^2(z)}{\left(1-\frac{y_k^2(z)}{z}\right)^2}}{1 - \frac{2y_k(z)}{\left(1-\frac{y_k^2(z)}{z}\right)^2}} . \tag{357}$$

The results from [14] are applied to this parameter as well. Instead of computing $N_u D_u^p S_k(z,u,v)$ one need only find $N_u D_u^p A_l(z,u,v)$ for $1 \le l \le k$ since

$$\begin{aligned} N_u D_u^p S_k(z,u,v) &= \frac{1 - zv\varphi'(T_k(z))}{1 - z\varphi'(T_k(z))} N_u D_u^p A_k(z,u,v) \\ &\quad + \sum_{l=1}^{k-1} \frac{\prod_{s=l}^{k-1} z(1-v)\varphi'(T_s(z))}{\prod_{s=l}^{k} (1 - z\varphi'(T_s(z)))} N_u D_u^p A_l(z,u,v) . \end{aligned} \tag{358}$$

From [14] we have that

$$N_u D_u S_k(z,u,v) \sim \frac{1 - zv\varphi'(T_k(z))}{1 - z\varphi'(T_k(z))} N_u D_u^p A_k(z,u,v) , \tag{359}$$

for $p \ge 2$ fixed, $m = \mathcal{O}(\sqrt{n}\,)$ and $n \to \infty$. Using the formula (350), one can show the following for the random variable $Y_{n,p}$ which counts the Steiner distance of p randomly selected nodes in monotonic non-crossing trees of size n

$$\mathbb{P}\{Y_{n,p} = m\} = \frac{[z^n v^m] N_u D_u^p S_k(z,u,v)}{p!\binom{n}{p}[z^n]T_k(z)} \sim \frac{2m^{2p-3}}{n^{p-1}(p-2)!}\left(\frac{\sigma_k}{\sqrt{2}}\right)^{2(p-1)} e^{-\frac{\sigma_k^2 m^2}{2n}} , \tag{360}$$

where σ_k is as defined in the section for the size of the ancestor tree. Upon setting $m = x\sqrt{n} + o(\sqrt{n}\,)$ we obtain

$$\begin{aligned} \sqrt{n}\,\mathbb{P}\{Y_{n,p} = m\} &\sim \frac{2}{(p-1)!}\left(\frac{\sigma_k}{\sqrt{2}}\right)^{2p} x^{2p-1} e^{-\frac{\sigma_k^2 x^2}{2}} \\ &= g(p-1, 2, \frac{\sqrt{2}}{\sigma_k}; x) , \end{aligned} \tag{361}$$

which means that the limiting distribution of the normalised random variable $\frac{Y_{n,p}}{\sqrt{n}}$ is asymptotically for fixed $p \geq 1$ and $n \to \infty$ a generalised Gamma distribution with parameters $(p-1, 2, \frac{\sqrt{2}}{\sigma_k}; x)$.

Certain classes of monotonically labelled trees (binary, t-ary, ordered) and their properties were introduced by Prodinger and Urbanek in [38]. In this chapter we have used the models proposed in their paper to analyse our usual parameters. In order to produce general results it was useful to consider particular cases first. Although sometimes tedious, this method enabled us to gauge the behaviour of the parameters.

Very recently, Gittenberger and Panholzer, [14], have derived general results for monotonically labelled random trees which we have applied in our study, especially when limiting distributions were determined.

5

Conclusion

There's nothing that keeps its youth,
So far as I know, but a tree and truth.

The Deacon's Masterpiece
Oliver Wendell Holmes

The results presented in this thesis pertain to specific classes of increasing and monotonic trees and they constitute only a very small fraction of the vast domain that tree structures occupy in areas such as analysis of algorithms, combinatorics and theoretical computer science. Trees appear everywhere, from the studies of chemical compounds (for example [34], [42]) to financial, genetic and statistical modelling.

In recent years, many general techniques have been found which make it possible to learn characteristics of complicated new tree structures. Some notable advances are the methods employed for determining limiting distributions for just about any conceivable tree parameter. The development of the "quasi-power theorem", as described in [18], is particularly useful in this respect.

The tools required for studying tree structures rely on many fields of mathematics such as algebra, approximation theory, complex analysis and differential equations. Our approach has been through generating functions and singularity analysis. These enable one to understand the behaviour of several parameters from which we have presented two that describe the distances between nodes, namely the size of the ancestor tree and the Steiner distance. Naturally, there are many other ways of understanding trees, for instance from algebraic, geometric and graph theoretic (see [4]) points of view.

Although tree structures have been the subject of much study there are many important problems to be solved. For example, accurate results about a basic parameter, tree height, have only recently been produced. Distances between nodes and limiting distributions in priority trees (one of the most useful structures for implementing queues) have not appeared as yet.

It is through the analysis of trees that many interesting methods and theories have come to light and as their understanding grows so does their value. Whichever way one comes across these structures, there will always be scope for further investigation. Indeed, Knuth, [22], refers to trees as being

> *the most important nonlinear structures that arise in computer algorithms.*

6

Appendix

> There is, I conceive, scarcely any tree that may not be advantageously used in the various combinations of form and color.
>
> William Gilpin

We present some numerical experiments which deal one of the parameters analysed in the thesis. Consider the generating function for the size of the ancestor tree in monotonic ordered trees. We will illustrate it for small values of the tree size n and number of labels p. If $n = 3$ say, then p can have four values $p = 0, \ldots, 3$. In the figures which follow, we present all the possible cases and their respective contribution to the size of the ancestor tree.

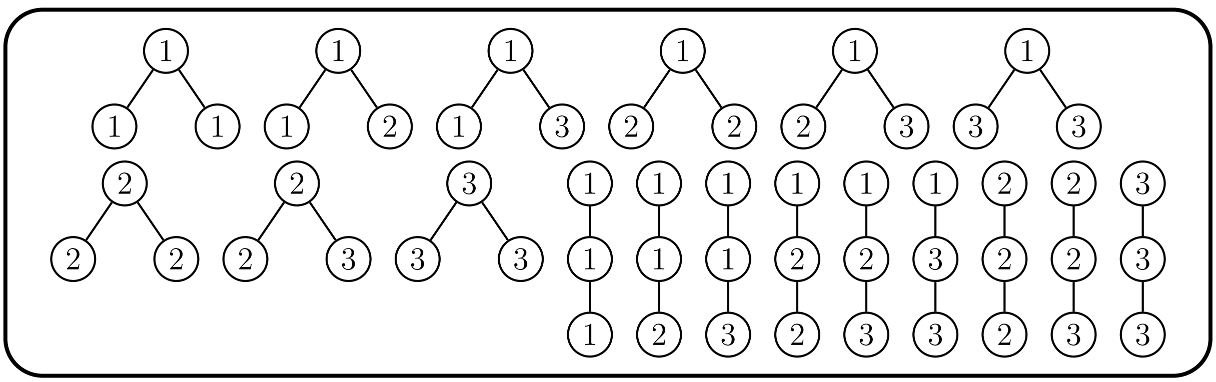


Figure 6.1: All monotonic ordered trees of size $n = 3$; the contribution to the generating function for the size of the ancestor tree is $18z^3$.

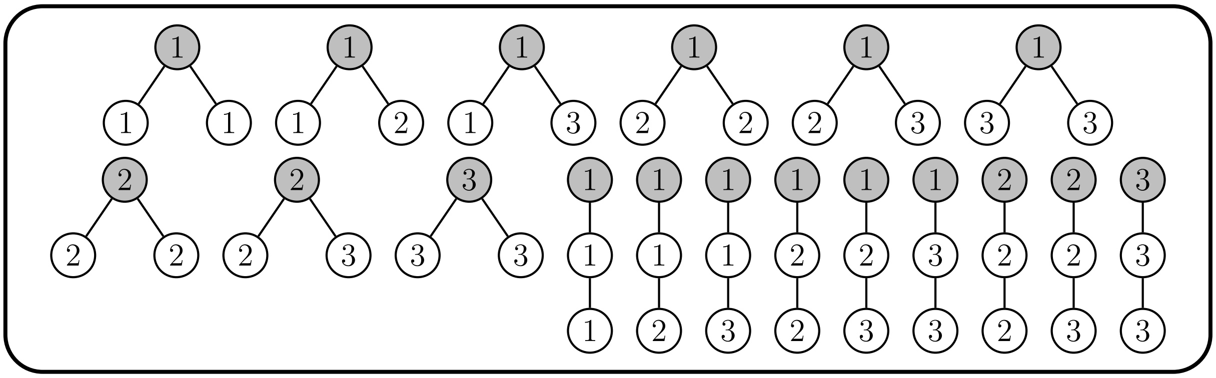


Figure 6.2: All monotonic ordered trees of size 3 with the the root labelled; the size of the ancestor tree is 1 and the contribution to the generating function is $18uvz^3$.

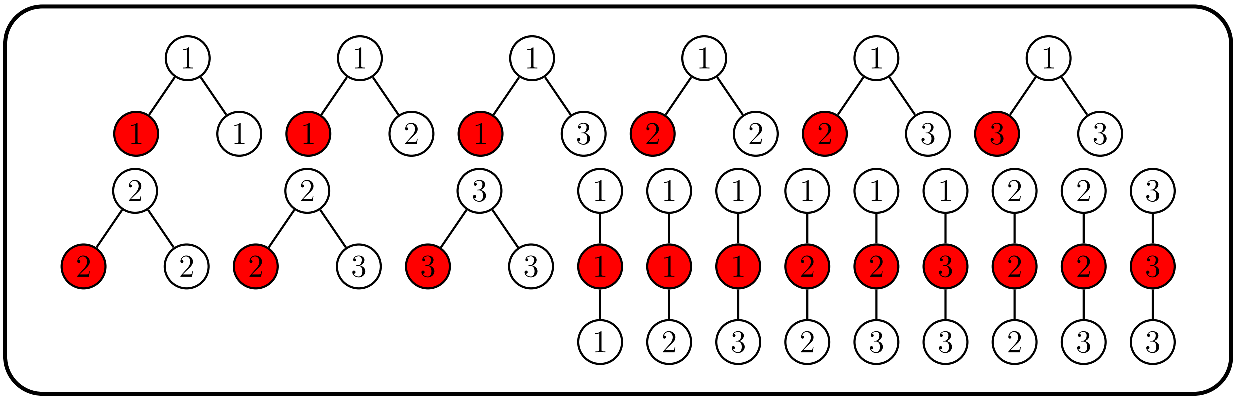


Figure 6.3: All monotonic ordered trees of size 3 with one node, other than the root, labelled; the size of the ancestor tree is 2 and the contribution to the generating function is $18uv^2z^3$.

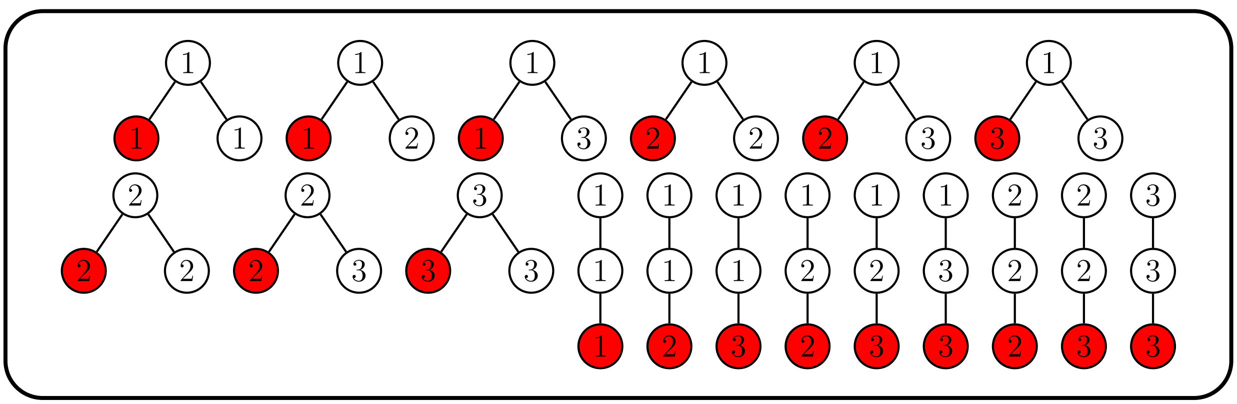


Figure 6.4: All monotonic ordered trees of size $n = 3$ with one node, other than the root, labelled; size of the ancestor tree for the first nine trees is 2 and their contribution to the generating function is $9uv^2z^3$; size of the ancestor tree for the last nine trees is 3 and their contribution is $9u^2v^3z^3$.

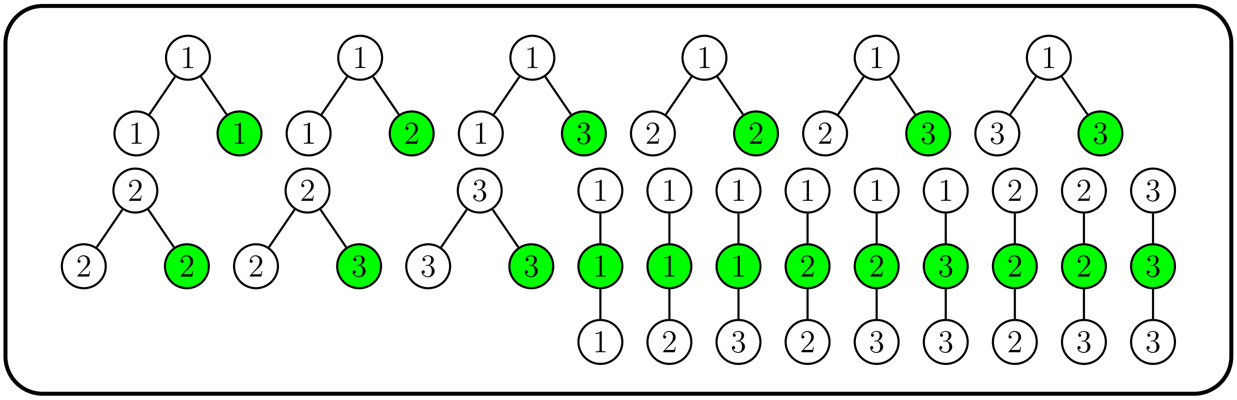


Figure 6.5: All monotonic ordered trees of size 3 with one node, other than the root, labelled; the size of the ancestor tree is 2 and the contribution to the generating function is $18uv^2z^3$.

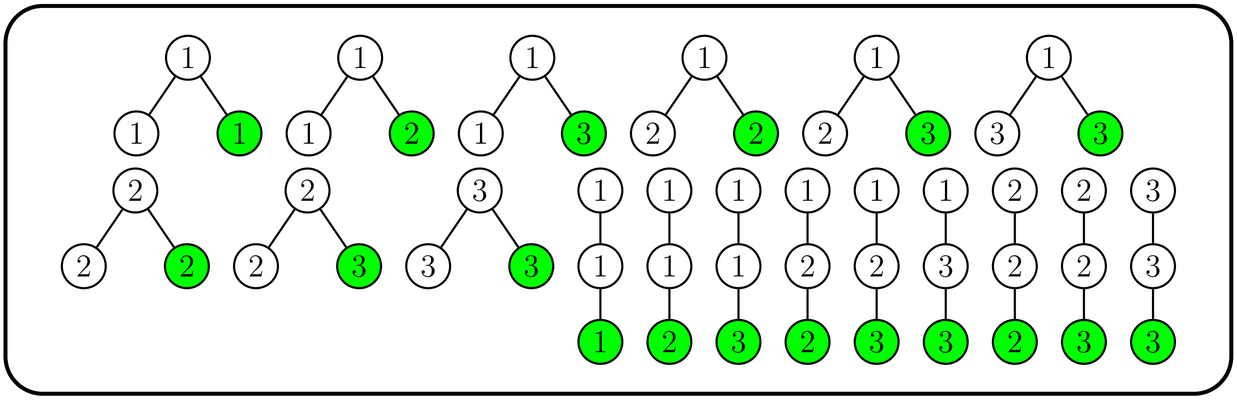


Figure 6.6: All monotonic ordered trees of size 3 with one node, other than the root, labelled; the size of the ancestor tree for the first nine trees is 2 and their contribution is $9uv^2z^3$; the size of the ancestor tree for the last nine trees is 3 and their contribution is $9u^2v^3z^3$.

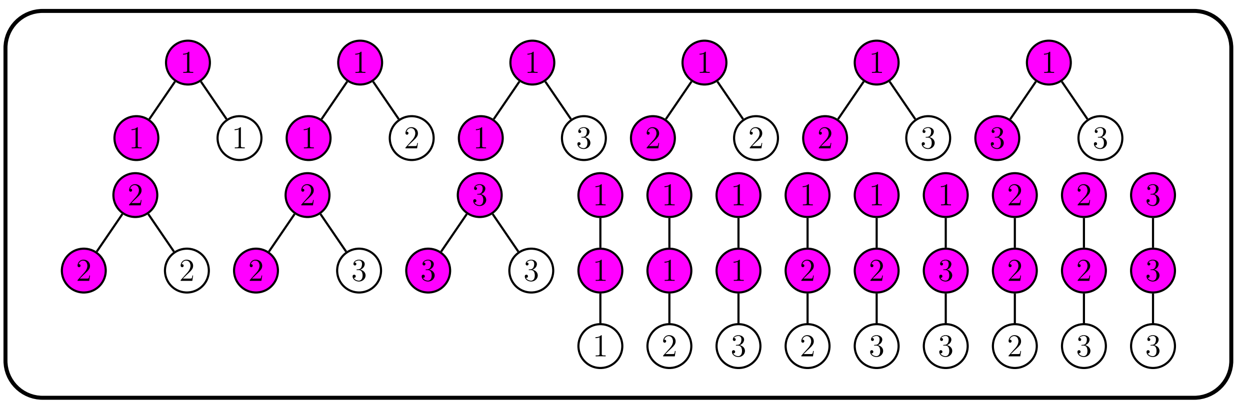


Figure 6.7: All monotonic ordered trees of size 3 with the root and one other node labelled; the size of the ancestor tree is 2 and the contribution to the generating function is $18u^2v^2z^3$.

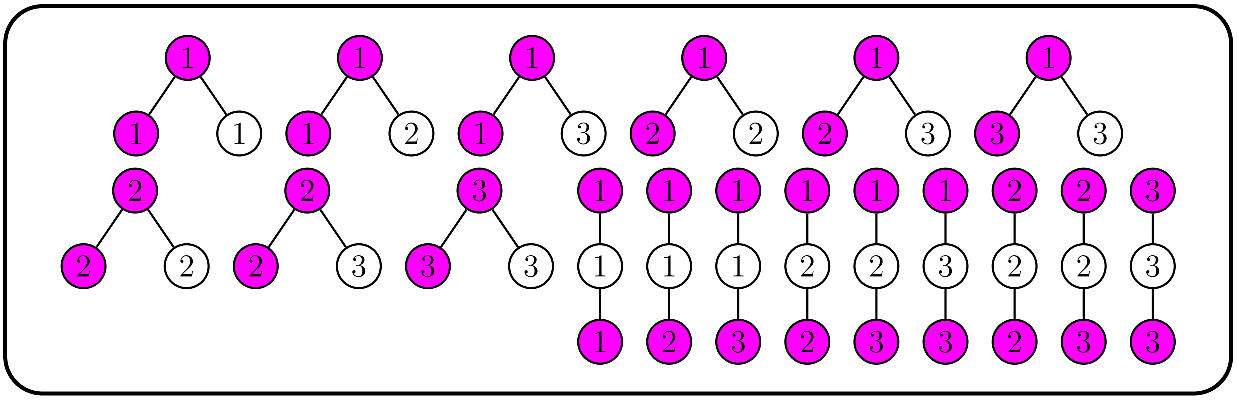


Figure 6.8: All monotonic ordered trees of size $n = 3$ with the root and one other node labelled; the size of the ancestor tree for the first nine trees is 2 and their contribution is $9u^2v^2z^3$; the size of the ancestor tree for the last nine trees is 3 and their contribution is $9u^2v^3z^3$.

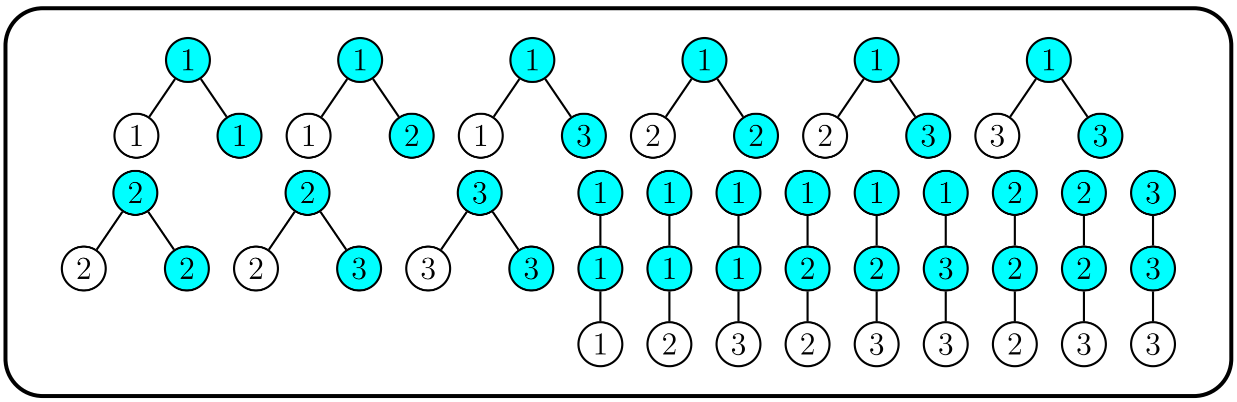


Figure 6.9: All monotonic ordered trees of size 3 with the root and one other node labelled; the size of the ancestor tree is 2 and the contribution to the generating function is $18u^2v^2z^3$.

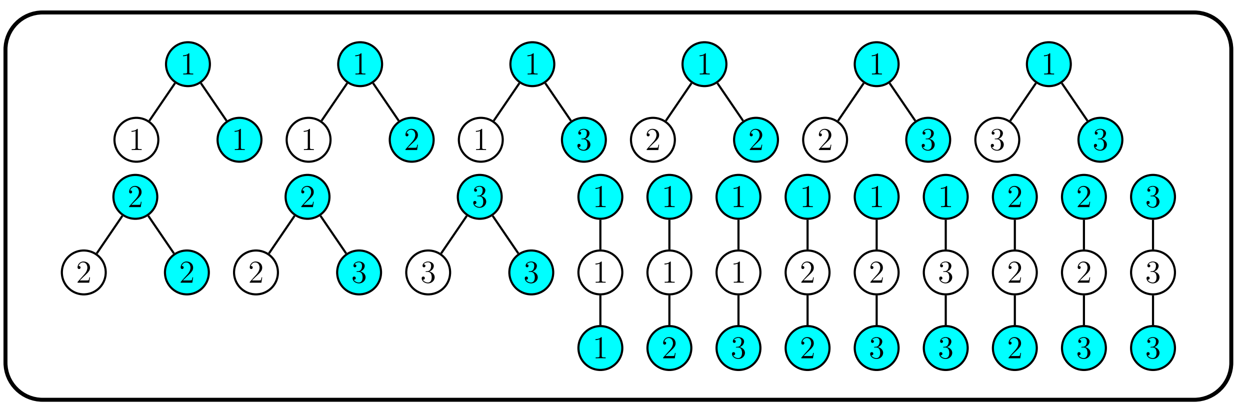


Figure 6.10: All monotonic ordered trees of size $n = 3$ with the root and one other node labelled; the size of the ancestor tree for the first nine trees is 2 and their contribution is $9u^2v^2z^3$; the size of the ancestor tree for the last nine trees is 3 and their contribution is $9u^2v^3z^3$.

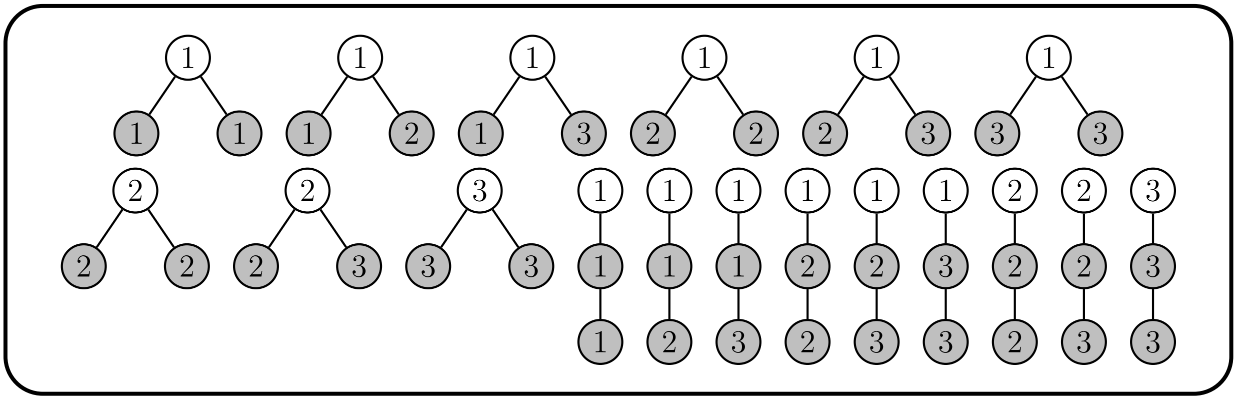


Figure 6.11: All monotonic ordered trees of size 3 with two nodes other than the root labelled; the size of the ancestor tree is 3 and their contribution to the generating function is $18u^2v^3z^3$.

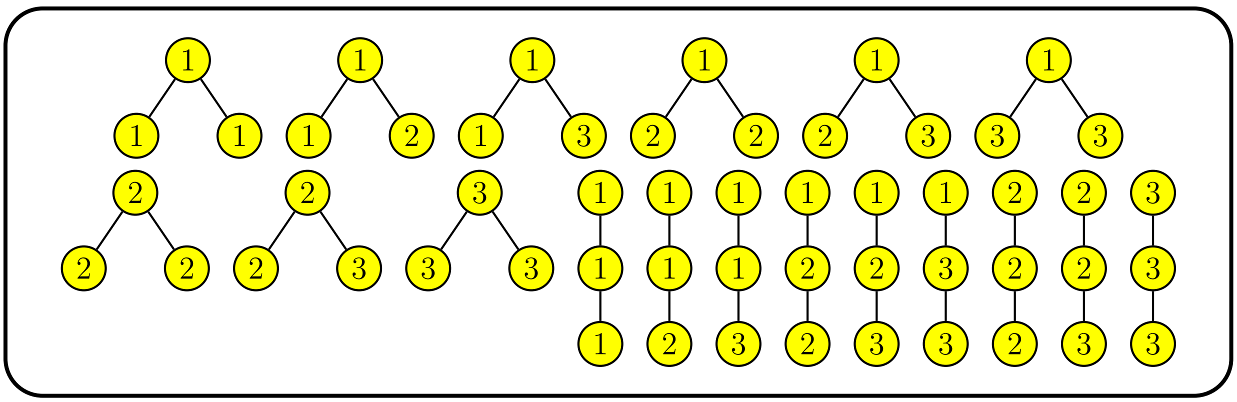


Figure 6.12: All monotonic ordered trees of size 3 with three nodes labelled; the size of the ancestor tree is 3 and their contribution to the generating function is $18u^3v^3z^3$.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, 1972.

[2] F. Bergeron, F. Flajolet, and B. Salvy. Varieties of increasing trees. *Lecture Notes in Computer Science*, 581:24–48, 1992.

[3] N. G. De Bruijn. *Asymptotic Methods in Analysis*. Dover, 1981.

[4] P. Dankelmann, O. R. Oellermann, and H. C. Swart. The average Steiner distance of a graph. *Journal of Graph Theory*, 22:15–22, 1996.

[5] M. Drmota. The height distribution of leaves in rooted trees. *Discrete Mathematics and Applications*, 4:45–58, 1994.

[6] M. Drmota. Combinatorics and asymptotics on trees. *Cubo Journal*, 6:105–136, 2004.

[7] P. Flajolet. Analyse d'algorithmes de manipulation d'arbres et des fichiers. *Thèse Paris-Sud-Orsay*, 1979.

[8] P. Flajolet and M. Noy. Analytic combinatorics of non-crossing configurations. *INRIA Rocquencourt*, Research Report number 3196, 1997.

[9] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM Journal of Discrete Mathematics*, 3:216–240, 1990.

[10] P. Flajolet and R. Sedgewick. The average case analysis of algorithms: complex asymptotics and generating functions. *INRIA Rocquencourt*, Research Report number 2026, 1993.

[11] P. Flajolet and R. Sedgewick. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, 1996.

[12] P. Flajolet and R. Sedgewick. The average case analysis of algorithms: multivariate asymptotics and limit distributions. *INRIA Rocquencourt*, Research Report number 3162, 1997.

[13] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Appears on the following website `http://pauillac.inria.fr/algo/flajolet/Publications/books.html`, 2005.

[14] B. Gittenberger and A. Panholzer. Some results for monotonically labelled simply generated trees. *Discrete Mathematics and Theoretical Computer Science*, 2005.

[15] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, 1994.

[16] D. H. Greene and D. E. Knuth. *Mathematics for the Analysis of Algorithms*. Birkhäuser, 1982.

[17] F. Harary and E. M. Palmer. *Graphical Enumeration*. Academic Press, 1973.

[18] H.-K. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19:329–343, 1998.

[19] S. Janson. The Wiener index of simply generated random trees. *Random Structures and Algorithms*, 22(4):337–358, 2003.

[20] P. Kirschenhofer. On the average shape of monotonically labelled tree structures. *Discrete Applied Mathematics*, 7:161–181, 1984.

[21] P. Kirschenhofer and H. Prodinger. On the average oscillation of the contour of monotonically labelled ordered trees. *Proceedings of the Second World Conference on Mathematics at the Service of Man (Las Palmas)*, pages 394–399, 1982.

[22] D. E. Knuth. *The Art of Computer Programming: Volume 1, Fundamental Algorithms*. Addison-Wesley, 2002.

[23] H. Mahmoud and R. Neininger. Distribution of distances in random binary search trees. *Annals of Applied Probability*, 13(1):253 –276, 2003.

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

[25] K. Morris. On some parameters in monotonically labelled trees. In *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities*, Trends in Mathematics, pages 261–263. Birkhäuser, 2004.

[26] K. Morris, A. Panholzer, and H. Prodinger. On some parameters in heap ordered trees. *Combinatorics, Probability and Computing*, 13:677–696, 2004.

[27] K. Morris and H. Prodinger. Monotone tree structures revisited. *submitted*, 2005.

[28] R. Neininger. The Wiener index of random trees. *Combinatorics, Probability and Computing*, 11(6):587–597, 2002.

[29] R. Otter. The number of trees. *Annals of Mathematics*, 49:583–599, 1948.

[30] A. Panholzer. The distribution of the size of the ancestor tree and of the induced spanning subtree for random trees. *Random Structures and Algorithms*, 25:179–207, 2004.

[31] A. Panholzer and H. Prodinger. Bijections for ternary trees and non-crossing trees. *Discrete Mathematics*, 250:181–195, 2002.

[32] A. Panholzer and H. Prodinger. Analysis of some statistics for increasing tree families. *Discrete Mathematics and Theoretical Computer Science*, 6:437–460, 2004.

[33] A. Panholzer and H. Prodinger. Spanning tree size in random binary search trees. *Annals of Applied Probability*, 14:718–733, 2004.

[34] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Mathematica*, 68:145–254, 1937.

[35] H. Prodinger. Depth and path length of heap ordered trees. *International Journal of Foundations of Computer Science*, 7:293–299, 1996.

[36] H. Prodinger. The level of nodes in heap ordered trees. Appears on the following website `http://math.sun.ac.za/~prodinger/abstract/abs_126.htm`, 1997.

[37] H. Prodinger. Lecture notes on combinatorics. Appears on the following website `http://math.sun.ac.za/~prodinger/teach.htm`, 2001.

[38] H. Prodinger and F. J. Urbanek. On monotone functions of tree structures. *Discrete Applied Mathematics*, 5:223–239, 1983.

[39] B. Salvy and P. Zimmermann. Gfun: a Maple package for the manipulation of generating and holonomic functions in one variable. *ACM Transactions on Mathematical Software*, 20(2):163–177, 1994.

[40] R. P. Stanley. *Enumerative Combinatorics, Volume 1*. Cambridge University Press, 2002.

[41] R. P. Stanley. *Enumerative Combinatorics, Volume 2*. Cambridge University Press, 2002.

[42] H. Wiener. Structural determination of paraffin boiling points. *Journal of American Chemical Society*, 69:17–20, 1947.

[43] H. S. Wilf. *generatingfunctionology*. Academic Press, 1990.