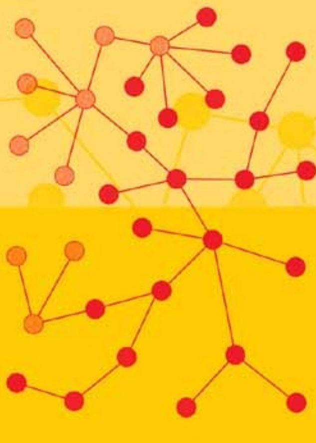
Michael Drmota

Random Trees

An Interplay between Combinatorics and Probability



SpringerWienNewYork

SpringerWienNewYork

Michael Drmota

Random Trees

An Interplay between Combinatorics and Probability

SpringerWienNewYork

Univ.-Prof. Dipl.-Ing. Dr. techn. Michael Drmota
Institute of Discrete Mathematics and Geometry, Vienna University
of Technology, Austria

Printed in Germany

SpringerWienNewYork is part of Springer Science + Business Media
springer.at

Typesetting: Camera ready by the author
Printing: Strauss GmbH, 69509 Mörlenbach, Germany

Printed on acid-free and chlorine-free bleached paper

With 72 black & white figures

SPIN: 12161386

Library of Congress Control Number: 2008942098

ISBN 978-3-211-75355-2 SpringerWienNewYork

To Gabriela, Heidi, Hanni and Peter

Preface

Trees are a fundamental object in graph theory and combinatorics as well as a basic object for data structures and algorithms in computer science. During the last years research related to (random) trees has been constantly increasing and several asymptotic and probabilistic techniques have been developed in order to describe characteristics of interest of large trees in different settings.

The purpose of this book is to provide a thorough introduction into various aspects of trees in random settings and a systematic treatment of the involved mathematical techniques. It should serve as a reference book as well as a basis for future research. One major conceptual aspect is to connect combinatorial and probabilistic methods that range from counting techniques (generating functions, bijections) over asymptotic methods (singularity analysis, saddle point techniques) to various sophisticated techniques in asymptotic probability (convergence of stochastic processes, martingales). However, the reading of the book requires just basic knowledge in combinatorics, complex analysis, functional analysis and probability theory of master degree level. It is also part of concept of the book to provide full proofs of the major results even if they are technically involved and lengthy.

Due to the diversity of the topic of the book it is impossible to present an exhaustive treatment of all known models of random trees and of all important aspects that have been considered so far. For example, we do not deal with the simulation of random trees. The choice of the topics reflects the author's taste and experience. It is slightly leaning on the combinatorial side and analytic methods based on generating functions play a dominant role in most of the parts of the book. Nevertheless, the general goal is to describe the limiting behaviour of large trees in terms of continuous random objects. This ranges from central (or other) limit theorems for simple tree statistics to functional limit theorems for the shape of trees, for example, encoded by the horizontal or vertical profile. The majority of the results that we present in this book is very recent.

There are several excellent books and survey articles dealing with some aspects on combinatorics on trees and graphs resp. with probabilistic methods in these topics which complement the present book. One of the first ones was Harary and Palmer book *Graphical enumeration* [98]. Around the same time Knuth published the first three volumes of *The Art of Computer Programming* [128, 129, 130] where several classes of trees related to algorithms from computer science are systematically investigated. His books with Green *Mathematics for the analysis of algorithms* [96] and the one with Graham and Patashnik *Concrete Mathematics* [95] complement this programme. In parallel asymptotic methods in combinatorics, many of them based on generating functions, became more and more important. The articles by Bender *Asymptotic methods in enumeration* [7] and Odlyzko *Asymptotic enumeration methods* [165] are excellent surveys on this topic. This development is highlighted by Flajolet and Sedgewick's recent (monumental) monograph *Analytic Combinatorics* [84]. Computer science and in particular the mathematical analysis of algorithms was always a driving force for developing concepts for the asymptotic analysis of trees (see also the books by Kemp [122], Hofri [102], Sedgewick and Flajolet [191], and by Szpankowski [197]). Moreover, several concepts of random trees arose naturally in this scientific process (see for example Mahmoud's book *Evolution of random search trees* [146], and Pittel's, Devroye's or Janson's work).

However, combinatorics and problems of computer science, though important, are not the only origin of random tree concepts. There was at least a second (and almost independent) line of research concerning conditioned Galton-Watson trees. Here one starts with a Galton-Watson branching process and conditions on the size of the resulting trees. For example, Kolchin's book *Random Mappings* [132] summarises many results from the *Russian school.* This work is complemented by the *American school* represented by Aldous [3, 5] and Pitman [171] where stochastic processes related to the Brownian motion play an important role. The invention of the continuum random tree as well as the ISE (integrated super-Brownian excursion) by Aldous are breakthroughs. Actually these continuous limit objects are quite universal concepts. It seems that they also appear as limit objects for several kinds of random planar maps and other related discrete objects. There are even more general settings where Lévy processes are used (see the recent survey articles *Random Trees and Applications* [135] and *Random Real Trees* [136] by Le Gall and the book *Probability and Real Trees* [75] by Evans). By the way, the study of random graphs is completely different from that of random trees (compare with the books by Bollobás [21], Janson, Łuczak and Ruciński [116], and Kolchin [133]). Nevertheless, there is a very interesting paper *The Birth of the Giant Component* [115] which uses analytic methods that are very close to tree methods.

This book is divided into nine chapters. The first two of them are providing some background whereas the remaining chapters 3–9 are devoted to more specific and (more or less) self contained topics on random trees and on related subjects. Of course, they will use basic notions from Chapter 1 and some of the methods from Chapter 2.

In Chapter 1 we survey several classes of random trees that are considered here: combinatorial tree classes like planted plane trees, Galton-Watson trees, recursive trees, and search trees including binary search trees and digital trees.

Chapter 2 is a second introductory chapter. It collects some basic facts on combinatorics with generating functions and provides an analytic treatment of generating functions that satisfy a functional equation (or a system of functional equations) leading to asymptotics and central limit theorems. It is probably not necessary to study all parts of this chapter in a first reading but to use it as a reference chapter.

The first purpose of Chapter 3 is tree counting, to obtain explicit formulas for the numbers of trees of given size with possible and asymptotic information on these numbers in those cases, where no or no simple explicit formula is available. The analysis of several combinatorial classes of trees and also of Galton-Watson trees is based on generating functions and their analytic properties that are discussed in Chapter 2. The recursive structure of (rooted) trees usually leads to a functional equation for the corresponding generating functions. By extending these counting procedures with the help of bivariate generating functions one can also study (so-called) additive statistics on these tree classes like the number of nodes of given degree or more generally the number of occurrences of a given pattern. In all these cases we derive a central limit theorem.

The general topic of Chapters 4–7 is the limiting behaviour of the profile and related statistics of different classes of random trees. Starting from a natural (vertex) labelling on a discrete object, for example the distance to a root vertex in a tree, the profile is the value distribution of the labels. More precisely, if a random discrete object has size n then the profile $(X_{n,k})$ is given by the numbers $X_{n,k}$ of vertices with label k. The idea behind is that the profile $(X_{n,k})$ describes the shape of the random object. It is therefore natural to search for a proper limiting object of the profile after a proper scaling.

In Chapter 4 we discuss the depth profile (induced by the distance to the root) of Galton-Watson trees with bounded offspring variance which can be approximated by the local time of the Brownian excursion of duration 1. This property is closely related to the convergence of normalised Galton-Watson trees to the continuum random tree introduced by Aldous [2, 3, 4]. The proof method that we use here follows the same principles as those of the previous chapters. We use multivariate generating functions and analytic methods. Interestingly these methods can be applied to unlabelled rooted trees, too, where we obtain the same approximation result. And the only successful approach to the latter class of trees – also called Pólya trees – is based on generating functions in combination with Pólya's theory of counting. Thus, Pólya trees look like Galton-Watson trees although they are definitely not of that kind.

Chapter 5 considers again Galton-Watson trees but a different kind of profile that is induced by a random walk on the tree. We fix an integer valued distribution η with zero mean. Then, given a tree T, every edge e of T is endowed with an independent copy η_e of η. The label of a node is then defined as the sum of η_e over all edges e on the path to the root. There are several motivations to study such random models. For example, if η has only values ± 1 or 0 and ± 1 then the resulting trees are closely related to random triangulations and quadrangulations. Furthermore, the random variables η_e can be seen as random increments in an embedding of the tree in the space. This idea is originally due to Aldous [5] and gave rise of the ISE, the integrated super-Brownian excursion, which acts as the limiting occupation measure of the induced label distribution. The final result is that the corresponding profile can be approximated by the (random) density of the ISE. This result reaches very far and is out of scope of this book but, nevertheless, there are special cases which are of particular interest and capable for the framework of the present book. By the use of explicit generating functions of unexpected form the analysis recovers one-dimensional versions of the functional limit theorem and also leads to integral representations for several parameters of the ISE. These observations are due to Bousquet-Mélou [23].

Chapter 6 deals with recursive trees and their variants (plane oriented recursive trees, binary and m-ary search trees). The interesting feature of these kinds of trees is that they can be seen from different points of views: They can be seen as a combinatorial object (where usual counting procedures apply) as well as the result of a (stochastic) growth process. Interestingly their asymptotic structure is completely different from that of Galton-Watson trees. They are so-called $\log n$ trees which means that their expected height is of order $\log n$ (in contrast to Galton-Watson trees with expected height of order $\sqrt{n}$). We provide a unified approach to several basic statistics like the degree distribution. However, the main focus is again the profile. Here one observes that most vertices are concentrated around few levels so that a (possible) limiting object of the normalised project is not related to some functional of the Brownian motion. Nevertheless, the normalised profile $X_{n,k}/\mathbb{E}\,X_{n,k}$ can be approximated by $X(k/\log n)$, where $X(t)$ is now a random analytic function. We also deal with the height and its concentration properties.

Tries and digital search trees are two other classes of $\log n$ trees which are discussed in Chapter 7. Their construction is based on digital keys and not on the order structure of the keys as in the case of binary search trees. Again, most vertices are concentrated around few levels of order $\log n$ but the profile behaves differently. It is even more concentrated around its mean value than the profile of binary search trees or recursive trees. The normalised profile $X_{n,k}/\mathbb{E}\,X_{n,k}$ (of tries) converges to 1 and we observe a central limit theorem.

Chapter 8 is devoted to the so-called contraction method which was developed to handle stochastic recurrence relations which naturally appear in the stochastic analysis of recursive algorithms like Quicksort. Such recurrences also appear in the analysis of the profile of recursive trees and binary search trees (and their variants). The idea is that after normalisation the recurrence relation stabilises to a (stochastic) fixed point equation that can be solved uniquely by Banach's fixed point theorem in a properly chosen Banach space setting. Here we restrict ourselves to an L_2 setting with the Wasserstein metric. We mainly follow the work by Rösler, Rüschendorf, Neininger [158, 161, 162, 186, 187].

The final Chapter 9 deals with planar graphs. At first sight planar graphs and trees have nothing in common but there are strong similarities in the combinatorial and asymptotic analysis. For example the 2-connected parts of a connected (planar) graph have a tree structure which is reflected by the structure of the corresponding generating functions. In particular in the asymptotic analysis one can use the same techniques from Chapter 2 as for combinatorial tree classes in Chapter 3. Besides the asymptotic counting problem the major goal of this chapter is to study the degree distribution of random planar graphs or equivalently the expected number of vertices of given degree where we can again use asymptotic tree counting techniques. This chapter is based on recent work by Giménez, Noy and the author [63, 64].

Of course, such a book project cannot be completed without help and support from many colleagues and friends. In particular I am grateful to Mireille Bousquet-Mélou, Luc Devroye, Philippe Flajolet, Bernhard Gittenberger, Alexander Iksanov, Svante Janson, Christian Krattenthaler, Jean-François Marckert, Marc Noy, Ralph Neininger, Alois Panholzer, and Wojciech Szpankowski. I also thank Frank Emmert-Streib for helping me to design the book cover.

Finally I want to thank Veronika Kraus, Johannes Morgenbesser, and Christoph Strolz for their careful reading of the manuscript and for several hints to improve the presentation and Barbara Doležal-Rainer for her support in type setting. I also want to thank Stephen Soehnlen from Springer Verlag for his constant support in this book project and his patience.

I am especially indebted to my family to whom this book is dedicated.

Vienna, November 2008 *Michael Drmota*

Contents

1

Classes of Random Trees

In this first chapter we survey several types of random trees. We start with basic notions on trees and the description of several concepts of tree counting problems. In particular we distinguish between rooted and unrooted, plane and non-plane, and labelled and unlabelled trees. It is also possible to modify the counting procedure by putting certain weights on trees, for example, by using the degree distribution.

We consider classical combinatorial tree classes like planted plane trees or labelled rooted trees. Furthermore we discuss simply generated trees which can be also considered as conditioned Galton-Watson trees and cover several classes of the classical (rooted) trees. We introduce unlabelled trees (also called Pólya trees) that do not fall into this class but behave similarly to simply generated trees. Recursive trees (and more generally increasing trees) are labelled rooted trees where each path starting at the root has increasing labels. All these kinds of trees give rise to a natural probability distribution based on combinatorics by assuming that every tree of size n (of a certain class) is equally likely.

Trees occur also in the context of algorithms from computer science, for example, as data structures. Here the structure of the tree is determined by the input data of the algorithm. Prominent examples are binary search trees, digital search trees or tries. From a combinatorial point of view these kinds of trees are just binary trees. However, if we assume some probability distribution on the input data this induces a probability distribution on the corresponding trees. Moreover, one usually has a tree evolution process by inserting more and more data.

1.1 Basic Notions

Trees are defined as connected graphs without cycles, and their properties are basics of graph theory. For example, a connected graph is a tree, if and only if the number of edges equals the number of nodes minus 1. Furthermore, each pair of nodes is connected by a unique path.

The degree $d(v)$ of a node v in a tree is the number of nodes that are adjacent to v or the number of neighbours of v.

Nodes of degree ≤ 1 are usually called *leaves* or *external nodes* and the remaining ones *internal nodes*.

1.1.1 Rooted Versus Unrooted trees

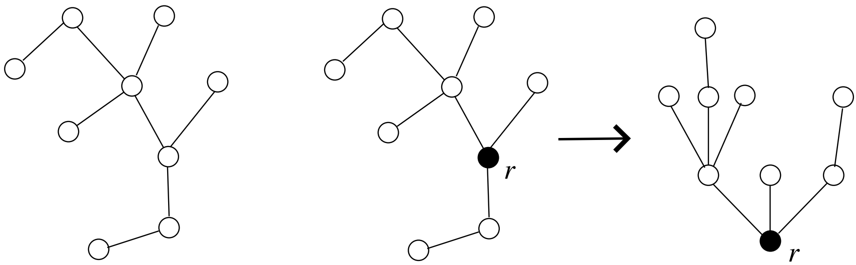


Fig. 1.1. Tree and rooted tree

If we mark a specific node r in a tree T, which we denote the *root* of T, we call the tree itself *rooted tree*. A rooted tree may be described easily in terms of *generations* or *levels*. The root is the 0-th generation. The neighbours of the root constitute the first generation, and in general the nodes at distance k from the root form the k-th generation (or level). If a node of level k has neighbours of level $k+1$ then these neighbours are also called *successors*. The number of successors of a node v is also called the *out-degree* $d^+(v)$. For all nodes v different from the root we have $d(v) = d^+(v) + 1$.

Furthermore, if v is a node in a rooted tree T then v may be considered as the root of a subtree T_v of T that consists of all iterated successors of v. This means that rooted trees can be constructed in a recursive way. Due to that property counting problems on rooted trees are usually easier than on unrooted trees.

Remark 1.1 *Rooted trees also have various applications in computer science. They naturally appear as data structures, e.g. the recursive structure of folders in any computer is just a rooted tree. Furthermore, fundamental algorithms such as Quicksort or the Lempel-Ziv data compression algorithm are closely* *related to rooted trees, namely to binary and digital search trees which are also used to store (and search for) data. Rooted trees even occur in information theory. For example, prefix free codes on an alphabet of order m are encoded as the set of leaves in m-ary trees.*

1.1.2 Plane Versus Non-Plane trees

Trees are planar graphs since they can be embedded into the plane without crossings. Nevertheless, a tree may have different embeddings (compare with Figure 1.2). This makes a difference in counting problems. When we say that we are counting *planar trees* we mean that we are counting all possible different embeddings into the plane.

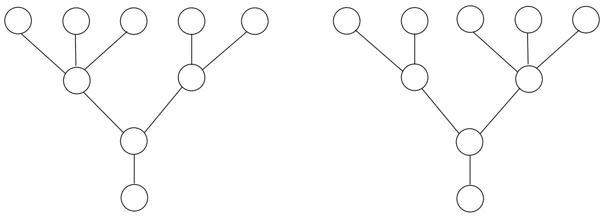


Fig. 1.2. Two different embeddings of a tree

In the context of rooted trees it is common to use the term *plane tree* or *ordered tree* when successors of the root and recursively the successors of each node are equipped with a *left-to-right-order*. Alternatively one can give the successors a rank so that one can speak of the j-th successor ($j \geq 1$). Of course, this induces a natural embedding into the half-plane (compare with Figure 1.3). Note that this notion is different from considering all embeddings into the plane, since it is not allowed to rotate the subtrees of the root cyclically around the root.

1.1.3 Labelled Versus Unlabelled Trees

We also distinguish between labelled trees, where the nodes are labelled by different numbers, and unlabelled trees, where nodes are indistinguishable. This is particularly important for the counting problem. For example, there is only one unlabelled tree with three nodes whereas there are three different labelled trees of size 3 with labels $1, 2, 3$ (see Figure 1.4).

There is much latitude in choosing labels on trees. The simplest model is to assume that the nodes of a trees of size n are labelled by the numbers $1, 2, \ldots, n$, but there are many other ways to do so. For so-called embedded trees one only assumes that the labels of adjacent vertices differ (at most) by

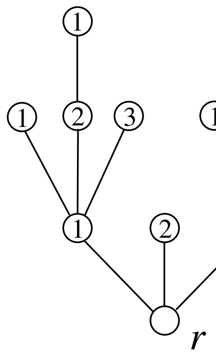


Fig. 1.3. Plane rooted tree

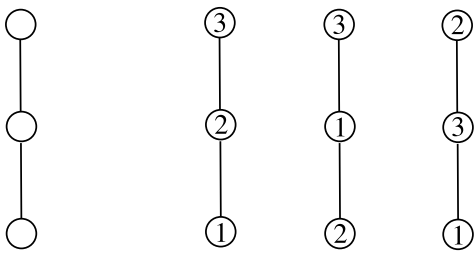


Fig. 1.4. Unlabelled versus labelled trees

1. Another possibility is to put labels consistently with the structure of the tree. For example, recursive trees have the property that the root is labelled by 1 and the labels on all paths away from the root are strictly increasing.

1.2 Combinatorial Trees

Let $\mathcal{T}$ be a class of finite trees which is defined by a structural condition (for example that the trees are binary). We then consider the subclasses $\mathcal{T}_n$ of $\mathcal{T}$ that consist of trees of size n and introduce a probability model on $\mathcal{T}_n$ by assuming that every tree T in $\mathcal{T}_n$ is equally likely. By this construction we get special kinds of random trees. Moreover, every parameter on trees (such as the number of leaves or the diameter) is then a random variable.

For simplicity we start with rooted trees since they have a recursive description.

1.2.1 Binary Trees

Binary trees are rooted trees, where each node is either a leaf (that is, it has no successor) or it has two successors. Usually these two successors are distinguishable: the left successor and the right successor, that is, we are dealing with plane trees. The leaves of a binary tree are also called *external* nodes and those nodes with two successors *internal* nodes. It is clear that a binary tree with n internal nodes has $n+1$ external nodes. Thus, the total number of nodes is always odd.

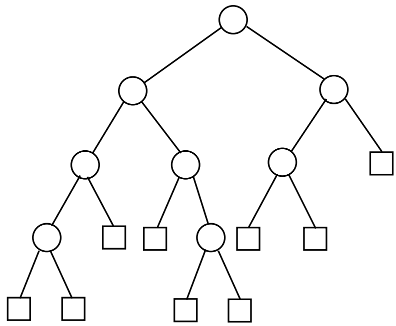


Fig. 1.5. Binary tree

A very important issue is that binary trees (and many other kinds of rooted trees) have a recursive structure. More precisely we can use the following *recursive definition* of binary trees:

> A binary tree $\mathcal{B}$ is either just an external node or an internal node (the root) with two subtrees that are again binary trees.

Formally we can write this in the form

$$\mathcal{B} = \square + \circ \times \mathcal{B} \times \mathcal{B}, \tag{1.1}$$

where $\mathcal{B}$ denotes the system of binary trees; $\square$ represents an external and $\circ$ an internal node.

In fact, this recursive description is the key for the analysis of many properties of binary (and similarly defined) trees. In particular, this formal equation has a direct translation into an equation for the corresponding generating (or counting) function $b(x)$ of the form $b(x) = 1 + xb(x)^2$. We discuss this translation in detail in Chapter 2.

A direct generalisation of binary trees is m-ary rooted trees, where $m \ge 2$ is a fixed integer. As in the binary case ($m = 2$) we just take into account the number n of internal nodes. The number of leaves is then given by $(m-1)n+1$ and the total number of nodes by $mn+1$.

Interestingly it is relatively easy to find explicit formulas for the numbers $b_n^{(m)}$ of m-ary trees with n internal nodes:

$$b_n^{(m)} = \frac{1}{(m-1)n+1}\binom{mn}{n}.$$

The set $\mathcal{T}_n$ of m-ary trees with n internal nodes then constitutes a set of *random trees* if we assume that every m-ary tree in $\mathcal{T}_n$ is equally likely, namely of probability $1/b_n^{(m)}$.

Note that in the binary case the number of trees is precisely the n-th *Catalan number*

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

It is also possible to consider binary and more generally m-ary trees, where the left-to-right-order of the successors is not taken into account. However, the counting problem of these classes of trees is much more involved (compare with Sections 1.2.5 and 3.1.5).

1.2.2 Planted Plane Trees

Another interesting class of trees are *planted plane trees*. Sometimes they are also called *Catalan trees*. Planted plane trees are again rooted trees, where each node has an arbitrary number of successors with a natural left-to-right-order (this again means that we are considering plane trees). The term *planted* comes from the interpretation that the root is connected (or planted) to an additional *phantom* node that is not taken into account (see Figure 1.6). Usually we will not even depict this additional node when we deal with planted trees. However, it is quite useful to define the degree of the root r by $d(r) = d^+(r) + 1$ which means that the additional (planted) node is considered a neighbour node. This has the advantage that in this case all nodes have the property $d(v) = d^+(v) + 1$.

The numbers p_n of planted plane trees with $n \ge 1$ nodes are given by

$$p_n = \frac{1}{n}\binom{2n-2}{n-1}.$$

This is precisely the $(n-1)$-st Catalan number C_{n-1} which explains the term *Catalan tree*. By the way, the relation $p_{n+1} = b_n$ has a natural interpretation (see Section 3.1.2).

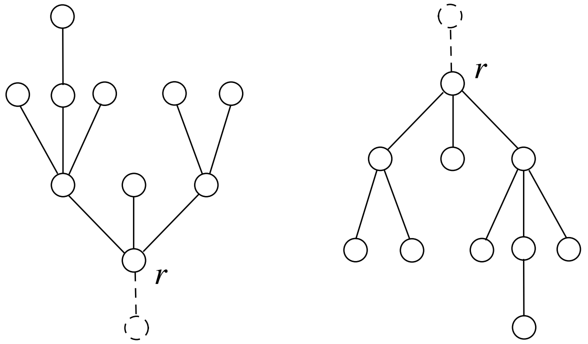


Fig. 1.6. Planted plane tree

1.2.3 Labelled Trees

We recall that a tree T of size n is labelled if the n nodes are labelled by $1, 2, \ldots, n$.[1] The counting problem of labelled trees is different from that of unlabelled trees. There is, however, an easy connection between rooted and unrooted labelled trees. There are exactly n different ways to make an unrooted tree to a rooted one by choosing one of the labelled nodes. Thus, the number of rooted labelled trees of size n equals the number of unrooted labelled trees exactly n times. Consequently it is sufficient to consider rooted labelled trees which has the advantage that one can use the recursive structure.

Note that if we do not care about the embedding in the plane or about the left to right order of the successors, an unrooted labelled tree can be interpreted as a spanning tree of the complete graph K_n with nodes $1, 2, \ldots, n$ (see Figure 1.7).

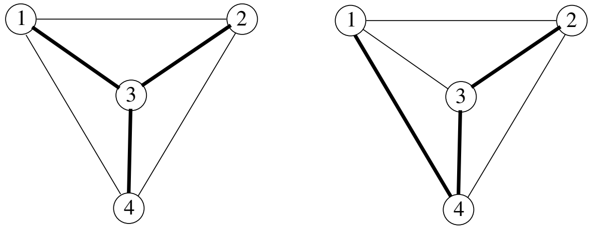


Fig. 1.7. 2 of 16 possible spanning trees of K_4

[1] Other kinds of labelled trees like recursive trees or well-labelled trees will be discussed in the sequel.

It is a well known fact that the number of unrooted labelled trees of size n equals n^{n-2} (usually called Cayley's formula). Hence, there are n^{n-1} different rooted labelled trees of size n. Sometimes these trees are called *Cayley trees* (but this term is also used for infinite regular trees).

1.2.4 Labelled Plane Trees

It is also of interest to count the number of different planar embeddings of labelled trees. There is even an explicit formula, namely for $n \ge 2$ there are

$$\frac{(2n-3)!}{(n-1)!}$$

different planar embeddings of labelled trees of size n (and $n(2n-3)!/(n-1)!$ different planar embeddings of rooted labelled trees of size n). For example, for $n = 4$ there are $4^2 = 16$ different labelled trees but $5!/3! = 20$ different planar embeddings.

1.2.5 Unlabelled Trees

Let $\tilde{\mathcal{T}}$ denote the set of unlabelled unrooted trees and $\mathcal{T}$ be the set of unlabelled rooted trees. Here we do not care about the possible embeddings into the plane. We just think of trees in the graph-theoretical sense.

These kinds of trees are relatively difficult to count. Let us denote by $\tilde{t}_n$ and t_n the corresponding numbers of those trees of size n, for example we have

$$\tilde{t}_1 = 1,\ \tilde{t}_2 = 1,\ \tilde{t}_3 = 1,\ \tilde{t}_4 = 2 \quad \text{and} \quad t_1 = 1,\ t_2 = 1,\ t_3 = 2,\ t_4 = 4.$$

However, if there is no direct recursive relation one has to take into account all symmetries. Nevertheless, this problem can be solved by using generating functions and Pólya's theory of counting [176] (see Section 3.1.5). For that reason these trees are also called *Pólya trees*.

In order to give an impression of the kind of problems one has to face we just state that the generating functions

$$\tilde{t}(x) = \sum_{n\ge 1} \tilde{t}_n x^n \quad \text{and} \quad t(x) = \sum_{n\ge 1} t_n x^n$$

satisfy the relations

$$t(x) = x \exp\left(t(x) + \frac{1}{2} t(x^2) + \frac{1}{3} t(x^3) + \cdots \right) \tag{1.2}$$

and

$$\tilde{t}(x) = t(x) - \frac{1}{2} t(x)^2 + \frac{1}{2} t(x^2). \tag{1.3}$$

It seems that there is no proper explicit formula for t_n and $\tilde{t}_n$. However, there are asymptotic expansions for them and by using extensions of the mentioned counting procedure it is also possible to study several shape characteristics of these kinds of trees.

1.2.6 Unlabelled Plane Trees

We already mentioned that a tree usually has several different embeddings into the plane. Planted plane trees are, in particular, designed to take into account all possible planar embeddings of planted rooted trees.

It is, however, another non-trivial step to count all embeddings of unlabelled rooted trees and all embeddings of unlabelled trees. Again we have to take into account symmetries. Fortunately Pólya's theory can be applied here, too. As in the case of unlabelled trees we do not get explicit formulas but asymptotic expansions (see Section 3.1.6).

1.2.7 Simply Generated Trees – Galton-Watson Trees

Simply generated trees are weighted versions of rooted trees and have been introduced by Meir and Moon [151]. The idea is to put a weight to a rooted tree according to its degree distribution.

Let ϕ_j, $j \ge 0$, be a sequence of non-negative real numbers, called the *weight sequence*. Usually one assumes that $\phi_0 > 0$ and $\phi_j > 0$ for some $j \ge 2$. We then define the weight $\omega(T)$ of a finite rooted ordered tree T by

$$\omega(T) = \prod_{v \in V(T)} \phi_{d^+(v)} = \prod_{j \ge 0} \phi_j^{D_j(T)},$$

where $d^+(v)$ denotes the out-degree of the vertex v (or the number of successors) and $D_j(T)$ the number of nodes in T with j successors. The numbers

$$y_n = \sum_{|T|=n} \omega(T)$$

are then the weighted numbers of trees of size n. It is natural to define a probability distribution on the set $\mathcal{T}_n$ by

$$\pi_n(T) = \frac{\omega(T)}{y_n} \qquad (T \in \mathcal{T}_n). \tag{1.4}$$

It is convenient to introduce the generating series

$$\Phi(x) = \phi_0 + \phi_1 x + \phi_2 x^2 + \cdots = \sum_{j \ge 0} \phi_j x^j .$$

In Section 3.1.4 we will show that the generating function $y(x) = \sum_{n\ge 1} y_n x^n$ satisfies the equation

$$y(x) = x\,\Phi(y(x)).$$

This equation is the key for the asymptotic analysis of these kinds of trees.

If we replace ϕ_j by $\tilde{\phi}_j = ab^j \phi_j$, which is the same as replacing $\Phi(x)$ by $\tilde{\Phi}(x) = a\Phi(bx)$ for two numbers $a, b > 0$, then $\omega(T)$ is replaced by

$$\tilde{\omega}(T) = \prod_{j\ge 0} \left(ab^j\phi_j\right)^{D_j(T)} = a^{|T|}b^{|T|-1}\omega(T).$$

Note that $\sum_j jD_j(T) = |T| - 1$. Hence, $\tilde{y}_n = a^n b^{n-1} y_n$ and the probability distribution π_n on $\mathcal{T}_n$ is the same for $\tilde{\Phi}(x)$ and $\Phi(x)$ (for every n). Usually only these distributions are important, and we may then freely make this type of modification of ϕ_j.

Simply generated trees generalise several of the above examples of combinatorial trees.

Example 1.2 *If $\phi_j = 1$ for all $j \ge 0$, that is, $\Phi(x) = 1/(1-x)$, then all planted plane trees have weight $\omega(T) = 1$ and y_n is the number of planted plane trees. Thus, π_n is the uniform distribution on planted plane trees of size n.*

Example 1.3 *Binary trees (counted according to their internal nodes) are also covered by this approach. If we set $\phi_0 = 1$, $\phi_1 = 2$, $\phi_2 = 1$, and $\phi_j = 0$ for $j \ge 3$, that is, $\Phi(x) = (1+x)^2$, then nodes with one successor get weight 2. This takes into account that binary trees (where external nodes are disregarded) have two kinds of nodes with one successor, namely those with a left branch but no right branch and those with a right branch but no left branch. Thus, π_n is the uniform distribution on all binary trees with n internal nodes.*

Similarly, m-ary trees are covered with the help of the weights $\phi_j = \binom{m}{j}$ or with $\Phi(x) = (1+x)^m$.

Example 1.4 *If $\phi_0 = \phi_1 = \phi_2 = 1$ and $\phi_j = 0$ for $j \ge 3$ or $\Phi(x) = 1+x+x^2$, then we get so-called Motzkin trees. Here only rooted trees, where all nodes have less than 3 successors, get (a non-zero) weight $\omega(T) = 1$: y_n is the number of Motzkin trees with n nodes and π_n is the uniform distribution on Motzkin trees of size n.*

Example 1.5 *If we set $\phi_j = 1/j!$ then*

$$n! \cdot y_n = n^{n-1}$$

denotes precisely the number of labelled rooted non-plane trees. The weight $\phi_j = 1/j!$ disregards all possible orderings of the successors of a vertex of out-degree j and the factor $n!$ corresponds to all possible labellings of n nodes. Hence, π_n yields the uniform distribution on labelled rooted trees.

Interestingly there is an intimate relation to *Galton-Watson branching processes*. Let ξ be a non-negative integer-valued random variable, the so-called offspring distribution. The Galton-Watson branching process starts with a single individual (generation 0); each individual has a number of children distributed as independent copies of ξ. If Z_k denotes the size of the generation k, then a formal description of the process $(Z_k)_{k\ge 0}$ is $Z_0 = 1$, and for $k \ge 1$

$$Z_k = \sum_{j=1}^{Z_{k-1}} \xi_j^{(k)},$$

where the $(\xi_j^{(k)})_{k,j}$ are i.i.d.[2] random variables distributed as ξ.

It is clear that Galton-Watson branching processes can be represented by ordered (finite or infinite) rooted trees T such that the sequence Z_k is just the number of nodes at level k and $\sum_{k\ge 0} Z_k$ (which is called the *total progeny*) is the number of nodes $|T|$ of T. We denote by $\nu(T)$ the probability that a specific tree T occurs. If $\mathbb{P}\{\xi = 0\} = 0$ then the total progeny is infinite with probability 1. Thus we always assume that $\mathbb{P}\{\xi = 0\} > 0$.

The generating function $y(x) = \sum_{n\ge 1} y_n x^n$ of the numbers

$$y_n = \mathbb{P}\{|T| = n\} = \sum_{|T|=n} \nu(T)$$

satisfies the functional equation

$$y(x) = x\,\Phi(y(x)),$$

where

$$\Phi(t) = \mathbb{E}\, t^{\xi} = \sum_{j\ge 0} \phi_j t^j$$

with $\phi_j = \mathbb{P}\{\xi = j\}$. Observe that

$$\nu(T) = \prod_{j\ge 0} \phi_j^{D_j(T)} = \omega(T).$$

The *weight* of T is now the *probability* of T.

If we condition the Galton-Watson tree T on $|T| = n$, we thus get the probability distribution (1.4) on $\mathcal{T}_n$. Hence, the conditioned Galton-Watson trees are simply generated trees with $\phi_j = \mathbb{P}\{\xi = j\}$ as above. We have here $\Phi(1) = \sum_j \phi_j = 1$, but this is no real restriction. In fact, if $(\phi_j)_{j\ge 0}$ is any sequence of non-negative weights satisfying the very weak condition $\Phi(x) = \sum_{j\ge 0} \phi_j x^j < \infty$ for some $x > 0$, then we can replace (as above) ϕ_j by $ab^j\phi_j$ with $b = x$ and $a = 1/\Phi(x)$ and thus the simply generated tree is the same as the conditioned Galton-Watson tree with offspring distribution $\mathbb{P}\{\xi = j\} = \phi_j x^j/\Phi(x)$. Consequently, for all practical purposes, simply generated trees are the same as conditioned Galton-Watson trees.

The argument above also shows that the distribution of a conditioned Galton-Watson tree is not changed if we replace the offspring distribution ξ by $\tilde{\xi}$ with $\mathbb{P}\{\tilde{\xi} = j\} = \mathbb{P}\{\xi = j\} = \tau^j/\Phi(\tau)$ and thus $\tilde{\Phi}(x) = \Phi(\tau x)/\Phi(\tau)$ for any $\tau > 0$ with $\Phi(\tau) < \infty$. (Such modifications are called conjugate or tilted distributions.)

[2] The letters "i.i.d." abbreviate "independent and identically distributed".

Note that

$$\mu = \Phi'(1) = \mathbb{E}\,\xi$$

is the expected value of the offspring distribution. If $\mu < 1$, the Galton-Watson branching process is called sub-critical, if $\mu = 1$, then it is critical, and if $\mu > 1$, then it is supercritical. From a combinatorial point of view we do not have to distinguish between these three cases. Namely, if we replace the offspring distribution by a conjugate distribution as above, the new expected value is

$$\tilde{\Phi}'(1) = \frac{\tau\Phi'(\tau)}{\Phi(\tau)}.$$

We can thus always assume that the Galton-Watson process is critical, provided only that there exists $\tau > 0$ with

$$\tau\Phi'(\tau) = \Phi(\tau) < \infty,$$

a weak condition that is satisfied for all interesting classes of Galton-Watson trees.

It is usually convenient to choose a critical version, which explains why the equation $\tau\Phi'(\tau) = \Phi(\tau)$ appears in most asymptotic results. A heuristic reason is that the probability of the event $|T| = n$ that we condition on typically decays exponentially in the subcritical and supercritical cases but only as $n^{-1/2}$ in the critical case, and it seems advantageous to condition on an event of not too small probability.

Example 1.6 *For planted plane trees (as in Example 1.2) we start with* $\Phi(x) = 1/(1-x)$*. The equation* $\tau\Phi'(\tau) = \Phi(\tau)$ *is* $\tau(1-\tau)^{-2} = (1-\tau)^{-1}$*, which is solved by* $\tau = \frac{1}{2}$*. Random planted plane trees are thus conditioned Galton-Watson trees with the critical offspring distribution given by* $\Phi(x) = (1-x/2)^{-1}/2 = 1/(2-x)$*, or* $\mathbb{P}\{\xi = j\} = 2^{-j-1}$ *(for* $j \geq 0$*), a geometric distribution.*

Example 1.7 *Similarly random binary trees are obtained with a binomial offspring distribution* $\mathrm{Bi}(2, 1/2)$ *with* $\Phi(x) = (1+x)^2/4$*, and more generally random m-ary trees are obtained with offspring law* $\mathrm{Bi}(m, 1/m)$ *with* $\Phi(x) = ((m-1+x)/m)^m$*.*

Example 1.8 *For Motzkin trees the critical offspring distribution* ξ *is uniform on* $\{0, 1, 2\}$ *with* $\Phi(x) = (1+x+x^2)/3$*.*

Example 1.9 *For uniform rooted labelled trees the critical* ξ *has a Poisson distribution* $\mathrm{Po}(1)$ *with* $\Phi(x) = e^{x-1}$*.*

Finally we remark that for a critical offspring distribution ξ, its variance is given by

$$\sigma^2 = \mathbb{V}\mathrm{ar}\,\xi = \mathbb{E}\,\xi^2 - 1 = \mathbb{E}(\xi(\xi-1)) = \Phi''(1).$$

Starting with an arbitrary sequence $(\phi_j)_{j\geq 0}$ and modifying it as above the get a critical probability distribution, we obtain the variance

$$\sigma^2 = \tilde{\Phi}''(1) = \frac{\tau^2\Phi''(\tau)}{\Phi(\tau)},$$

where $\tau > 0$ is such that $\tau\Phi'(\tau) = \Phi(\tau) < \infty$ (assuming this is possible). We will see that this quantity appears in several asymptotic results.

1.3 Recursive Trees

Recursive trees are rooted labelled trees, where the root is labelled by 1 and the labels of all successors of any node v are larger than the label of v (see Figure 1.8).

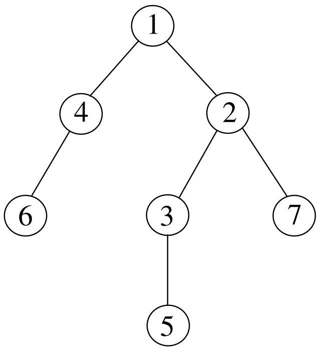


Fig. 1.8. Recursive tree

1.3.1 Non-Plane Recursive Trees

Usually one does not take care of the possible embeddings of a recursive tree into the plane. In this sense recursive trees can be seen as the result of the following evolution process. Suppose that the process starts with a node carrying the label 1. This node will be the root of the tree. Then attach a node with label 2 to the root. The next step is to attach a node with label 3. However, there are two possibilities: either to attach it to the root or to the node with label 2. Similarly one proceeds further. After having attached the nodes with labels $1, 2, \ldots, k$, attach the node with label $k+1$ to one of the existing nodes.

Obviously, every recursive tree of size n is obtained in a unique way. Moreover, the labels represent something like the history of the evolution process.

Since there are exactly k ways to attach the node with label $k+1$, there are exactly $(n-1)!$ possible trees of size n.

The natural probability distribution on recursive trees of size n is to assume that each of these $(n-1)!$ trees is equally likely. This probability distribution is also obtained from the evolution process by attaching successively each new node to one of the already existing nodes with equal probability.

Remark 1.10 *Historically, recursive trees appear in various contexts. They are used to model the spread of epidemics (see [155]) or to investigate and construct family trees of preserved copies of ancient manuscripts (see [157]). Other applications are the study of the schemes of chain letters or pyramid games (see [88]).*

1.3.2 Plane Oriented Recursive Trees

Note that the left-to-right-order of the successors of the nodes in a recursive tree was not relevant in the above counting procedure. It is, however, relatively easy to consider all possible embeddings as plane rooted trees. These kind of trees are usually called plane oriented recursive trees (PORTs).

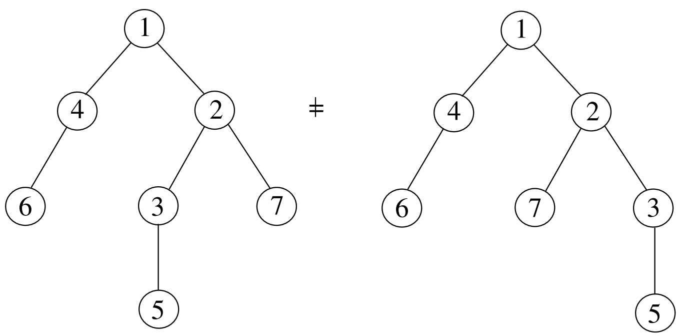


Fig. 1.9. Two different plane oriented trees

They can again be seen as the result of an evolution process, where the left-to-right-order of the successors is taken into account. More precisely, if a node v has out-degree d, then there are $d+1$ possible ways to attach a new node to v. Hence, the number of different plane oriented recursive trees with n nodes equals

$$1 \cdot 3 \cdot \ldots \cdot (2n-3) = (2n-3)!! = \frac{1}{2^{n-1}} \frac{(2(n-1))!}{(n-1)!}.$$

As above, the natural probability distribution on plane oriented recursive trees of size n is to assume that each of these $(2n-3)!!$ trees is equally likely.

This probability distribution is also obtained from the evolution process by attaching each node with probability proportional to the out-degree plus 1 to the already existing nodes.

1.3.3 Increasing Trees

The probabilistic model of simply generated trees was to define a weight that reflects the degree distribution of rooted trees. The same idea can be applied to recursive and to plane oriented recursive trees. The resulting classes of trees are called *increasing trees*. They have been first introduced by Bergeron, Flajolet, and Salvy [12].

As above we define the weight $\omega(T)$ of a recursive or a plane oriented recursive tree T by

$$\omega(T) = \prod_{v \in V(T)} \phi_{d^+(v)} = \prod_{j \ge 0} \phi_j^{D_j(T)},$$

where $d^+(v)$ denotes the out-degree of the vertex v (or the number of successors) and $D_j(T)$ the number of nodes in T with j successors. Then we set

$$y_n = \sum_{T \in \mathcal{J}_n} \omega(T),$$

where $\mathcal{J}_n$ denotes the set of recursive or plane oriented recursive trees of size n. The natural probability distribution on the set $\mathcal{J}_n$ of increasing trees is then given by

$$\pi_n(T) = \frac{\omega(T)}{y_n} \qquad (T \in \mathcal{J}_n).$$

As in the case of simply generated trees it is also possible to introduce generating series. We set

$$\Phi(x) = \phi_0 + \phi_1 x + \phi_2 \frac{x^2}{2!} + \phi_2 \frac{x^3}{3!} + \cdots$$

in the case of recursive trees and

$$\Phi(x) = \phi_0 + \phi_1 x + \phi_2 x^2 + \phi_3 x^3 + \cdots$$

in the case of plane oriented recursive trees. The generating function

$$y(z) = \sum_{n \ge 0} y_n \frac{z^n}{n!}$$

satisfies the differential equation

$$y'(z) = \Phi(y(z)), \quad y(0) = 0.$$

In the interest of clarity we state how the general concept specialises.

1. Recursive trees (that is, every non-planar recursive tree gets weight 1) are given by $\Phi(x) = e^x$. Here $y_n = (n-1)!$ and $y(z) = \log(1/(1-z))$.
2. Plane oriented recursive trees are given by $\Phi(x) = 1/(1-x)$. This means that every planar recursive tree gets weight 1. Here $y_n = (2n-3)!! = 1 \cdot 3 \cdot 5 \cdots (2n-3)$ and $y(z) = 1 - \sqrt{1-2z}$.
3. Binary recursive trees are defined by $\Phi(x) = (1+x)^2$. We have $y_n = n!$ and $y(z) = 1/(1-z)$. The probability model that is induced by this (planar) binary increasing trees is exactly the standard permutation model of *binary search trees* that is discussed in Section 1.4.1.

Note that the probability distribution on $\mathcal{J}_n$ is not automatically given by an evolution process as it is definitely the case for recursive trees and plane oriented recursive trees. It is interesting that there are precisely three families of increasing trees, where the probability distribution π_n is also induced by a (natural) tree evolution process.

1. $\Phi(x) = \phi_0 e^{\frac{\phi_1}{\phi_0}x}$ with $\phi_0 > 0$, $\phi_1 > 0$.
2. $\Phi(x) = \phi_0 \left(1 - \dfrac{\phi_1}{r\phi_0}x\right)^{-r}$ for some $r > 0$ and $\phi_0 > 0$, $\phi_1 > 0$.
3. $\Phi(x) = \phi_0 \left(1 + (\phi_1/(d\phi_0))x\right)^d$ for some $d \in \{2, 3, \ldots\}$ and $\phi_0 > 0$, $\phi_1 > 0$.

The corresponding tree evolution process runs as follows:[3] The starting point is (again) a node (the root) with label 1. Now assume that a tree T of size n is present. We attach to every node v of T a *local weight* $\rho(v) = (k+1)\phi_{k+1}\phi_0/\phi_k$ when v has k successors and set $\rho(T) = \sum_{v \in V(T)} \rho(v)$. Observe that in a planar tree there are $k+1$ different ways to attach a new (labelled) node to an (already existing) node with k successors. Now choose a node v in T according to the probability distribution $\rho(v)/\rho(T)$ and then independently and uniformly one of the $k+1$ possibilities to attach a new node there (when v has k successors). This construction ensures that in these three particular cases a tree T of size n, which occurs with probability proportional to $\omega(T)$, generates a tree T' of size $n+1$ with probability that is proportional to $\omega(T)\phi_{k+1}\phi_0/\phi_k$, which equals $\omega(T')$. Thus, this procedure induces the same probability distribution on $\mathcal{J}_n$ as the one mentioned above, where a tree $T \in \mathcal{J}_n$ has probability $\omega(T)/y_n$.

Note that if we are only interested in the distributions π_n, then we can work (without loss of generality) with some special values for ϕ_0 and ϕ_1. It is sufficient to consider the generating functions

1. $\Phi(x) = e^x$,
2. $\Phi(x) = (1-x)^{-r}$ for some $r > 0$,
3. $\Phi(x) = (1+x)^d$ for some $d \in \{2, 3, \ldots\}$.

The first class is just the class of recursive trees. The second class can be interpreted as generalised plane oriented recursive trees, since the probability

[3] In the interest of brevity we only discuss the plane version.

of choosing a node with out-degree j is proportional to $j + r$. For $r = 1$ we get (usually) plane oriented recursive trees. The trees in the third class are so-called d-ary recursive trees; they correspond to an interesting tree evolution process that we shortly describe for $d = 3$.

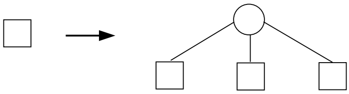


Fig. 1.10. Substitution in 3-ary recursive trees

We consider 3-ary trees and distinguish (as in the case of binary trees) between internal and external nodes. We define the size of the tree by the number of internal nodes. The evolution process starts with an empty tree, that is, with just an external node. The first step in the growth process is to replace this external node by an internal one with three successors that are external ones (see Figure 1.10). Then with probability 1/3 one of these three external nodes is selected and again replaced by an internal node with 3 successors. In this way one continues. In each step one of the external nodes is replaced (with equal probability) by an internal node with 3 successors.

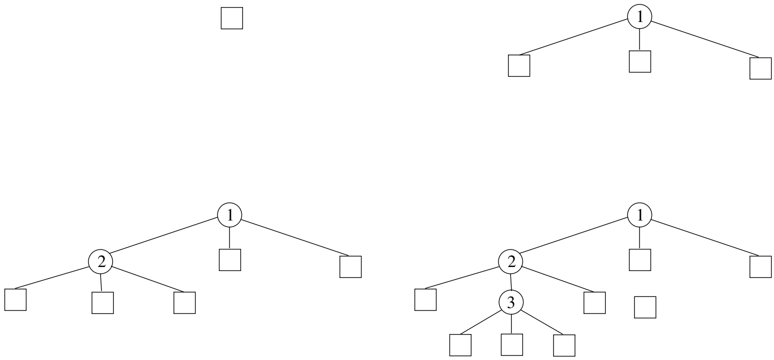


Fig. 1.11. Evolution process of 3-ary recursive trees

1.4 Search Trees

Search trees are used in computer science for storing and searching data. There are several concepts (compare with [146]). We just mention a few standard probabilistic models that are used to analyse these kinds of trees and the algorithms that are related with them.

1.4.1 Binary Search Trees

The origin of binary search trees dates to a fundamental problem in computer science: the dictionary problem. In this problem a set of records is given where each can be addressed by a key. The binary search tree is a data structure used for storing the records. Basic operations include *insert* and *search*.

Binary search trees are plane binary trees generated by a random permutation (or list) π of the numbers $\{1, 2, \ldots, n\}$. The elements of $\{1, 2, \ldots, n\}$ serve as keys. The keys are stored in the internal nodes of the tree. Starting with one of the keys (for example with $\pi(1)$) one first compares $\pi(1)$ with $\pi(2)$. If $\pi(2) < \pi(1)$, then $\pi(2)$ becomes root of the left subtree; otherwise, $\pi(2)$ becomes root of the right subtree. When having constructed a tree with nodes $\pi(1), \ldots, \pi(k)$, the next node $\pi(k+1)$ is inserted by comparison with the existing nodes in the following way: start with the root as current node. If $\pi(k+1)$ is less than the current node, then descend into the left subtree, otherwise into the right subtree. Now continue with the root of the chosen subtree as current, according to the same rule. Finally, attach $n+1$ external nodes (= leaves) at the possible places. Figure 1.12 shows an example of a binary search tree (without and with external nodes) for the input keys $(4, 6, 3, 5, 1, 8, 2, 7)$.

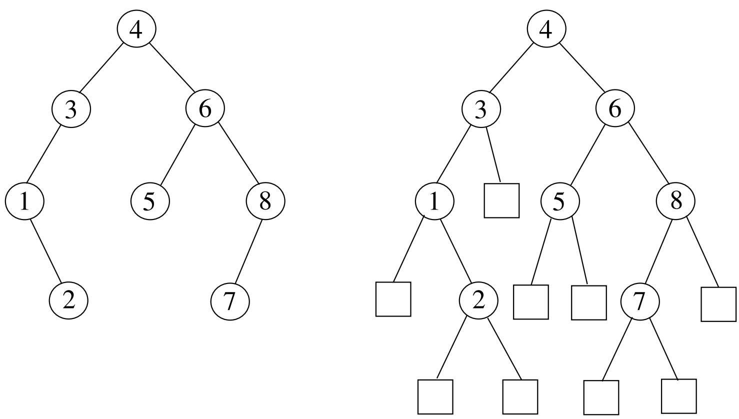


Fig. 1.12. Binary search tree

Alternatively one can describe the construction of the binary search tree recursively in the following way. If $n > 1$, we select (as above) a pivot (for example $\pi(1)$) and subdivide the remaining keys into two sublists I_1, I_2:

$$I_1 = (x \in \{\pi(2), \dots, \pi(n)\} : x < \pi(1)) \quad \text{and}$$
$$I_2 = (x \in \{\pi(2), \dots, \pi(n)\} : x > \pi(1)) .$$

The pivot $\pi(1)$ is put to the root and by recursively applying the same procedure, the elements of I_1 constitute the left subtree of the root and the elements of I_2 the right subtree. This is precisely the standard Quicksort algorithm.

At the moment there is no randomness involved. Every input sequence induces uniquely and deterministically a binary search tree. However, if we assume that the input data follow some probabilistic rule, then this induces a probability distribution on the corresponding binary search trees. The most common probabilistic model is the random permutation model, where one assumes that every permutation of the input data $1, 2, \dots, n$ is equally likely.

By assuming this standard probability model, there is, however, a completely different point of view to binary search trees, namely the tree evolution process of 2-ary recursive trees (compare with the description of 3-ary recursive trees in Section 1.3.3). Here one starts with the empty tree (just consisting of an external vertex). Then in a first step this external node is replaced by an internal one with two attached external nodes. In a second step one of these two external nodes is again replaced by an internal one with two attached external nodes. In this way one continues. In each step one of the existing external nodes is replaced by an internal one (plus two externals) with equal probability.

It is easy to explain that these two models actually produce the same kinds of random trees. Suppose that the keys $1, \dots, n$ are replaced by n real numbers $x_1, \dots, x_n$ that are ordered, that is, $x_1 < x_2 < \cdots < x_n$. Suppose that we have already constructed a binary search tree T according to some permutation π of $x_1, \dots, x_n$. The choice of an external node of T and replacing it by an internal one corresponds to the choice of one of the $n+1$ intervals $(-\infty, x_1)$, (x_1, x_2), …, (x_n, ∞) and choosing a number x^* of one of these intervals and working out the binary search tree algorithm to the list of $n+1$ elements, where x^* is appended to the list π (compare with Figure 1.12). However, this procedure also produces equally likely random permutations of $n+1$ elements from a random permutation of n elements.

1.4.2 Fringe Balanced m-Ary Search Trees

There are several generalisations of binary search trees. The search trees that we consider here, are characterised by two integer parameters $m \geq 2$ and $t \geq 0$. As binary search trees they are built from a set of n distinct keys taken from some totally ordered sets such as real numbers or integers. For our purposes we again assume that the keys are the integers $1, \dots, n$. The search tree is an m-ary tree where each node has at most m successors; moreover, each node stores one or several of the keys, up to at most $m-1$ keys in each node. The parameter t affects the structure of the trees; higher values of t

tend to make the tree more balanced. The special case $m = 2$ and $t = 0$ corresponds to binary search trees.

To describe the construction of the search tree, we begin with the simplest case $t = 0$. If $n = 0$, the tree is empty. If $1 \le n \le m-1$ the tree consists of a root only, with all keys stored in the root. If $n \ge m$ we select $m-1$ keys that are called *pivots*. The pivots are stored in the root. The $m-1$ pivots split the set of the remaining $n-m+1$ keys into m sublists $I_1, \ldots, I_m$: if the pivots are $x_1 < x_2 < \cdots < x_{m-1}$, then $I_1 := (x_i : x_i < x_1)$, $I_2 := (x_i : x_1 < x_i < x_2)$, $\ldots$, $I_m := (x_i : x_{m-1} < x_i)$. We then recursively construct a search tree for each of the sets I_i of keys (ignoring it if I_i is empty), and attach the roots of these trees as children of the root in the search tree from left to right.

In the case $t \ge 1$, the only difference is that the pivots are selected in a different way. We now select $mt + m - 1$ keys at random, order them as $y_1 < \cdots < y_{mt+m-1}$, and let the pivots be $y_{t+1}, y_{2(t+1)}, \ldots, y_{(m-1)(t+1)}$. In the case $m \le n < mt+m-1$, when this procedure is impossible, we select the pivots by some supplementary rule (depending only on the order properties of the keys). Usually one aims that the corresponding subtree that is generated here is as balanced as possible. This explains the notion *fringe balanced tree.* In particular, in the case $m = 2$, we let the pivot be the median of $2t + 1$ selected keys (when $n \ge 2t + 1$).

The standard probability model is again to assume that every permutation of the keys $1, \ldots, n$ is equally likely. The choice of the pivots can then be deterministic. For example, one always chooses the first $mt + m - 1$ keys. It is then easy to describe the splitting at the root of the tree by the random vector $\mathbf{V}_n = (V_{n,1}, V_{n,2}, \ldots, V_{n,m})$, where $V_{n,k} := |I_k|$ is the number of keys in the k-th subset, and thus the number of nodes in the k-th subtree of the root (including empty subtrees).

We thus always have, provided $n \ge m$,

$$V_{n,1} + V_{n,2} + \cdots + V_{n,m} = n - (m-1) = n + 1 - m$$

and elementary combinatorics, counting the number of possible choices of the $mt + m - 1$ selected keys, showing that the probability distribution is, for $n \ge mt + m - 1$ and $n_1 + n_2 + \cdots + n_m = n - m + 1$,

$$\mathbb{P}\{\mathbf{V}_n = (n_1, \ldots, n_m)\} = \frac{\binom{n_1}{t} \cdots \binom{n_m}{t}}{\binom{n}{mt+m-1}}. \tag{1.5}$$

(The distribution of $\mathbf{V}_n$ for $m \le n < mt + m - 1$ is not specified.)

In particular, for $n \ge mt + m - 1$, the components $V_{n,j}$ are identically distributed, and another simple counting argument yields, for $n \ge mt+m-1$ and $0 \le \ell \le n - 1$,

$$\mathbb{P}\{V_{n,j} = \ell\} = \frac{\binom{\ell}{t}\binom{n-\ell-1}{(m-1)t+m-2}}{\binom{n}{mt+m-1}}. \tag{1.6}$$

For usual binary search tree with $m = 2$ and $t = 0$ we have $V_{n,1}$ and $V_{n,2} = n - 1 - V_{n-1}$ which are uniformly distributed on $\{0, \ldots, n-1\}$.

1.4.3 Digital Search Trees

Digital search trees are intended for the same kind of problems as binary search trees. However, they are not constructed from the total order structure of the keys for the data stored in the internal nodes of the tree but from digital representations (or binary sequences) which serve as keys.

Consider a set of records, numbered from 1 to n and let $x_1, \ldots, x_n$ be binary sequences for each item (that represent the keys). We construct a binary tree – the digital search tree – from such a sequence as follows. First, the root is left empty, we can say that it stores the empty word.[4] Then the first item occupies the right or left child of the root depending whether its first symbol is 1 or 0. After having inserted the first k items, we insert item $k+1$: Choose the root as current node and look at the binary key x_{k+1}. If the first digit is 1, descend into the right subtree, otherwise into the left one. If the root of the subtree is occupied, continue by looking at the next digit of the key. This procedure terminates at the first unoccupied node where the $(k+1)$-st item is stored.

For example, consider the items

$$\begin{aligned} x_1 &= 110011\cdots \\ x_2 &= 100110\cdots \\ x_3 &= 010010\cdots \\ x_4 &= 101110\cdots \\ x_5 &= 000110\cdots \\ x_6 &= 010111\cdots \\ x_7 &= 000100\cdots \\ x_8 &= 100101\cdots . \end{aligned}$$

If we apply the above described procedure we end up with the binary tree depicted in Figure 1.13. As in the case of binary search trees we can append external nodes to make it a complete binary tree.

The standard probabilistic model – the Bernoulli model – is to assume that the keys $x_1, \ldots, x_n$ are binary sequences, where the digits 0 and 1 are drawn independently and identically distributed with probability p for 1 and probability $q = 1 - p$ for 0. The case $p = q = \frac{1}{2}$ is called symmetric.

There are several natural generalisations of this basic model. Instead of a binary alphabet one can use an m-ary one leading to m-ary digital search trees. One can also change the probabilistic model by using, for example, discrete Markov processes to generate the key sequences or so-called dynamic sources that are based on dynamical systems $T : [0,1] \to [0,1]$ (compare with [41, 206]).

[4] Sometimes the first item is stored to the root. The resulting tree is slightly different but (in a proper probabilistic model) both variants have the same asymptotic behaviour.

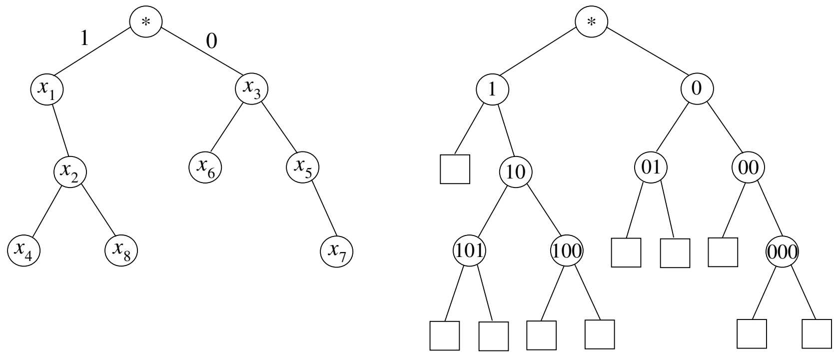


Fig. 1.13. Digital search tree

1.4.4 Tries

The construction idea of Tries[5] is similar to that of digital search trees except that the records are stored in the leaves rather than in the internal nodes. Again a 1 indicates a descent into the right subtree, and 0 indicates a descent into the left subtree. Insertion causes some rearrangement of the tree, since a leaf becomes an internal node. In contrast to binary search trees and digital search trees, the shape of the trie is independent of the actual order of insertion. The position of each item is determined by the shortest unique prefix of its key. If we use the same input data as for the example of a digital search tree, then we obtain the trie that is depicted in the left part of Figure 1.14.

An alternative description runs: Given a set $\mathcal{X}$ of strings, we partition $\mathcal{X}$ into two parts, $\mathcal{X}_L$ and $\mathcal{X}_R$, such that $x_j \in \mathcal{X}_L$ (respectively $x_j \in \mathcal{X}_R$) if the first symbol of x_j is 0 (respectively 1). The rest of the trie is defined recursively in the same way, except that the splitting at the k-th level depends on the k-th symbol of each string. The first time that a branch contains exactly one string, a leaf is placed in the trie at that location (denoting the placement of the string into the trie), and no further branching takes place from such a portion of the trie.

This description implies also a recursive definition of tries. As above consider a sequence of n binary strings. If $n = 0$, then the trie is empty. If $n = 1$, then a single (external) node holding this item is allocated. If $n \geq 1$, then the trie consists of a root (internal) node directing strings to the 2 subtrees according to the first letter of each string, and string directed to the same subtree are themselves tries, however, constructed from the second letter on.

Patricia tries are a slight modification of tries. Consider the case when several keys share the same prefix, but all other keys differ from this prefix already in their first position. Then the edges corresponding to this prefix may

---

[5] The notion *trie* was suggested by Fredkin [86] as it being part of re*trie*val.

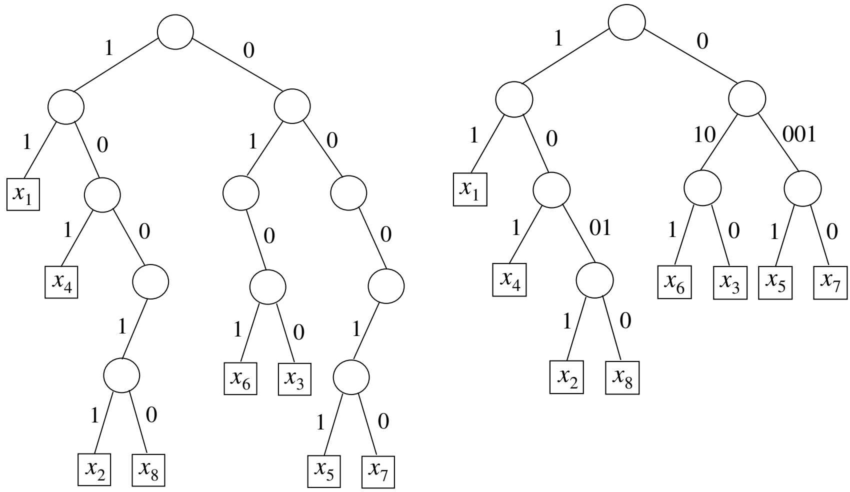


Fig. 1.14. Trie and Patricia trie

be contracted to one single edge. This method of construction leads to a more efficient structure (compare with Figure 1.14).

As in the case of digital search trees we can construct m-ary tries by using strings over an m-ary alphabet leading to m-ary trees.

Finally, if the input strings follow some probabilistic rule (coming, for example, from a Bernoulli or Markov source) then we obtain random tries and random Patricia tries.

2

Generating Functions

Generating functions are not only a useful tool to count combinatorial objects but also an analytic object that can be used to obtain asymptotics. They can be used to encode the distribution of random variables that are related to counting problems and, hence, asymptotic methods can be applied to obtain probabilistic limit theorems like central limit theorems.

In this chapter we survey some properties of generating functions following the mentioned *categories* above. First we collect some useful facts on generating functions in relation to counting problems, in particular, how certain combinatorial constructions have their counterparts in relations for generating functions. Next we provide a short introduction into singularity analysis of generating functions and its applications to asymptotics.

One major goal is to provide analytic and asymptotic properties of a generating function when it satisfies a functional equation and more generally when it is related to the solution of a system of functional equations. This situation occurs naturally in combinatorial problems with a recursive structure (as in tree counting problems) because a recursive relation usually translates into a functional equation for the corresponding generating function.

It turns out that solutions of functional equations typically have a finite radius of convergence R and – what is even more remarkable – that the kind of singularity at $x = R$ is of so-called square root type. This means that the generating function can be represented as a power series in $\sqrt{R-x}$. This explains that square root type singularities are omnipresent in the asymptotics of tree enumeration problems.

2.1 Counting with Generating Functions

Generating functions are quite natural in the context of tree counting since (rooted) trees have a recursive structure that usually translates to recurrence relations for corresponding counting problems. Besides generating functions are a proper tool for solving recurrence equations.

In order to give an idea how generating functions can be used to count trees we consider *binary trees*. Recall that binary trees are rooted trees, where each node is either a leaf or it has two distinguishable successors: the left successor and the right successor. The leaves of a binary tree are called *external* nodes and those nodes with two successors *internal* nodes. As already mentioned a binary tree with n internal nodes has $n + 1$ external nodes. Thus, the total number of nodes is always odd.

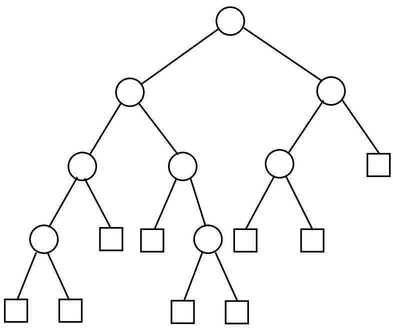


Fig. 2.1. Binary tree

We prove an explicit formula for the number of binary trees with the help of generating functions.

Theorem 2.1. *The number b_n of binary trees with n internal nodes is given by the Catalan number*

$$b_n = \frac{1}{n+1}\binom{2n}{n}.$$

Proof. Suppose that a binary tree has $n+1$ internal nodes. Then the left and right subtrees are also binary trees (with k and $n-k$ internal nodes, where $0 \le k \le n$). Thus, one gets directly the recurrence for the corresponding numbers:

$$b_{n+1} = \sum_{k=0}^{n} b_k b_{n-k}. \tag{2.1}$$

The initial value is $b_0 = 1$ (where the tree consists just of the root).

This recurrence can be solved using the generating function

$$b(x) = \sum_{n\geq 0} b_n x^n.$$

By (2.1) we find the relation

$$b(x) = 1 + xb(x)^2 \tag{2.2}$$

and consequently an explicit representation of the form

$$b(x) = \frac{1-\sqrt{1-4x}}{2x}. \tag{2.3}$$

Hence (by using the notation $[x^n]a(x) = a_n$ for the n-th coefficient of a power series $a(x) = \sum_{n\geq 0} a_n x^n$) we obtain

$$\begin{aligned} b_n &= [x^n]\frac{1-\sqrt{1-4x}}{2x} \\ &= -\frac{1}{2}[x^{n+1}](1-4x)^{\frac{1}{2}} \\ &= -\frac{1}{2}\binom{\frac{1}{2}}{n+1}(-4)^{n+1} \\ &= \frac{1}{n+1}\binom{2n}{n}. \end{aligned}$$

By inspecting the proof of Theorem 2.1 one observes that the recurrence relation (2.1) – together with its initial condition – is exactly a translation of a recursive description of binary trees (that was given in Section 1.2.1: a binary tree $\mathcal{B}$ is either just an external node or an internal node (the root) with two subtrees that are again binary trees.

It is also worth mentioning that the formal restatement of this recursive definition,

$$\mathcal{B} = \Box + \circ \times \mathcal{B} \times \mathcal{B} = \Box + \circ \times \mathcal{B}^2, \tag{2.4}$$

leads to a corresponding relation (2.2) for the generating function:

$$b(x) = 1 + xb(x)^2,$$

compare also with the schematic Figure 2.2.

2.1.1 Generating Functions and Combinatorial Constructions

We will now provide a more systematic treatment of combinatorial constructions and generating functions. The presentation is inspired by the work of Flajolet and his co-authors (see in particular the monograph [84] where this concept is described in much more detail).

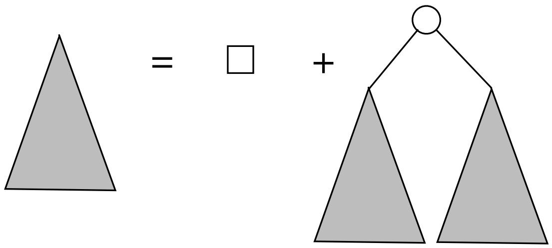


Fig. 2.2. Binary tree

Definition 2.2. *The ordinary generating function (ogf) of a sequence* $(a_n)_{n\geq 0}$ *(of complex numbers) is the formal power series*

$$a(x) = \sum_{n\geq 0} a_n x^n. \tag{2.5}$$

Similarly the exponential generating function (egf) of the sequence $(a_n)_{n\geq 0}$ *is given by*

$$\hat{a}(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}. \tag{2.6}$$

We use the notation

$$[x^n]a(x) = a_n$$

to *extract* the coefficient of x^n in a generating function.

It is clear that certain algebraic operations on the sequence a_n have their counterpart on the level of generating functions. The two tables in Figure 2.3 collect some of them.

A generating function $a(x)$ represents an analytic function for $|x| < R$, where

$$R = \left(\limsup_{n\to\infty} |a_n|^{\frac{1}{n}}\right)^{-1}$$

denotes the radius of convergence. Thus, if $R > 0$ then we can either use a differentiation to represent the sequence

$$a_n = \frac{a^{(n)}(0)}{n!},$$

or we use Cauchy's formula

$$a_n = \frac{1}{2\pi i}\int_\gamma a(x)\,\frac{dx}{x^{n+1}},$$

where γ is a closed curve inside the region of analyticity of $a(x)$ with winding number $+1$ around the origin.

	sequence	ogf
sum	$c_n = a_n + b_n$	$c(x) = a(x) + b(x)$
product	$c_n = \sum_{k=0}^{n} a_k b_{n-k}$	$c(x) = a(x)b(x)$
partial sums	$c_n = \sum_{k=0}^{n} a_k$	$c(x) = \frac{1}{1-x} a(x)$
marking	$c_n = n a_n$	$c(x) = x a'(x)$
scaling	$c_n = \gamma^n a_n$	$c(x) = a(\gamma x)$

	sequence	egf
sum	$c_n = a_n + b_n$	$\hat{c}(x) = \hat{a}(x) + \hat{b}(x)$
product	$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}$	$\hat{c}(x) = \hat{a}(x)\hat{b}(x)$
binomial sums	$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k$	$\hat{c}(x) = e^x \hat{a}(x)$
marking	$c_n = n a_n$	$\hat{c}(x) = x \hat{a}'(x)$
scaling	$c_n = \gamma^n a_n$	$\hat{c}(x) = \hat{a}(\gamma x)$

Fig. 2.3. Basic relations between sequences and their generating functions

Another point of view of generating functions is that they can be considered as a power series generated by certain *combinatorial objects.*

Let C be a (countable) set of objects, for example, a set of graphs and

$$|\cdot| : C \to \mathbb{N}$$

a *weight function* that assigns to every element $c \in C$ a weight or size $|c|$. We assume that the sets

$$C_n := |\cdot|^{-1}(\{n\}) = \{c \in C : |c| = n\} \qquad (n \in \mathbb{N})$$

are all finite. Set $c_n = |C_n|$. Then the ordinary generating function $c(x)$ of the pair $(C, |\cdot|)$ that we also call *combinatorial structure* is given by

$$c(x) = \sum_{c \in C} x^{|c|} = \sum_{n \geq 0} c_n x^n,$$

and the exponential generating function $\hat{c}(x)$ by

$$\hat{c}(x) = \sum_{c \in C} \frac{x^{|c|}}{|c|!} = \sum_{n \geq 0} c_n \frac{x^n}{n!}.$$

The choice of ordinary generating functions or exponential generating functions depends on the kind of problem. As a rule unlabelled (or unordered) structures should be counted with the help of ordinary generating functions and labelled (or ordered) structures with exponential generating functions.

Example 2.3 *The ogf of binary trees, where the weight is the number of internal nodes, is given by*

$$b(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

Example 2.4 *The egf of permutations of finite sets, where the weight is the size of the finite set, is given by*

$$\hat{p}(x) = \sum_{n \ge 0} n! \frac{x^n}{n!} = \frac{1}{1-x}.$$

One major aspect in the use of generating functions is that certain combinatorial constructions have their counterparts in relations of the corresponding generating functions. The recursive description of binary trees (2.4) is a nice example to support this statement.

We again use the notation $A, B, C, \ldots$ for sets of combinatorial objects with corresponding size functions $|\cdot|$. First suppose that the objects that we consider are (in some sense) unlabelled or unordered. We have the following basic operations:

1. If A and B are disjoint then $C = A + B = A \cup B$ denotes the union of A and B.
2. $C = A \times B$ denotes the Cartesian product. The size function of a pair $c = (a, b)$ is given by $|c| = |a| + |b|$.
3. Suppose that A contains no object of size 0 and that the sets A, $A \times A$, $A \times A \times A, \ldots$ are disjoint. Then
$$C = A^* := \{\varepsilon\} + A + A \times A + A \times A \times A + \cdots$$
is the set of (finite) sequences of elements of A (ε denotes the empty object of size zero).
4. Let $C = \mathcal{P}_{\mathrm{fin}}(A)$ denote the set of all finite subsets of A. The size of a subset $\{a_1, a_2, \ldots, a_k\} \subseteq A$ is given by $|\{a_1, a_2, \ldots, a_k\}| = |a_1| + |a_2| + \cdots + |a_k|$.
5. Let $C = \mathcal{M}_{\mathrm{fin}}(A)$ denote the set of all finite multisets $\{a_1^{j_1}, a_2^{j_2}, \ldots, a_k^{j_k}\}$ of A, that is, the element a_i is taken j_i times ($1 \le i \le k$). Its size is given by $|\{a_1^{j_1}, a_2^{j_2}, \ldots, a_k^{j_k}\}| = j_1|a_1| + j_2|a_2| + \cdots + j_k|a_k|$.
6. Suppose that B has no object of zero size. Then the composition $C = A(B)$ of A and B is given by
$$C = A(B) := A_0 + A_1 \times B + A_2 \times B \times B + \cdots,$$
where $A_n = \{a \in A \,|\, |a| = n\}$. The size of $(a, b_1, b_2, \ldots, b_n) \in A_n \times B \times B \times \cdots \times B$ is defined by $|b_1| + |b_2| + \cdots + |b_n|$. The combinatorial interpretation of this construction is that an element $a \in A$ of size $|a| = n$ is substituted by an n-tuple of elements of B.

As already indicated these combinatorial constructions have counterparts in relations of generating functions:

combinat. constr.	ogf
$C = A + B$	$c(x) = a(x) + b(x)$
$C = A \times B$	$c(x) = a(x)b(x)$
$C = A^*$	$c(x) = \frac{1}{1-a(x)}$
$C = \mathcal{P}_{\text{fin}}(A)$	$c(x) = e^{a(x) - \frac{1}{2}a(x^2) + \frac{1}{3}a(x^3) \mp \cdots}$
$C = \mathcal{M}_{\text{fin}}(A)$	$c(x) = e^{a(x) + \frac{1}{2}a(x^2) + \frac{1}{3}a(x^3) + \cdots}$
$C = A(B)$	$c(x) = a(b(x))$

We just indicate the proofs. For example, the ogf of the Cartesian product $C = A \times B$ is given by

$$c(x) = \sum_{(a,b) \in A \times B} x^{|(a,b)|} = \sum_{(a,b) \in A \times B} x^{|a|+|b|} = \sum_{a \in A} x^{|a|} \cdot \sum_{b \in B} x^{|b|} = a(x) \cdot b(x).$$

This implies that $1/(1 - a(x))$ represents the ogf of A^*.

The generating function of the finite subsets of A is given by

$$c(x) = \prod_{a \in A} \left(1 + x^{|a|}\right) = \prod_{n \geq 1} (1 + x^n)^{a_n}.$$

By taking logarithms and expanding the series $\log(1 + x^n)$ one obtains the proposed representation $c(x) = e^{a(x) - \frac{1}{2}a(x^2) + \frac{1}{3}a(x^3) \mp \cdots}$. Similarly the ogf of finite multisets is given by

$$c(x) = \prod_{a \in A} \left(\frac{1}{1 - x^{|a|}}\right) = \prod_{n \geq 1} \left(\frac{1}{1 - x^n}\right)^{a_n}.$$

There are similar constructions of so-called labelled or ordered combinatorial objects with corresponding relations for their exponential generating functions. We call a combinatorial object c of size n labelled if it is formally of the form $c = \tilde{c} \times \pi$, where $\pi \in \mathfrak{S}_n$ is a permutation. For example, we can think of a graph $\tilde{c}$ of n vertices and the permutation π represents a labelling of its nodes. We list some of the combinatorial constructions where we have to take care of the permutations involved.

1. If A and B are disjoint then $C = A + B = A \cup B$ denotes the union of A and B.
2. The *labelled product* $C = A * B$ of two labelled structures is defined as follows. Suppose that $a = \tilde{a} \times \pi \in A$ has size $|a| = k$ and $b = \tilde{b} \times \sigma \in B$ has size $|b| = m$. Then we define $a * b$ as the set of objects $((\tilde{a}, \tilde{b}), \tau)$, where $\tau \in \mathfrak{S}_{k+m}$ runs over all permutations that are consistent with π and σ in the following way: there is a partition $\{j_1, j_2, \ldots, j_k\}, \{\ell_1, \ell_2, \ldots, \ell_m\}$ of $\{1, 2, \ldots, k + m\}$ with $j_1 < j_2 < \cdots < j_k$ and $\ell_1 < \ell_2 < \cdots < \ell_m$ such that

$$\tau(1) = j_{\pi(1)},\ \tau(2) = j_{\pi(2)}, \ldots, \tau(k) = j_{\pi(k)} \quad \text{and}$$
$$\tau(k+1) = \ell_{\sigma(1)},\ \tau(k+2) = \ell_{\sigma(2)}, \ldots, \tau(k+m) = \ell_{\sigma(m)}.$$

Finally we set

$$A * B = \bigcup_{a\in A,\ b\in B} a * b.$$

The size of $((\tilde{a}, \tilde{b}), \tau)$ is given by $|((\tilde{a}, \tilde{b}), \tau)| = |a| + |b|$.

3. Suppose that A contains no object of size 0 and that the sets A, $A * A$, $A * A * A, \ldots$ are disjoint. Then

$$C = A^* := \{\varepsilon\} + A + A * A + A * A * + \cdots$$

is the set of (finite labelled) sequences of elements of A.

4. Similarly we define unordered labelled sequences by

$$C = e^A = \{\epsilon\} + A + \frac{1}{2!} A * A + \frac{1}{3!} A * A * A + \cdots,$$

where the short hand notation $\frac{1}{n!} A * A * \cdots * A$ means that we do not take care of the order of the n elements in the sequence $A * A * \cdots * A$.

5. Suppose that B has no object of size zero. Then the composition $C = A(B)$ of A and B is given by

$$C = A(B) := A_0 + A_1 \times B + A_2 \times B \times B + \cdots,$$

where $A_n = \{a \in A \,|\, |a| = n\}$.

The corresponding relations for the exponential generating functions are listed in the following table:

combinat. constr.	egf
$C = A + B$	$\hat{c}(x) = \hat{a}(x) + \hat{b}(x)$
$C = A * B$	$\hat{c}(x) = \hat{a}(x)\hat{b}(x)$
$C = A^*$	$\hat{c}(x) = \frac{1}{1-\hat{a}(x)}$
$C = e^A$	$\hat{c}(x) = e^{\hat{a}(x)}$
$C = A(B)$	$\hat{c}(x) = \hat{a}(\hat{b}(x))$

The only case that has to be explained is the labelled product $C = A * B$. If $|a| = k$ and $|b| = m$ then there are exactly $\binom{k+m}{k}$ possible ways to partition $\{1, 2, \ldots, k+m\}$ into two sets of size k and m. Thus $a * b$ has size $\binom{k+m}{k}$ and consequently

$$\hat{c}(x) = \sum_{a\in A, b\in B} \binom{|a| + |b|}{|a|} \frac{x^{|a|+|b|}}{(|a| + |b|)!} = \sum_{a\in A} \frac{x^{|a|}}{|a|!} \cdot \sum_{b\in B} \frac{x^{|b|}}{|b|!} = \hat{a}(x) \cdot \hat{b}(x).$$

2.1.2 Pólya's Theory of Counting

Counting problems on more involved unlabelled or unordered structures are usually much more difficult to handle than those on labelled structures. For example, it is possible to count labelled planar graphs (using generating functions), but the corresponding counting problem for unlabelled planar graphs has not been solved so far. The problem arises from symmetries which are difficult to deal with. Nevertheless, Pólya's theory of counting gives a method to handle such situations. We adopt the notation of [138].

Definition 2.5. *Let D be a finite set (of size n) and $\mathfrak{G} \le \mathfrak{S}_D$ a subgroup of the group $\mathfrak{S}_D$ of permutations of the set D. Then the cycle index $P_{\mathfrak{G}}(x_1, x_2, \ldots, x_n)$ of $\mathfrak{G}$ is defined by*

$$P_{\mathfrak{G}}(x_1, x_2, \ldots, x_n) := \frac{1}{|\mathfrak{G}|} \sum_{\pi \in \mathfrak{G}} x_1^{\lambda_1(\pi)} x_2^{\lambda_2(\pi)} \cdots x_n^{\lambda_n(\pi)},$$

where $\lambda_j(\pi)$ denotes the number of cycles of length j in the canonical cycle decomposition of the permutation π.

Example 2.6 *The cycle index of the trivial subgroup $\mathfrak{E}_n = \{\mathrm{id}\}$ of $\mathfrak{S}_n$ is given by*

$$P_{\mathfrak{E}_n}(x_1, x_2, \ldots, x_n) = x_1^n,$$

whereas the cycle index of $\mathfrak{S}_n$ has the following representation:

$$P_{\mathfrak{S}_n}(x_1, x_2, \ldots, x_n) \\ = \frac{1}{n!} \sum_{k_1 + 2k_2 + \cdots + nk_n = n} \frac{n!}{k_1! k_2! \cdots k_n! \cdot 1^{k_1} 2^{k_2} \cdots n^{k_n}} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}.$$

Note that the ogf of the cycle indices of $\mathfrak{S}_n$ has a nice representation:

$$\sum_{n \ge 0} P_{\mathfrak{S}_n}(x_1, x_2, \ldots, x_n) t^n = e^{tx_1 + \frac{1}{2}t^2 x_2 + \frac{1}{3}t^3 x_3 + \cdots}.$$

The term $\frac{1}{k} t^k x_k$ can be considered as the egf of the cycle index of a cyclic permutation of length k. Hence, the generating function

$$tx_1 + \frac{1}{2}t^2 x_2 + \frac{1}{3}t^3 x_3 + \cdots$$

describes all possible cyclic permutations. Since permutations can be seen as an unordered sequence of cyclic permutations, the generating function

$$e^{tx_1 + \frac{1}{2}t^2 x_2 + \frac{1}{3}t^3 x_3 + \cdots}$$

represents all cycle indices of permutations. Due to the factor $1/n!$ in the definition of the cycle index we finally have a ogf of the cycle indices of $\mathfrak{S}_n$.

Example 2.7 *The cycle index of the cyclic group $\mathfrak{C}_n$ (of order n, considered as a subgroup of $\mathfrak{S}_n$) is given by*

$$P_{\mathfrak{C}_n}(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{d|n} \varphi(d)\, x_d^{n/d},$$

where φ denotes Euler's totient function.

Also consider the set $M = R^D$, that is, the set of all functions $f : D \to R$. Let $\mathfrak{G}$ be a subgroup of $\mathfrak{S}_D$.

We call two functions $f, g \in M$ equivalent with respect to $\mathfrak{G}$ (and denote this by $f \sim g$) if there is a permutation $\pi \in \mathfrak{G}$ with $f(\pi(x)) = g(x)$ for all $x \in D$.

We consider a weight function $w : R \to W$ and define the weight of $f \in M = R^D$ by

$$w(f) = \prod_{x \in D} w(f(x)).$$

It is clear that $f \sim g$ implies $w(f) = w(g)$, that is, the weight $w(\mathbf{c})$ can be defined for an equivalence class $\mathbf{c} \in M/{\sim}$, too.

Pólya's theorem now gives an explicit representation of the sum of all weights of equivalence classes (a proof based on Burnside's lemma can be found in [138]).

Theorem 2.8. *Suppose that R and D are finite sets and let $\mathfrak{G}$ be a subgroup of $\mathfrak{S}_D$, and set $M = R^D$. Then we have*

$$\sum_{\mathbf{c} \in M/\sim} w(\mathbf{c}) = P_{\mathfrak{G}}\left(\sum_{r \in R} w(r), \sum_{r \in R} w(r)^2, \ldots, \sum_{r \in R} w(r)^{|D|} \right),$$

where this equality is interpreted in the ring $\mathbb{R}[W]$.

For example, it follows that the number of equivalence classes is given by $|M/{\sim}| = P_{\mathfrak{G}}(|R|, |R|, \ldots, |R|)$.

However, we now want to apply this concept to combinatorial constructions and generating functions. Let $(A, |\cdot|)$ be a combinatorial structure and let $k \geq 1$ be a fixed integer. Consider the set of k-tuples A^k, where the size $|(a_1, a_2, \ldots, a_k)|$ is defined by the sum of the sizes $|a_1| + |a_2| + \cdots + |a_k|$. Let $\mathfrak{G}$ be a subgroup of $\mathfrak{S}_k$ and define $(a_1, a_2, \ldots, a_k) \sim (a_1', a_2', \ldots, a_k')$ if there is $\pi \in \mathfrak{S}_k$ with $a_j = a'_{\pi(j)}$ $(1 \leq j \leq k)$. This is the same construction as above by setting $D = \{1, 2, \ldots, k\}$ and $R = A$ with the only difference that R is now a countable set. Furthermore, we define the weight function $w : A \to \mathbb{R}[x]$ by $w(a) = x^{|a|}$.

If we set $C = A^k/{\sim}$, then $(C, |\cdot|)$ is again a combinatorial structure and Pólya's theorem shows how the ordinary generating function $c(x)$ is related to $a(x)$.

Theorem 2.9. *Let* $(A, |\cdot|)$ *be a combinatorial structure and* $C = A^k/\sim$ *be defined as above. Then the ordinary generating function* $c(x)$ *is given by*

$$c(x) = P_{\mathfrak{G}}(a(x), a(x^2), \ldots, a(x^k)). \tag{2.7}$$

Proof. The proof is immediate. We only have to note that Theorem 2.8 formally extends to countable sets R. For example, we first apply Theorem 2.8 for $R_N = \{a \in A : |a| \le N\}$ and $M_N = R_N^k$. This leads to a corresponding formula

$$c_N(x) = P_{\mathfrak{G}}(a_N(x), a_N(x^2), \ldots, a_N(x^k)),$$

where $a_N(x)$ and $c_N(x)$ are the ogf's of A_N and $C_N = R_N^k/\sim$. Obviously we have $[x^n]a_N(x) = [x^n]a(x)$ and $[x^n]c_N(x) = [x^n]c(x)$ for all $n \le N$. Thus, we can formally let $N \to \infty$ and obtain (2.7).

In particular, we can apply this construction to the cyclic group $\mathfrak{C}_k$ and the symmetric group $\mathfrak{S}_k$:

$$c(x) = P_{\mathfrak{C}_k}(a(x), a(x^2), \ldots, a(x^k)), \tag{2.8}$$

or

$$c(x) = P_{\mathfrak{S}_k}(a(x), a(x^2), \ldots, a(x^k)). \tag{2.9}$$

If we use the cyclic group then C can be interpreted as the set of cycles of length k of elements of A and if we use the whole group $\mathfrak{S}_k$ then C can be seen as the set of multisets (of total size k) of elements of A. Note that the representation (2.9) can also be deduced from the multiset construction of unlabelled structures:

$$e^{a(x)+\frac{1}{2!}a(x^2)+\frac{1}{3!}a(x^3)+\cdots}.$$

If we introduce another variable t then the k-th coefficient

$$\begin{aligned}
&[t^k]e^{ta(x)+t^2\frac{1}{2!}a(x^2)+t^3\frac{1}{3!}a(x^3)+\cdots} \\
&\quad = \frac{1}{k!}\sum_{\ell_1+2\ell_2+\cdots+k\ell_k=k} \frac{k!}{\ell_1!\ell_2!\cdots\ell_k!\cdot 1^{\ell_1}2^{\ell_2}\cdots k^{\ell_k}}\, a(x)^{\ell_1}a(x^2)^{\ell_2}\cdots a(x^k)^{\ell_k} \\
&\quad = P_{\mathfrak{S}_k}(a(x), a(x^2), \ldots, a(x^k))
\end{aligned}$$

collects all multisets of size k.

It is also of interest to consider k-tuples $(a_1, a_2, \ldots, a_k)$ of different elements $a_1, a_2, \ldots, a_k \in A$. There is no general result here as for the unrestricted case. Nevertheless there are some particular cases that can be treated similarly. For example, if C denotes the set of k-tuples $(a_1, a_2, \ldots, a_k)$ with the property that all a_j are different we have

$$c(x) = a(x)^k - \sum_{\ell=2}^{k}(-1)^\ell \binom{k}{l} a(x)^{k-\ell}a(x^\ell).$$

This is just a direct application of the principle of inclusion and exclusion.

Finally, if C denotes the sets of size k of elements of A (that is, we consider k-tuples $(a_1, a_2, \ldots, a_k)$ with the property that all a_j are different and the full group $\mathfrak{S}_k$) we obtain

$$c(x) = P_{\mathfrak{S}_k}(a(x), -a(x^2), \ldots, (-1)^{k-1}a(x^k)).$$

Here we just have to observe that this generating function is given by the k-th coefficient

$$[t^k]e^{ta(x) - t^2\frac{1}{2!}a(x^2) + t^3\frac{1}{3!}a(x^3) \mp \cdots}.$$

2.1.3 Lagrange Inversion Formula

Let $a(x) = \sum_{n\ge 0} a_n x^n$ be a power series with $a_0 = 0$ and $a_1 \ne 0$. The Lagrange inversion formula provides an explicit representation of the coefficients of the inverse power series $a^{[-1]}(x)$ which is defined by $a(a^{[-1]}(x)) = a^{[-1]}(a(x)) = x$.

Theorem 2.10. *Let* $a(x) = \sum\limits_{n\ge 0} a_n x^n$ *be a formal power series with* $a_0 = 0$ *and* $a_1 \ne 0$. *Let* $a^{[-1]}(x)$ *be the inverse power series and* $g(x)$ *an arbitrary power series. Then the* n*-th coefficient of* $g(a^{[-1]}(x))$ *is given by*

$$[x^n]g(a^{[-1]}(x)) = \frac{1}{n}[u^{n-1}]g'(u)\left(\frac{u}{a(u)}\right)^n \qquad (n \ge 1).$$

In tree enumeration problems the following variant is more appropriate.

Theorem 2.11. *Let* $\Phi(x)$ *be a power series with* $\Phi(0) \ne 0$ *and* $y(x)$ *the (unique) power series solution of the equation*

$$y(x) = x\Phi(y(x)).$$

Then $y(x)$ *is invertible and the* n*-th coefficient of* $g(y(x))$ *(where* $g(x)$ *is an arbitrary power series) is given by*

$$[x^n]g(y(x)) = \frac{1}{n}[u^{n-1}]g'(u)\Phi(u)^n \qquad (n \ge 1).$$

Theorems 2.10 and 2.11 are equivalent. If $a(x) = x/\Phi(x)$ then $a^{[-1]}(x) = y(x)$, where $y(x)$ satisfies the equation $y(x) = x\Phi(y(x))$.

We give an immediate application of Theorem 2.11. We have already observed that the generating function $b(x)$ of binary trees satisfies the functional equation (2.2). If we set $\tilde{b}(x) = b(x) - 1$ then

$$\tilde{b}(x) = x(1 + \tilde{b}(x))^2.$$

By Theorem 2.11 with $\Phi(x) = (1+x)^2$ we obtain (for $n \ge 1$)

$$b_n = [x^n]\tilde{b}(x) = \frac{1}{n}[u^{n-1}](1+u)^{2n}$$
$$= \frac{1}{n}\binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}.$$

Proof. Since Theorems 2.10 and 2.11 are equivalent, we only have to prove Theorem 2.10. We just present an analytic proof for complex coefficients. The resulting analytic identities are formal ones, too.

We start with Cauchy's formula and use the substitution $u = a^{[-1]}(x)$. Note that if γ is a contour with winding number 1 around the origin then $\gamma' = a^{[-1]}(\gamma)$ has the same property:

$$[x^n]g(a^{[-1]}(x)) = \frac{1}{2\pi i}\int_\gamma \frac{g(a^{[-1]}(x))}{x^{n+1}}\,dx$$
$$= \frac{1}{2\pi i}\int_{\gamma'} g(u)\frac{a'(u)}{a(u)^{n+1}}\,du.$$

Since

$$\left(g(u)\frac{1}{a(u)^n}\right)' = g'(u)\frac{1}{a(u)^n} - n\,g(u)\frac{a'(u)}{a(u)^{n+1}},$$

it follows that

$$\int_{\gamma'} g(u)\frac{a'(u)}{a(u)^{n+1}}\,du = \frac{1}{n}\int_{\gamma'} g'(u)\frac{1}{a(u)^n}\,du,$$

and consequently

$$[x^n]g(a^{[-1]}(x)) = \frac{1}{n}\int_{\gamma'} g'(u)\frac{u^n}{a(u)^n}\,\frac{du}{u^n}$$
$$= \frac{1}{n}[u^{n-1}]g'(u)\left(\frac{u}{a(u)}\right)^n.$$

2.2 Asymptotics with Generating Functions

Generating functions $a(x) = \sum_{n\ge 0} a_n x^n$ can be considered, too, as power series in a complex variable x. If the radius of convergence is positive then one can apply Cauchy's formula

$$a_n = \frac{1}{2\pi i}\int_\gamma a(x)\,\frac{dx}{x^{n+1}},$$

where the contour γ is contained in the region of analyticity of $a(x)$ and encircles the $x = 0$ once. Thus, the analytic behaviour of $a(x)$ can be used to obtain information on the order of magnitude of a_n. In what follows we follow this idea by showing that certain kinds of singularities of $a(x)$ induce corresponding asymptotics for the coefficients a_n. This *singularity analysis* was introduced in a systematic way by Flajolet and Odlyzko [83] and will be the starting point of the following treatment. However, the main focus will be the analysis of functions $y(x)$ that satisfy a functional equation of the form $y(x) = F(x, y(x))$. We determine the kind of singularity (that will be of square root type) and extend this concept to a *combinatorial central limit theorem* (Theorem 2.23).

Note that Lagrange's inversion formula can be used to obtain explicit representations for the coefficients of the solution of the functional equation $y = x\Phi(y)$. The functional equation $y = F(x, y)$ mentioned above is therefore a natural generalisation that occurs frequently in tree enumeration problems, too.

2.2.1 Asymptotic Transfers

The basic property that allows an asymptotic transfer between the analytic behaviour of a generating function $a(x)$ and to its coefficients is the so-called *Transfer Lemma* by Flajolet and Odlyzko [83]. Note that an essential assumption is that $a(x)$ has an analytic continuation to a so-called Delta-domain Δ that is depicted in Figure 2.4.

Lemma 2.12. *Let*

$$a(x) = \sum_{n\ge 0} a_n x^n$$

be analytic in a region

$$\Delta = \Delta(x_0, \eta, \delta) = \{x : |x| < x_0 + \eta,\ |\arg(x/x_0 - 1)| > \delta\},$$

in which x_0 *and* η *are positive real numbers and* $0 < \delta < \pi/2$. *Furthermore suppose that there exists a real number* α *such that*

$$a(x) = O\left((1 - x/x_0)^{-\alpha}\right) \qquad (x \in \Delta).$$

Then

$$a_n = O\left(x_0^{-n} n^{\alpha-1}\right).$$

Proof. One uses Cauchy's formula

$$a_n = \frac{1}{2\pi i}\int_\gamma \frac{a(x)}{x^{n+1}}\,dx,$$

where γ is a suitable chosen path of integration around the origin. In particular, one can use $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, where (see also Figure 2.4)

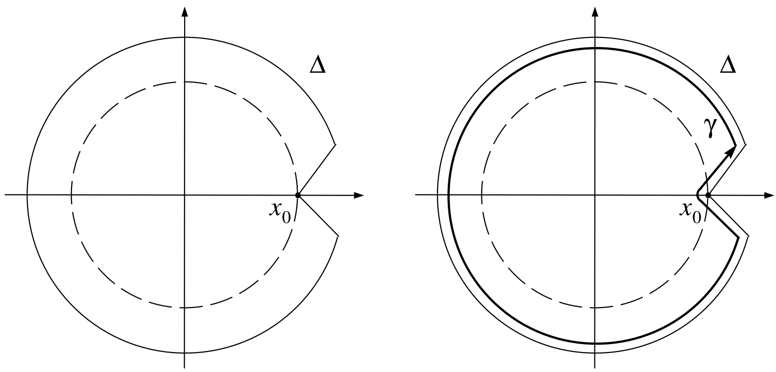


Fig. 2.4. Delta-domain and path of integration

$$\gamma_1 = \left\{ x = x_0 \left(1 + e^{-i\delta} \frac{-i + (\eta n - t)}{n} \right) : 0 \le t \le \eta n \right\},$$
$$\gamma_2 = \left\{ x = x_0 \left(1 - \frac{e^{i\varphi}}{n} \right) : -\frac{\pi}{2} + \delta \le \varphi \le \frac{\pi}{2} - \delta \right\},$$
$$\gamma_3 = \left\{ x = x_0 \left(1 + e^{i\delta} \frac{i + t}{n} \right) : 0 \le t \le \eta n \right\},$$

and γ_4 is a circular arc centred at the origin and making γ a closed curve.

It is easy to show that the bound $|a(x)| \le C|1 - x/x_0|^{-\alpha}$ implies

$$\left| \int_{\gamma_1 \cup \gamma_2 \cup \gamma_3} \frac{a(x)}{x^{n+1}}\, dx \right| \le C x_0^{-n} n^{\alpha-1} \cdot \left(\pi + 2 \int_0^\infty \frac{e^{-t\cos\delta}}{1 + t^\alpha}\, dt \right)$$
$$= O\left(x_0^{-n} n^{\alpha-1} \right)$$

whereas the integral over γ_4 is exponentially smaller:

$$\left| \int_{\gamma_4} \frac{a(x)}{x^{n+1}}\, dx \right| = O((x_0(1+\eta))^{-n}).$$

Remark 2.13 *A slight modification of the proof Lemma 2.12 also shows that*

$$a(x) = o\left((1 - x/x_0)^{-\alpha} \right) \qquad (x \to x_0,\, x \in \Delta)$$

implies

$$a_n = o\left(x_0^{-n} n^{\alpha-1} \right) \qquad (n \to \infty).$$

Next we discuss the case $a(x) = (1-x)^{-\alpha}$, where the n-th coefficient equals the binomial coefficient $(-1)^n \binom{-\alpha}{n}$.

Lemma 2.14. *For any compact set C in the complex plane $\mathbb{C}$ we have*

$$(-1)^n \binom{-\alpha}{n} = \frac{n^{\alpha-1}}{\Gamma(\alpha)} + O\left(n^{\Re\alpha-2}\right) \tag{2.10}$$

uniformly for $\alpha \in C$ *as* $n \to \infty$.

Proof. By Cauchy's formula we have

$$(-1)^n \binom{-\alpha}{n} = \frac{1}{2\pi i} \int_\gamma (1-x)^{-\alpha} x^{-n-1}\, dx,$$

where γ is a suitable closed curve around the origin. Note that the function $a(x) = (1-x)^{-\alpha}$ has an analytic continuation to $\mathbb{C} \setminus (1,\infty)$. Thus, we have more flexibility as in the proof of Lemma 2.12. We use $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, where

$$\gamma_1 = \left\{ x = 1 - \frac{i + \sqrt{n} - t}{n} : 0 \le t \le \sqrt{n} \right\},$$
$$\gamma_2 = \left\{ x = 1 - \frac{1}{n} e^{-i\varphi} : -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2} \right\},$$
$$\gamma_3 = \left\{ x = 1 + \frac{i+t}{n} : 0 \le t \le \sqrt{n} \right\},$$

and γ_4 is a circular arc centred at the origin and making γ a closed curve.

The easiest part is to estimate the integral over γ_4:

$$\left| \frac{1}{2\pi i} \int\limits_{\gamma_4} (1-x)^{-\alpha} x^{-n-1}\, dx \right| \le (1+n^{-\frac12})^{-n} \max\left(n^{\frac12 \Re\alpha}, (2+n^{-\frac12})^{-\Re\alpha} \right) e^{2\pi|\alpha|}.$$

On the remaining part $\gamma_1 \cup \gamma_2 \cup \gamma_3$ we use the Substitution $x = 1 + \frac{t}{n}$, where t varies on a corresponding curve $H_1 \cup H_2 \cup H_3$ that can be considered as a finite part of a so-called Hankel contour H (see Figure 2.5). The notion *Hankel contour* is adopted from the path of integration that is used for Hankel's integral representation for

$$\frac{1}{\Gamma(s)} = \frac{1}{2\pi i} \int_H (-t)^{-s} e^{-t}\, dt$$

that we use in a moment.

We approximate x^{-n-1} by $e^{-t}(1 + O(t^2/n))$. Now the integral over $\gamma_1 \cup \gamma_2 \cup \gamma_3$ is asymptotically given by

$$\begin{aligned}\frac{1}{2\pi i} \int_{\gamma_1\cup\gamma_2\cup\gamma_3} (1-x)^{-\alpha} x^{-n-1}\, dx &= \frac{n^{\alpha-1}}{2\pi i} \int_{H_1\cup H_2\cup H_3} (-t)^{-\alpha} e^{-t}\, dt \\ &\quad + \frac{n^{\alpha-2}}{2\pi i} \int_{H_1\cup H_2\cup H_3} (-t)^{-\alpha} e^{-t} \cdot \mathcal{O}(t^2)\, dt \\ &= n^{\alpha-1} I_1 + n^{\alpha-2} I_2.\end{aligned}$$

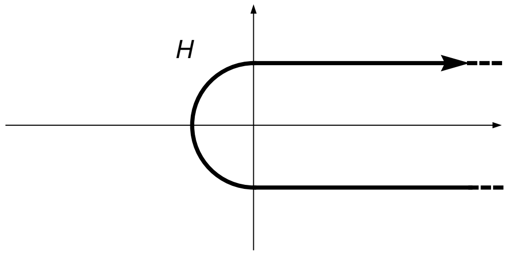


Fig. 2.5. Hankel contour of integration

Now I_1 approximates $1/\Gamma(\alpha)$ by

$$I_1 = \frac{1}{\Gamma(\alpha)} + O\left(\int_{\sqrt{n}}^{\infty} e^{2\pi|\alpha|}(1+t^2)^{-\frac{1}{2}\Re\alpha} e^{-t}\, dt\right)$$
$$= \frac{1}{\Gamma(\alpha)} + O\left(e^{2\pi|\alpha|}(1+n^2)^{-\frac{1}{2}\Re\alpha} e^{-\sqrt{n}}\right).$$

Finally, I_2 can be estimated by

$$I_2 = O\left(\int_0^{\infty} e^{2\pi|\alpha|}(1+t^2)^{1-\frac{1}{2}\Re\alpha}\, dt + O(1)\right).$$

Thus, for any compact set C in $\mathbb{C}$ we have

$$(-1)^n \binom{-\alpha}{n} = \frac{n^{\alpha-1}}{\Gamma(\alpha)} + \mathcal{O}(n^{\alpha-2})$$

uniformly for $\alpha \in C$ as $n \to \infty$.

This transfer lemma has two direct corollaries:

Corollary 2.15 *Suppose that a function is analytic in a region of the form Δ and that it has an expansion of the form*

$$a(x) = C\left(1 - \frac{x}{x_0}\right)^{-\alpha} + O\left(\left(1 - \frac{x}{x_0}\right)^{-\beta}\right) \qquad (x \in \Delta),$$

where $\beta < \Re(\alpha)$. Then we have

$$a_n = [x^n]a(x) = C\,\frac{n^{\alpha-1}}{\Gamma(\alpha)} x_0^{-n} + O\left(x_0^{-n} n^{\max\{\Re(\alpha)-2,\beta-1\}}\right). \qquad (2.11)$$

Proof. We just have to use Lemma 2.14 for the leading term and to apply Lemma 2.12 for the remainder term.

Corollary 2.16 *Suppose that $a(x)$ is analytic in a Delta-region Δ such that*

$$a(x) \sim C\left(1-\frac{x}{x_0}\right)^{-\alpha}$$

for $x \to x_0$ with $x \in \Delta$, where α is a complex number different from $\{0,-1,-2,\ldots\}$. Then, as $n \to \infty$

$$[x^n]a(x) \sim C\frac{n^{\alpha-1}}{\Gamma(\alpha)}x_0^{-n}.$$

Proof. Instead of Lemma 2.12 we now use the properties mentioned in Remark 2.13 for estimating the remainder term.

Remark 2.17 *Transfer principles like the one presented in Corollary 2.16 are not only valid for functions that behave like $C(1-x/x_0)^{-\alpha}$. For example one can add slowly varying factors (compare with [83]). In particular, if it is a purely logarithmic behaviour of the form*

$$a(x) \sim C\,\log\left(1-\frac{x}{x_0}\right)$$

for $x \in \Delta$ then we have

$$[x^n]a(x) \sim \frac{C}{n}\,x_0^{-n}.$$

Another important issue is that in fact all estimates in the proof of Lemma 2.12 are explicit and, thus, we obtain uniform estimates if we have an additional parameter.

Lemma 2.18. *Suppose that $a(x;w)$ is a power series in x and a parameter $w \in W$ such that there is an expansion of the form*

$$a(x;w) = C(w)\left(1-\frac{x}{x_0(w)}\right)^{-\alpha(w)} + O\left(\left(1-\frac{x}{x_0(w)}\right)^{-\beta(w)}\right),$$

that is uniform for $w \in W$ and $x \in \Delta(x_0(w))$ with functions $C(w)$, $x_0(w)$, $\alpha(w)$, and $\beta(w)$ that remain bounded and satisfy $\beta(w) < \Re(\alpha(w))$ for all $w \in W$. Then we get

$$[x^n]a(x;w) = C(w)\frac{n^{\alpha(w)-1}}{\Gamma(\alpha(w))}x_0(w)^{-n} + O\left(x_0(w)^{-n}n^{\max\{\Re(\alpha(w))-2,\beta(w)-1\}}\right) \tag{2.12}$$

uniformly for $w \in W$ as $n \to \infty$.

Proof. We just have to use two facts. First, the expansion (2.10) holds uniformly if α varies in some compact subset of the complex plane. Second, the estimates of Lemma 2.12 are explicit and also uniform.

2.2.2 Functional Equations

We already mentioned that functional equations of the form $y = F(x, y)$ occur naturally in tree enumeration problems. Informally they reflect the recursive nature of trees when one considers the subtrees of the root.

The next theorem shows that solutions of functional equations of that type (usually) have a square root singularity. The history of this theorem goes back to Bender [7], Canfield [31] and Meir and Moon [152]. It will have several applications within this book, in particular in Chapters 3, 4 and 9.

Theorem 2.19. *Suppose that $F(x, y)$ is an analytic function in x, y around $x = y = 0$ such that $F(0, y) = 0$ and that all Taylor coefficients of F around 0 are real and non-negative. Then there exists a unique analytic solution $y = y(x)$ of the functional equation*

$$y = F(x, y) \tag{2.13}$$

with $y(0) = 0$ that has non-negative Taylor coefficients around 0.

If the region of convergence of $F(x, y)$ is large enough such that there exist positive solutions $x = x_0$ and $y = y_0$ of the system of equations

$$\begin{aligned} y &= F(x, y), \\ 1 &= F_y(x, y) \end{aligned}$$

with $F_x(x_0, y_0) \neq 0$ and $F_{yy}(x_0, y_0) \neq 0$, then $y(x)$ is analytic for $|x| < x_0$ and there exist functions $g(x)$, $h(x)$ that are analytic around $x = x_0$ such that $y(x)$ has a representation of the form

$$y(x) = g(x) - h(x)\sqrt{1 - \frac{x}{x_0}} \tag{2.14}$$

locally around $x = x_0$. We have $g(x_0) = y(x_0)$ and

$$h(x_0) = \sqrt{\frac{2x_0 F_x(x_0, y_0)}{F_{yy}(x_0, y_0)}}.$$

Moreover, (2.14) provides a local analytic continuation of $y(x)$ (for $\arg(x - x_0) \neq 0$).

If we assume that $[x^n]\, y(x) > 0$ for $n \geq n_0$, then $x = x_0$ is the only singularity of $y(x)$ on the circle $|x| = x_0$ and we obtain an asymptotic expansion for $[x^n]\, y(x)$ of the form

$$[x^n]\, y(x) = \sqrt{\frac{x_0 F_x(x_0, y_0)}{2\pi F_{yy}(x_0, y_0)}}\, x_0^{-n}\, n^{-3/2} \left(1 + O(n^{-1})\right). \tag{2.15}$$

Note that the assumptions $F_x(x_0, y_0) \ne 0$ and $F_{yy}(x_0, y_0) \ne 0$ are really necessary to obtain a representation of the form (2.14). If $F_x(x, y) = 0$ then $F(x, y)$ (and $y(x)$) would not depend on x. If $F_{yy}(x, y) = 0$ then F is linear in y:

$$F(x, y) = yF_1(x) + F_2(x), \tag{2.16}$$

and consequently

$$y(x) = \frac{F_2(x)}{1 - F_1(x)} \tag{2.17}$$

is explicit and surely not of the form (2.14). However, a representation of the form (2.17) (where $F_1(x) \not\equiv 0$) usually leads to almost the asymptotic expansions for the coefficients of $y(x)$ in the case covered by Theorem 2.19. Suppose that the radius r of convergence of $F_1(x)$ is large enough that there is $0 < x_0 < r$ with $F_1(x_0) = 1$ and that $[x^n]y(x) > 0$ for $n \ge n_0$. Then x_0 is the only singularity on the circle of convergence $|x| = x_0$ of $y(x)$ and one gets

$$[x^n]y(x) = \frac{F_2(x_0)}{x_0 F'(x_0)} x_0^{-n} + O((x_0 + \eta)^{-n})$$

for some $\eta > 0$.

Remark 2.20 *Theorem 2.19 is designed for combinatorial applications, where the non-negativity of the coefficients is automatically satisfied. However, as we will see in the subsequent proof, in order to obtain a local representation of the form (2.14) it is sufficient to check analyticity of F and the conditions*

$$\begin{aligned} y_0 &= F(x_0, y_0), \\ 1 &= F_y(x_0, y_0), \\ 0 &\ne F_x(x_0, y_0), \\ 0 &\ne F_{yy}(x_0, y_0). \end{aligned}$$

Proof. Firstly we show that there exists a unique (analytic) solution $y = y(x)$ of $y = F(x, y)$ with $y(0) = 0$. Since $F(0, y) = 0$, it follows that the functional mapping

$$y(x) \mapsto F(x, y(x))$$

is a contraction for small x. Thus the iteratively defined functions $y_0(x) \equiv 0$ and

$$y_{m+1}(x) = F(x, y_m(x)) \qquad (n \ge 0)$$

converge uniformly to a limit function $y(x)$ which is the unique solution of (2.13). By definition it is clear that $y_m(x)$ is an analytic function around 0 and has real and non-negative Taylor coefficients. Consequently, the uniform limit $y(x)$ is analytic, too, with non-negative Taylor coefficients.

It is also possible to use the implicit function theorem. Since

$$F_y(0, 0) = 0 \ne 1,$$

there exists a solution $y = y(x)$ of (2.13) which is analytic around 0.

However, it is useful to know that all Taylor coefficients of $y(x)$ are non-negative. Namely, it follows that if $y(x)$ is regular at x' (which is real and positive) then $y(x)$ is regular for all x with $|x| \le x'$.

Let x_0 denote the radius of convergence of $y(x)$. Then x_0 is a singularity of $y(x)$. The mapping

$$x \mapsto F_y(x, y(x))$$

is strictly increasing for real and non-negative x as long as $y(x)$ is regular. Note that $F_y(0, y(0)) = 0$. As long as $F_y(x, y(x)) < 1$ it follows from the implicit function theorem that $y(x)$ is regular even in a neighbourhood of x. Hence there exists a finite limit point x_1 such that $\lim_{x \to x_1 -} y(x) = y_1$ is finite and satisfies $F_y(x_1, y_1) = 1$. If $y(x)$ was regular at $x = x_1$, then

$$y'(x_1) = F_x(x_1, y(x_1)) + F_y(x_1, y(x_1))y'(x_1)$$

would imply $F_x(x_1, y(x_1)) = 0$ which is surely not true. Thus, $y(x)$ is singular at $x = x_1$ which implies that $x_1 = x_0$. Moreover $y_0 = y(x_0) = y_1$ is finite.

Now let us consider the equation $y - F(x, y) = 0$ around $x = x_0$ and $y = y_0$. We have $1 - F_y(x_0, y_0) = 0$, but $-F_{yy}(x_0, y_0) \ne 0$. Hence by the Weierstrass preparation theorem[1] there exist functions $H(x, y)$, $p(x)$, $q(x)$ which are analytic around $x = x_0$ and $y = y_0$ and satisfy $H(x_0, y_0) \ne 1$, $p(x_0) = q(x_0) = 0$ and

$$y - F(x, y) = H(x, y)((y - y_0)^2 + p(x)(y - y_0) + q(x))$$

locally around $x = x_0$ and $y = y_0$. Since $F_x(x_0, y_0) \ne 0$, we also have $q_x(x_0) \ne 0$. This means that any analytic function $y = y(x)$ which satisfies $y(x) = F(x, y(x))$ in a subset of a neighbourhood of $x = x_0$ with x_0 on its boundary is given by

$$y(x) = y_0 - \frac{p(x)}{2} \pm \sqrt{\frac{p(x)^2}{4} - q(x)}.$$

Since $p(x_0) = 0$ and $q_x(x_0) \ne 0$, we have

$$\frac{\partial}{\partial x} \left(\frac{p(x)^2}{4} - q(x) \right)_{x = x_0} \ne 0,$$

too. Thus there exists an analytic function $K(x)$ such that $K(x_0) \ne 0$ and

$$\frac{p(x)^2}{4} - q(x) = K(x)(x - x_0)$$

[1] The Weierstrass preparation theorem (see [121] or [84, Theorem B.5]) says that every non-zero function $F(z_1, \ldots, z_d)$ with $F(0, \ldots, 0) = 0$ that is analytic at $(0, \ldots, 0)$ has a unique factorisation $F(z_1, \ldots, z_d) = K(z_1, \ldots, z_d)W(z_1; z_2, \ldots, z_d)$ into analytic factors, where $K(0, \ldots, 0) \ne 0$ and $W(z_1; z_2, \ldots, z_d) = z_1^d + z_1^{d-1} g_1(z_2, \ldots, z_d) + \cdots + g_d(z_2, \ldots, z_d)$ is a so-called Weierstrass polynomial, that is, all g_j are analytic and satisfy $g_j(0, \ldots, 0) = 0$.

locally around $x = x_0$. This finally leads to a local representation of $y = y(x)$ of the kind

$$y(x) = g(x) - h(x)\sqrt{1 - \frac{x}{x_0}}, \tag{2.18}$$

in which $g(x)$ and $h(x)$ are analytic around $x = x_0$ and satisfy $g(x_0) = y_0$ and $h(x_0) < 0$.

In order to calculate $h(x_0)$ we use Taylor's theorem

$$\begin{aligned} 0 &= F(x, y(x)) \\ &= F_x(x_0, y_0)(x - x_0) + \frac{1}{2}F_{yy}(x_0, y_0)(y(x) - y_0)^2 + \cdots \qquad (2.19) \\ &= F_x(x_0, y_0)(x - x_0) + \frac{1}{2}F_{yy}(x_0, y_0)h(x_0)^2(1 - x/x_0) + O(|x - x_0|^{3/2}). \end{aligned}$$

By comparing the coefficients of $(x - x_0)$ we immediately obtain

$$h(x_0) = \sqrt{\frac{2x_0F_x(x_0, y_0)}{F_{yy}(x_0, y_0)}}.$$

We now want to apply the transfer lemma (Lemma 2.12). For this purpose we have to show that $y(x)$ can be continued analytically to a region of the form Δ. The representation (2.18) provides such an analytic continuation for x in a neighbourhood of x_0. Now suppose that $|x_2| = x_0$ and $|\arg(x_2)| \ge \delta$. Then the assumption $y_n > 0$ for $n \ge n_0$ implies that $|y(x_2)| < y(|x_2|) = y(x_0)$ and consequently

$$|F_y(x_2, y(x_2))| \le F_y(|x_2|, |y(x_2)|) < F_y(|x_2|, y(|x_2|)) = F_y(x_0, y_0) = 1.$$

Thus, $F_y(x_2, y(x_2)) \ne 1$ and the implicit function theorem shows that there exists an analytic solution $y = y(x)$ in a neighbourhood of x_2. For $|x| < x_0$ this solution equals the power series $y(x)$ and for $|x| \ge x_0$ it provides an analytic continuation to a region of the form Δ (by compactness it is sufficient to consider finitely many x_2 with $|x_2| = x_0$ and $|\arg(x_2)| \ge \delta$). So finally we can apply Lemma 2.12 (or (2.11) with $\alpha = -1/2$ and $\beta = -3/2$; the analytic part of $g(x)$ provides exponentially smaller contributions.) This completes the proof of (2.15).

2.2.3 Asymptotic Normality and Functional Equations

We start with a slight extension of Theorem 2.19, where we add an additional k-dimensional parameter $\mathbf{u} = (u_1, \ldots, u_k)$ (compare also with [59]). This concept turns out to be useful for studying the distribution of tree parameters like the number of leaves. However, it has many other applications.

The reader might skip most parts of this section in a first reading, since it is quite technical, and should focus on Theorem 2.23 which is the main result of Section 2.2.3.

Theorem 2.21. *Suppose that* $F(x,y,\mathbf{u}) = \sum_{n,m} F_{n,m}(\mathbf{u})x^n y^m$ *is an analytic function in* x, y *around* 0 *and* $\mathbf{u}$ *around* $\mathbf{0}$ *such that* $F(0,y,\mathbf{u}) \equiv 0$, *that* $F(x,0,\mathbf{u}) \not\equiv 0$, *and that all coefficients* $F_{n,m}(1)$ *of* $F(x,y,1)$ *are real and non-negative. Then the unique solution* $y = y(x,\mathbf{u}) = \sum_n y_n(\mathbf{u})x^n$ *of the functional equation*

$$y = F(x,y,\mathbf{u}) \tag{2.20}$$

with $y(0,\mathbf{u}) = 0$ *is analytic around* $\mathbf{0}$ *and has non-negative coefficients* $y_n(\mathbf{1})$ *for* $\mathbf{u} = \mathbf{1} = (1,\ldots,1)$.

If the region of convergence of $F(x,y,\mathbf{u})$ *is large enough such that there exist non-negative solutions* $x = x_0$ *and* $y = y_0$ *of the system of equations*

$$\begin{aligned} y &= F(x,y,\mathbf{1}), \\ 1 &= F_y(x,y,\mathbf{1}) \end{aligned}$$

with $F_x(x_0,y_0,1) \neq 0$ *and* $F_{yy}(x_0,y_0,\mathbf{1}) \neq 0$ *then there exist functions* $f(\mathbf{u})$, $g(x,\mathbf{u})$, $h(x,\mathbf{u})$ *which are analytic around* $x = x_0$, $\mathbf{u} = \mathbf{1}$ *such that* $y(x,\mathbf{u})$ *is analytic for* $|x| < x_0$ *and* $|u_j - 1| \le \epsilon$ *(for some* $\epsilon > 0$ *and* $1 \le j \le k$*) and has a representation of the form*

$$y(x,\mathbf{u}) = g(x,\mathbf{u}) - h(x,\mathbf{u})\sqrt{1 - \frac{x}{f(\mathbf{u})}} \tag{2.21}$$

locally around $x = x_0$, $\mathbf{u} = \mathbf{1}$.

If $y_n(1) > 0$ *for* $n \ge n_0$, *then we also get*

$$y_n(\mathbf{u}) = \sqrt{\frac{f(\mathbf{u})F_x(f(\mathbf{u}),y(f(\mathbf{u}),\mathbf{u}),\mathbf{u})}{2\pi F_{yy}(f(\mathbf{u}),y(f(\mathbf{u}),\mathbf{u}),\mathbf{u})}}\, f(\mathbf{u})^{-n} n^{-3/2}\left(1 + O(n^{-1})\right) \tag{2.22}$$

uniformly for $|u_j - 1| < \epsilon$, $1 \le j \le k$.

Proof. The proof is completely the same as that of Theorem 2.19. We just have to be aware of the additional analytic parameter $\mathbf{u}$, in particular we have to apply Lemma 2.18 for $w = \mathbf{u}$.

Interestingly there is a strong relation to random variables that are asymptotically Gaussian. A random variable Z is called Gaussian or normal with law $N(\mu,\sigma^2)$, if its distribution function is of the form

$$\mathbb{P}\{Z \le x\} = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where μ is real, σ is positive and real, and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt.$$

We have $\mathbb{E}\,Z = \mu$ and $\mathbb{V}\mathrm{ar}\,Z = \sigma^2$. Equivalently, its characteristic function is given by

$$\mathbb{E}\, e^{itZ} = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}.$$

A random vector $\mathbf{Z} = (Z_1, \ldots, Z_k)$ is normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (with law $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$), if and only if all linear forms $\mathbf{a}^{\mathrm{T}}\mathbf{Z} = a_1 Z_1 + \cdots + a_k Z_k$ are normally distributed with mean $\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu}$ and variance $\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a}$.

The Gaussian distribution is important because it is a *universal law* that appears as a limiting distribution for sums of independent and weakly dependent random variables.

We say, that a sequence of random variables X_n satisfies *a central limit theorem* with (scaling) mean μ_n and (scaling) variance σ_n^2, if

$$\mathbb{P}\{X_n \le \mu_n + x \cdot \sigma_n\} = \Phi(x) + o(1)$$

as $n \to \infty$. In terms of *weak convergence*[2] this is equivalent to

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1). \tag{2.23}$$

For example, if X_n is the sum $\xi_1 + \xi_2 + \cdots + \xi_n$ of i.i.d. random variables ξ_j (with finite second moment) we have (2.23) with $\mu = \mathbb{E}\,\xi_j$ and $\sigma = \sqrt{\mathbb{V}\mathrm{ar}\,\xi_j}$. This can be checked easily by Lévy's criterion. Note that the characteristic function of X_n is just the n-th power of the characteristic function of ξ_j:

$$\mathbb{E}\, e^{itX_n} = \mathbb{E}\, e^{it(\xi_1+\xi_2+\cdots+\xi_n)} = \left(\mathbb{E}\, e^{it\xi_j}\right)^n.$$

We state (and prove) a multivariate version of the so-called *Quasi Power Theorem* by H.-K. Hwang [104] (see also [84], similar theorems can be found in [6, 10]), which is very useful to prove a central limit theorem. The idea is that if the characteristic function of a sequence of random variables X_n behaves almost like powers of a function, then the distribution of X_n should be approximated by a corresponding sum of i.i.d. random variables and, thus, one can expect a central limit theorem. Note that the so-called probability generating function $\mathbb{E}\, u^X$ is related to the characteristic function by setting $u = e^{it}$.

We also include simple tail estimates. (There exists a more precise large deviation result for the univariate case [105].)

Theorem 2.22. *Let $\mathbf{X}_n$ be a k-dimensional random vector with the property that*[3]

[2] A sequence of random variables X_n converges weakly to a random variable X, if $\lim_{n\to\infty} \mathbb{P}\{X_n \le x\} = \mathbb{P}\{X \le x\}$ holds for all points of continuity of $F_X(x) = \mathbb{P}\{X \le x\}$. This is denoted by $X_n \xrightarrow{d} X$. Another possible definition is that $\lim_{n\to\infty} \mathbb{E}\, F(X_n) = \mathbb{E}\, F(X_n)$ for all bounded continuous functions F. It is sufficient to check this condition for exponential functions $F(x) = e^{itx}$. Hence, $X_n \xrightarrow{d} X$, if and only if $\lim_{n\to\infty} \mathbb{E}\, e^{itX_n} = \mathbb{E}\, e^{itX}$ (for all real numbers t) which is Lévy's criterion. Corresponding properties hold for random vectors.

[3] If $\mathbf{u} = (u_1, \ldots, u_N)$ and $\mathbf{x} = (x_1, \ldots, x_N)$ are vectors of real (or complex) numbers then $\mathbf{u}^{\mathbf{x}}$ abbreviates $u_1^{x_1} u_2^{x_2} \cdots u_N^{x_N}$.

$$\mathbb{E}\,\mathbf{u}^{\mathbf{X}_n} = e^{\lambda_n \cdot A(\mathbf{u}) + B(\mathbf{u})} \left(1 + O\left(\frac{1}{\phi_n}\right)\right) \tag{2.24}$$

holds uniformly in a complex neighbourhood of $\mathbf{u} = \mathbf{1}$*, where* λ_n *and* ϕ_n *are sequences of positive real numbers with* $\lambda_n \to \infty$ *and* $\phi_n \to \infty$*, and* $A(\mathbf{u})$ *and* $B(\mathbf{u})$ *are analytic functions in this neighbourhood of* $\mathbf{u} = \mathbf{1}$ *with* $A(\mathbf{1}) = B(\mathbf{1}) = \mathbf{0}$*. Then* $\mathbf{X}_n$ *satisfies a central limit theorem of the form*

$$\frac{1}{\sqrt{\lambda_n}} \left(\mathbf{X}_n - \mathbb{E}\,\mathbf{X}_n\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}\right) \tag{2.25}$$

and we have

$$\mathbb{E}\,\mathbf{X}_n = \lambda_n \boldsymbol{\mu} + O\left(1 + \lambda_n/\phi_n\right)$$

and

$$\mathbb{C}\mathrm{ov}\,X_n = \lambda_n \mathbf{\Sigma} + O\left(\left(1 + \lambda_n/\phi_n\right)^2\right),$$

where

$$\boldsymbol{\mu} = A_{\mathbf{u}}(\mathbf{1}) = (A_{u_j}(\mathbf{1}))_{1 \le j \le k}$$

and

$$\mathbf{\Sigma} = (A_{u_i u_j}(\mathbf{1}) + \delta_{ij} A_{u_j}(\mathbf{1}))_{1 \le i,j \le k}.$$

Finally if we additionally assume that $\lambda_n = \phi_n$ *there exist positive constants* c_1, c_2, c_3 *such that*

$$\mathbb{P}\left\{\|\mathbf{X}_n - \mathbb{E}\,\mathbf{X}_n\| \ge \epsilon\sqrt{\lambda_n}\right\} \le c_1 e^{-c_2 \epsilon^2} \tag{2.26}$$

uniformly for $\epsilon \le c_3\sqrt{\lambda_n}$*.*

Proof. We first consider the univariate case $k = 1$, that is, we have

$$\mathbb{E}\,u^{X_n} = e^{\lambda_n \cdot a(u) + b(u)} \left(1 + O\left(\frac{1}{\phi_n}\right)\right). \tag{2.27}$$

By assumption, we obtain for t in a neighbourhood of $t = 0$

$$\mathbb{E}\,e^{itX_n} = e^{it\lambda_n\mu - \frac{1}{2}t^2\lambda_n\sigma^2 + O(\lambda_n t^3) + O(t)} \left(1 + O\left(\frac{1}{\phi_n}\right)\right),$$

where $\mu = a'(1)$ and $\sigma^2 = a'(1) + a''(1)$. Set $Y_n = (X_n - \lambda_n\mu)/\sqrt{\lambda_n}$. Then, replacing t by $t/\sqrt{\lambda_n}$, one gets

$$\mathbb{E}\,e^{itY_n} = e^{-\frac{\sigma^2}{2}t^2 + O(t^3/\sqrt{\lambda_n}) + O(t/\sqrt{\lambda_n})} \left(1 + O\left(\frac{1}{\phi_n}\right)\right).$$

Thus, Y_n is asymptotically normal with mean zero and variance σ^2.

Next set $f_n(u) = \mathbb{E}\,u^{X_n}$. Then $f_n'(1) = \mathbb{E}\,X_n$. On the other hand, by Cauchy's formula, we have

$$f_n'(1) = \frac{1}{2\pi i} \int_{|u-1|=\rho} \frac{f_n(u)}{(u-1)^2}\, du.$$

In particular, we use the circle $|u-1| = 1/\lambda_n$ as the path of integration and get

$$\begin{aligned}
&\mathbb{E}\, X_n = \\
&\frac{1}{2\pi i} \int_{|u-1|=1/\lambda_n} \frac{1 + (\lambda_n a'(1) + b'(1))(u-1) + O(\lambda_n (u-1)^2)}{(u-1)^2} \left(1 + O\left(\frac{1}{\phi_n}\right)\right) du \\
&\quad = \lambda_n a'(1) + O\left(1 + \frac{\lambda_n}{\phi_n}\right).
\end{aligned}$$

We can manage the variance similarly. Set $g_n(u) = f_n(u) u^{-\lambda_n a'(1) - b'(1)}$. Then $\mathbb{V}\mathrm{ar}\, X_n = g_n''(1) + O\left(1 + \lambda_n^2/\phi_n^2\right)$. By using the approximation

$$\begin{aligned}
&\exp\left(\lambda_n (a(u) - a'(1)\log u) + (b(u) - b'(1)\log u\right) \\
&\qquad = 1 + (\lambda_n (a''(1) + a'(1)) + (b''(1) + b'(1)))\frac{(u-1)^2}{2} \\
&\qquad + O(\lambda_n (u-1)^3) + O(\lambda_n^2 (u-1)^4)
\end{aligned}$$

and a similar calculation as above one obtains

$$g_n''(1) = \lambda_n (a''(1) + a'(1)) + O\left(1 + \frac{\lambda_n}{\phi_n}\right)$$

and consequently

$$\mathbb{V}\mathrm{ar}\, X_n = \lambda_n (a''(1) + a'(1)) + O\left((1 + \lambda_n/\phi_n)^2\right).$$

If $\sigma^2 > 0$, then Y_n/σ and $(X_n - \mathbb{E}\, X_n)/\sqrt{\mathbb{V}\mathrm{ar}\, X_n}$ have the same limiting distribution. Hence, the central limit theorem follows.

In order to obtain tail estimates we proceed as follows. Suppose that $\lambda_n = \phi_n$. Then we get (similarly as above)

$$\mathbb{E}\, e^{t(X_n - \mathbb{E}\, X_n)/\sqrt{\lambda_n}} = e^{\frac{\sigma^2}{2} t^2 + O(t^3/\sqrt{\lambda_n}) + O(t/\sqrt{\lambda_n})} \left(1 + O\left(\frac{1}{\lambda_n}\right)\right).$$

Hence, there exist positive constants c', c'', c''' with

$$\mathbb{E}\, e^{t(X_n - \mathbb{E}\, X_n)/\sqrt{\lambda_n}} \le c' e^{c'' t^2}$$

for real t with $|t| \le c''' \sqrt{\lambda_n}$. By a Chernov type argument we get for every $t > 0$ the inequality

$$\mathbb{P}\{|Y| \ge \epsilon\} \le \left(\mathbb{E}\, e^{tY} + \mathbb{E}\, e^{-tY}\right) e^{-\epsilon t}.$$

We also get (2.26) (for $k = 1$) by choosing $c_1 = 2c'$, $t = \epsilon/(2c'')$, $c_2 = 1/(4c'')$, and $c_3 = 2c''c'''$.

Now recall that a random vector $\mathbf{Y}$ is normally distributed with mean zero and covariance matrix $\mathbf{\Sigma}$, if and only if $\mathbf{a}^{\mathrm{T}}\mathbf{Y} = a_1 Y_1 + \cdots + a_k Y_k$ is normally distributed with mean zero and variance $\mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{a}$.

Hence, if we assume that a sequence of random vectors $\mathbf{X}_n$ satisfies (2.24) then the random variable $X_n(\mathbf{a}) = \mathbf{a}^{\mathrm{T}}\mathbf{X}_n$ satisfies (2.27) with $a(u) = A(u^{a_1}, \ldots, u^{a_k})$ and $b(u) = B(u^{a_1}, \ldots, u^{a_k})$. Consequently $X_n(\mathbf{a})$ is asymptotically normal with $\mathbb{E}\, X_n(\mathbf{a}) = \lambda_n \mu + O\left(1 + \lambda_n/\phi_n\right)$ and $\mathbf{Var}\, X_n(\mathbf{a}) = \lambda_n \sigma^2 + O\left(1 + \lambda_n/\phi_n\right)$, where

$$\mu = a'(1) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\mu} \quad \text{and} \quad \sigma^2 = a'(1) + a''(1) = \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}\,\mathbf{a}.$$

The tail estimate (2.26) can also be derived from the one-dimensional case $k = 1$.

If $\|\mathbf{X}_n - \mathbb{E}\,\mathbf{X}_n\| \ge \epsilon\sqrt{\lambda_n}$, then there exists j with $\|X_n^{(j)} - \mathbb{E}\, X_n^{(j)}\| \ge \epsilon\sqrt{\lambda_n}/\sqrt{k}$. Hence

$$\mathbb{P}\left\{\|\mathbf{X}_n - \mathbb{E}\,\mathbf{X}_n\| \ge \epsilon\sqrt{\lambda_n}\right\} \le k c_1 e^{-c_2\epsilon^2/k}.$$

Thus, we also get (2.26) in the multidimensional case. We just have to adjust c_1 and c_2.

By combining Theorems 2.21 and 2.22 we immediately obtain the following *combinatorial central limit theorem.* (For simplicity we just state a univariate version and discuss the multivariate version afterwards.)

Theorem 2.23. *Suppose that X_n is a sequence of random variables such that*

$$\mathbb{E}\, u^{X_n} = \frac{[x^n]\, y(x,u)}{[x^n]\, y(x,1)},$$

where $y(x,u)$ is a power series, that is the (analytic) solution of the functional equation $y = F(x,y,u)$, where $F(x,y,u)$ satisfies the assumptions of Theorem 2.21. In particular, let $x = x_0 > 0$ and $y = y_0 > 0$ be the (minimal) solution of the system of equations

$$\begin{aligned} y &= F(x,y,1), \\ 1 &= F_y(x,y,1) \end{aligned}$$

and set

$$\begin{aligned} \mu &= \frac{F_u}{x_0 F_x}, \\ \sigma^2 &= \mu + \mu^2 + \frac{1}{x_0 F_x^3 F_{yy}}\Big(F_x^2(F_{yy}F_{uu} - F_{yu}^2) - 2F_xF_u(F_{yy}F_{xu} - F_{yx}F_{yu}) \\ &\qquad + F_u^2(F_{yy}F_{xx} - F_{yx}^2)\Big), \end{aligned}$$

where all partial derivatives are evaluated at the point $(x_0, y_0, 1)$*. Then we have*

$$\mathbb{E}\, X_n = \mu n + O(1) \quad \textit{and} \quad \mathbb{V}\mathrm{ar}\, X_n = \sigma^2 n + O(1)$$

and if $\sigma^2 > 0$ *then*

$$\frac{X_n - \mathbb{E}\, X_n}{\sqrt{\mathbb{V}\mathrm{ar}\, X_n}} \xrightarrow{d} N(0,1).$$

Proof. By Theorem 2.21 we have uniformly for u in a complex neighbourhood of 1

$$\begin{aligned}\mathbb{E}\, u^{X_n} &= \frac{[x^n]\, y(x,u)}{[x^n]\, y(x,1)} \\ &= \sqrt{\frac{f(u)F_x(f(u), y(f(u),u),u)F_{yy}(x_0,y_0,1)}{F_{yy}(f(u),y(f(u),u),u)x_0F_x(x_0,y_0,1)}} \left(\frac{x_0}{f(u)}\right)^n \left(1 + O(n^{-1})\right).\end{aligned}$$

Thus, we can apply Theorem 2.22 with $A(u) = -\log(f(u)/x_0)$, where $f(u) = x(u)$ and $y(u)$ are the solutions of the system

$$\begin{aligned} y &= F(x,y,u), \\ 1 &= F_y(x,y,u). \end{aligned}$$

In particular the constants μ and σ^2 are given by

$$\mu = -\frac{x'(1)}{x_0} \quad \text{and} \quad \sigma^2 = \mu + \mu^2 - \frac{x''(1)}{x_0}.$$

By implicit differentiation one gets (with some algebra)

$$x'(1) = -\frac{F_u(x_0,y_0,1)}{F_x(x_0,y_0,1)} = -\frac{F_u}{F_x}$$

and

$$x''(1) = -\frac{1}{F_x}\left(F_{xx}x'(1)^2 + F_{xy}x'(1)y'(1) + 2F_{ux}x'(1) + F_{uy}y'(1) + F_{uu}\right),$$

where

$$y'(1) = -\frac{F_{xy}x'(1) + F_{uy}}{F_{yy}}.$$

The explicit representations for μ and σ^2 follow immediately.

Remark 2.24 *If we have several variables* $(u_1, \ldots, u_k) =: \mathbf{u}$ *and a sequence of random vectors* $\mathbf{X}_n$ *with*

$$\mathbb{E}\, \mathbf{u}^{\mathbf{X}_n} = \frac{[x^n]\, y(x,\mathbf{u})}{[x^n]\, y(x,\mathbf{1})},$$

where $y(x, \mathbf{u})$ is a power series, which is the solution of the function equation $y = F(x, y, \mathbf{u})$, then we get

$$\mathbb{E}\,\mathbf{X}_n = \boldsymbol{\mu}\, n + O(1) \quad \textit{and} \quad \mathbb{C}\mathrm{ov}\,\mathbf{X}_n = \boldsymbol{\Sigma}\, n + O(1),$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ and $\boldsymbol{\Sigma} = (\sigma_{ij})_{1 \le i,j \le k}$ can be calculated by

$$\begin{aligned}
\mu_i &= \frac{F_{u_i}}{x_0 F_x}, \\
\sigma_{ij} &= \mu_i \mu_j + \mu_i \delta_{i,j} \\
&\quad + \frac{1}{x_0 F_x^3 F_{yy}} \Big(F_x^2 (F_{yy} F_{u_i u_j} - F_{y u_i} F_{y u_j}) - F_x F_{u_i} (F_{yy} F_{x u_j} - F_{yx} F_{y u_j}) \\
&\qquad - F_x F_{u_j} (F_{yy} F_{x u_i} - F_{yx} F_{y u_i}) + F_{u_i} F_{u_j} (F_{yy} F_{xx} - F_{yx}^2) \Big),
\end{aligned}$$

and we also have a central limit theorem of the form

$$\frac{1}{\sqrt{n}} \left(\mathbf{X}_n - \mathbb{E}\,\mathbf{X}_n \right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right).$$

We finally state a useful variant of a central limit theorem for random variables which are defined with the help of generating functions.

Theorem 2.25. *Suppose that a sequence of k-dimensional random vectors $\mathbf{X}_n$ satisfies*

$$\mathbb{E}\,\mathbf{u}^{\mathbf{X}_n} = \frac{c_n(\mathbf{u})}{c_n(\mathbf{1})},$$

where $c_n(\mathbf{u})$ is the coefficient of x^n of an analytic function

$$f(x, \mathbf{u}) = \sum_{n \ge 0} c_n(\mathbf{u}) x^n$$

that has a local singular representation of the form

$$f(x, \mathbf{u}) = g(x, \mathbf{u}) + h(x, \mathbf{u}) \left(1 - \frac{x}{\rho(\mathbf{u})} \right)^{\alpha}$$

for some real $\alpha \in \mathbb{R} \setminus \mathbb{N}$ and functions $g(x, \mathbf{u})$, $h(x, \mathbf{u}) \ne 0$ and $\rho(\mathbf{u}) \ne 0$ that are analytic around $x = x_0 > 0$ and $\mathbf{u} = \mathbf{1}$. Suppose also that $x = \rho(\mathbf{u})$ is the only singularity of $f(x, u)$ on the disc $|x| \le |\rho(\mathbf{u})|$, if $\mathbf{u}$ is sufficiently close to $\mathbf{1}$ and that there exists an analytic continuation of $f(x, \mathbf{u})$ to the region $|x| < |\rho(\mathbf{u})| + \delta$, $|\arg(x - \rho(\mathbf{u}))| > \epsilon$ for some $\delta > 0$ and $\epsilon > 0$.

Then $\mathbf{X}_n$ satisfies a central limit theorem

$$\frac{1}{\sqrt{n}} \left(\mathbf{X}_n - \mathbb{E}\,\mathbf{X}_n \right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

with

$$\mathbb{E}\,\mathbf{X}_n = \boldsymbol{\mu} n + O(1) \quad \textit{and} \quad \mathbb{C}\mathrm{ov}\,\mathbf{X}_n = \boldsymbol{\Sigma} n + O(1),$$

where

$$\boldsymbol{\mu} = -\frac{\rho_{\mathbf{u}}(\mathbf{1})}{\rho(\mathbf{1})}$$

and

$$\boldsymbol{\Sigma} = -\frac{\rho_{\mathbf{uu}}(\mathbf{1})}{\rho(\mathbf{1})} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} + \mathrm{diag}(\boldsymbol{\mu}).$$

Furthermore there exist positive constants c_1, c_2, c_3 *such that*

$$\mathbb{P}\left\{\|\mathbf{X}_n - \mathbb{E}\,\mathbf{X}_n\| \ge \epsilon\sqrt{n}\right\} \le c_1 e^{-c_2\epsilon^2}$$

uniformly for $\epsilon \le c_3\sqrt{n}$.

Proof. By Lemma 2.18 we get the asymptotic expansion

$$c_n(\mathbf{u}) = \frac{h(\rho(\mathbf{u}), \mathbf{u})}{\Gamma(-\alpha)} n^{-\alpha-1} \rho(\mathbf{u})^{-n} \left(1 + O\left(\frac{1}{n}\right)\right)$$

that is uniform for $\mathbf{u}$ in a complex neighbourhood of $\mathbf{u} = \mathbf{1}$. Hence,

$$\begin{aligned}\mathbb{E}\,\mathbf{u}^{\mathbf{X}_n} &= \frac{c_n(\mathbf{u})}{c_n(\mathbf{1})} \\ &= \frac{h(\rho(\mathbf{u}), \mathbf{u})}{h(\rho(\mathbf{1}), \mathbf{1})} \left(\frac{\rho(\mathbf{1})}{\rho(\mathbf{u})}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right)\end{aligned}$$

and consequently the result follows from Theorem 2.22.

2.2.4 Transfer of Singularities

The main objective of this section is to consider analytic functions $f(x,u)$ that have a local representation[4]

$$f(x,u) = g(x,u) - h(x,u)\sqrt{1 - \frac{x}{\rho(u)}} \tag{2.28}$$

that holds in a (complex) neighbourhood $U \in \mathbb{C}^2$ of (x_0, u_0) with $x_0 \ne 0$, $u_0 \ne 0$ and with $\rho(u_0) = x_0$ (we only have to cut off the half lines $\{x \in \mathbb{C} : \arg(x - \rho(u)) = 0\}$ in order to have an unambiguous value of the square root).

We derive certain closure properties which will be mainly used in Chapter 9 for the asymptotic analysis of planar graphs. It might be skipped in a first reading.

The functions $g(x,u)$ and $h(x,u)$ are analytic in U and $\rho(u)$ is analytic in a neighbourhood of u_0. In our context we can usually assume that x_0 and

[4] The reason for the negative sign in front of $h(x,u)$ is that the coefficients of $\sqrt{1 - x/\rho(u)}$ are negative (if $\rho(u) > 0$) so that the coefficients of $f(x,u)$ in x are non-negative (provided that $h(x,u) > 0$, too).

u_0 are positive real numbers. In Section 2.2.2 we have shown that solutions $f(x,u)$ of functional equations (usually) have a local expansion of this form.

Note that a function $f(x,u)$ of the form (2.28) can also be represented as

$$f(x,u) = \sum_{\ell \ge 0} a_\ell(u) \left(1 - \frac{x}{\rho(u)}\right)^{\ell/2} = \sum_{\ell \ge 0} a_\ell(u) X^\ell, \qquad (2.29)$$

where $X = \sqrt{1 - x/\rho(u)}$. The analytic function $g(x,u)$ and $h(x,u)$ are given by

$$g(x,u) = \sum_{k \ge 0} a_{2k}(u) \left(1 - \frac{x}{\rho(u)}\right)^k = \sum_{k \ge 0} (-1)^k a_{2k}(u) \rho(u)^{-k} \big(x - \rho(u)\big)^k,$$

and

$$h(x,u) = \sum_{k \ge 0} a_{2k+1}(u) \left(1 - \frac{x}{\rho(u)}\right)^k = \sum_{k \ge 0} (-1)^k a_{2k+1}(u) \rho(u)^{-k} \big(x - \rho(u)\big)^k.$$

In particular, the coefficients $a_\ell(u)$ are analytic in u (for u close to u_0) and the power series

$$\sum_{\ell \ge 0} a_\ell(u) X^\ell$$

converges uniformly and absolutely, if u is close to u_0 and $|X| < r$ (for some properly chosen $r > 0$). It also represents an analytic function of u and X (in that range).

In the following passage we analyse some properties of functions that have a singular expansion of the form (2.28).

Lemma 2.26. *Suppose that $f(x,u)$ has a singular expansion of the form (2.28) and that $H(x,u,z)$ is a function that is analytic at $(x_0, u_0, f(x_0,u_0))$ such that*

$$H_z(x_0, u_0, f(x_0, u_0)) \ne 0.$$

Then

$$f_H(x,u) = H(x,u,f(x,u))$$

has the same kind of singular expansion, that is

$$f_H(x,u) = \overline{g}(x,u) - \overline{h}(x,u) \sqrt{1 - \frac{x}{\rho(u)}}$$

for certain analytic functions $\overline{g}(x,u)$ and $\overline{h}(x,u)$.

If $\rho(u) \ne 0$ and $f(x,u)$ has an analytic continuation to the region $|x| \le |\rho(u)| + \epsilon$, $\arg(x/\rho(u) - 1) \ne 0$ for some $\epsilon > 0$, and if $H(x,u,z)$ is also analytic for $|x| < x_0 + \epsilon$, $|u| < u_0 + \epsilon$ and $|z| < f(x_0,u_0) + \epsilon$, then $f(x,u)$ and $f_H(x,u)$ have power series expansions

$$f(x,u) = \sum_{n\ge 0} a_n(u)x^n \quad \textit{and} \quad f_H(x,u) = \sum_{n\ge 0} b_n(u)x^n,$$

where $a_n(u)$ *and* $b_n(u)$ *satisfy*

$$\lim_{n\to\infty} \frac{b_n(u)}{a_n(u)} = H_z(\rho(u), u, f(\rho(u))). \tag{2.30}$$

Proof. We use the Taylor series expansion of $H(x,u,z)$ at $z = g(x,u)$,

$$H(x,y,z) = \sum_{\ell=0}^{\infty} H_\ell(x,u)(z - g(x,u))^\ell,$$

and substitute $z = f(x,u)$:

$$\begin{aligned} f_H(x,u) &= H(x,y,f(x,u)) \\ &= \sum_{\ell=0}^{\infty} H_\ell(x,u) \left(-h(x,u)\sqrt{1-\frac{x}{\rho(u)}}\right)^\ell \\ &= \sum_{k=0}^{\infty} H_{2k}(x,u) \left(-h(x,u)\sqrt{1-\frac{x}{\rho(u)}}\right)^{2k} \\ &+ \sum_{k=0}^{\infty} H_{2k+1}(x,u) \left(-h(x,u)\sqrt{1-\frac{x}{\rho(u)}}\right)^{2k+1} \\ &= \overline{g}(x,u) - \overline{h}(x,u)\sqrt{1-\frac{x}{\rho(u)}}, \end{aligned}$$

where $\overline{g}(x,u)$ and $\overline{h}(x,u)$ are analytic at (x_0,u_0). (Note that all $H_\ell(x,u)$ are analytic in (x,u) and all appearing series are absolutely convergent.)

Finally we can use Lemma 2.18 in order to obtain asymptotic results for $a_n(u)$ and $b_n(u)$ which implies (2.30).

Lemma 2.27. *Suppose that* $f(x,u)$ *has a singular expansion of the form (2.28) such that* $|\rho(u)|$ *is the radius of convergence of the function* $x \mapsto f(x,u)$, *if* u *is sufficiently close to* u_0. *Then the partial derivative* $f_x(x,u)$ *and the integral* $\int_0^x f(t,u)\,dt$ *have local singular expansions of the form*

$$f_x(x,u) = \frac{g_2(x,u)}{\sqrt{1-\frac{x}{\rho(u)}}} + h_2(x,u) \tag{2.31}$$

and

$$\int_0^x f(t,u)\,dt = g_3(x,u) + h_3(x,u)\left(1-\frac{x}{\rho(u)}\right)^{3/2}, \tag{2.32}$$

where $g_2(x,u)$, $g_3(x,u)$, $h_2(x,u)$, *and* $h_3(x,u)$ *are analytic at* (x_0,u_0).[5]

[5] We decide the signs in front of $g_2(x,u)$ and $h_3(x,u)$ to be positive, since $1/\sqrt{1-x}$ and $(1-x)^{3/2}$ have positive coefficients (with one exception: the linear term of $(1-x)^{3/2}$).

Proof. First, from (2.28) one derives

$$\begin{aligned} f_x(x,u) &= g_x(x,u) - h_x(x,u)\sqrt{1-\frac{x}{\rho(u)}} + \frac{h(x,u)}{2\rho(u)\sqrt{1-\frac{x}{\rho(u)}}} \\ &= \frac{\frac{h(x,u)}{2\rho(u)} - h_x(x,u)\left(1-\frac{x}{\rho(u)}\right)}{\sqrt{1-\frac{x}{\rho(u)}}} + g_x(x,u) \\ &= \frac{g_2(x,u)}{\sqrt{1-\frac{x}{\rho(u)}}} + h_2(x,u). \end{aligned}$$

The proof of the representation of the integral is a bit more complicated. We represent $f(x,u)$ in the form

$$f(x,u) = \sum_{j=0}^{\infty} a_j(u)\left(1-\frac{x}{\rho(u)}\right)^{j/2}. \tag{2.33}$$

Recall that the power series

$$\sum_{j=0}^{\infty} a_j(u) X^{\ell}$$

converges absolutely and uniformly in a complex neighbourhood of u_0: $|u - u_0| \le r$ and for $|X| \le r$ (for some $r > 0$). Hence, there exists $\eta > 0$ such that $\eta|\rho(u)| < r$ for all u with $|u-u_0| \le r$. By assumption there are no singularities of $f(x,u)$ in the range $|x| \le |\rho(u)|(1-\eta)$.

We now assume that x is close to x_0 such that $|1 - x/\rho(u)| < r$. Then we split the integral $\int_0^x f(t,u)\,dt$ into three parts:

$$\begin{aligned} \int_0^x f(t,u)\,dt &= \int_0^{\rho(u)(1-\eta)} f(t,u)\,dt + \int_{\rho(u)(1-\eta)}^{\rho(u)} f(t,u)\,dt + \int_{\rho(u)}^{x} f(t,u)\,dt \\ &= I_1(u) + I_2(u) + I_3(x,u). \end{aligned}$$

Since η is properly chosen, there are no singularities of $f(t,u)$ in the range $|t| \le |\rho(u)|(1-\eta)$. Hence, $I_1(u)$ is an analytic function in u.

Next, by using the series representation (2.33) we obtain

$$\begin{aligned} I_2(u) &= \int_{\rho(u)(1-\eta)}^{\rho(u)} \sum_{j=0}^{\infty} a_j(u)\left(1-\frac{t}{\rho(u)}\right)^{j/2} dt \\ &= -\sum_{j=0}^{\infty} a_j(u)\frac{2\rho(u)}{j+1}\left(1-\frac{t}{\rho(u)}\right)^{\frac{j+2}{2}}\Bigg|_{t=\rho(u)(1-\eta)}^{t=\rho(u)} \\ &= \sum_{j=0}^{\infty} a_j(u)\frac{2\rho(u)}{j+1}\eta^{\frac{j+2}{2}}, \end{aligned}$$

which represents an analytic function, too.

Finally, the third integral evaluates to

$$\begin{aligned} I_3(x,u) &= \int_{\rho(u)}^{x} \sum_{j=0}^{\infty} a_j(u) \left(1 - \frac{t}{\rho(u)}\right)^{j/2} dt \\ &= -\sum_{j=0}^{\infty} a_j(u) \frac{2\rho(u)}{j+1} \left(1 - \frac{t}{\rho(u)}\right)^{\frac{j+2}{2}} \Bigg|_{t=\rho(u)}^{t=x} \\ &= -\sum_{j=0}^{\infty} a_j(u) \frac{2\rho(u)}{j+1} \left(1 - \frac{x}{\rho(u)}\right)^{\frac{j+2}{2}} . \end{aligned}$$

This can be represented as

$$I_3(x,u) = \tilde{g}(x,u) + \tilde{h}(x,u) \left(1 - \frac{x}{\rho(u)}\right)^{3/2}$$

with analytic functions $\tilde{g}(x,u)$ and $\tilde{h}(x,u)$. Putting these three representations together, we get (2.32).

Another important feature is that we can switch between local expansions in terms of x and u.

Lemma 2.28. *Suppose that $f(x,u)$ has a local representation of the form (2.28) such that*

$$\rho(u_0) \neq 0 \quad \textit{and} \quad \rho'(u_0) \neq 0.$$

Then the singular expansion (2.28) can be rewritten as

$$f(x,u) = \tilde{g}(x,u) - \tilde{h}(x,u) \sqrt{1 - \frac{u}{R(x)}},$$

where $R(x)$ is the (analytic) inverse function of $\rho(u)$.

Proof. Since $\rho'(u_0) \neq 0$, it follows from the Weierstrass preparation theorem (in our context we use a *shifted version* and apply it for $d = 1$ which actually is a refined version of the implicit function theorem) that there exists an analytic function $K(x,u)$ with $K(x_0,u_0) \neq 0$ such that

$$\rho(u) - x = K(x,u)(R(x) - u),$$

where $R(x)$ is the (analytic) inverse function of $\rho(u)$ in a neighbourhood of x_0. This is because near (x_0,u_0) we have $R(x) = u$, if and only if $\rho(u) = x$.

Consequently

$$1 - \frac{x}{\rho(u)} = K(x,u) \frac{R(x)}{\rho(u)} \left(1 - \frac{u}{R(x)}\right),$$

and hence

$$f(x,u) = g(x,u) - h(x,u)\sqrt{K(x,u)\frac{R(x)}{\rho(u)}}\sqrt{1-\frac{u}{R(x)}}$$
$$= \tilde{g}(x,u) - \tilde{h}(x,u)\sqrt{1-\frac{u}{R(x)}}.$$

Remark 2.29 *Note that Lemma 2.28 has some flexibility. For example, if $f(x,u)$ has a singular expansion of the form*

$$f(x,u) = g(x,u) + h(x,u)\left(1-\frac{x}{\rho(u)}\right)^{3/2},$$

then we also get a singular expansion of the form

$$f(x,u) = \tilde{g}(x,u) + \tilde{h}(x,u)\left(1-\frac{u}{R(x)}\right)^{\frac{3}{2}}.$$

Similarly, if $f(x,u)$ is of the form

$$f(x,u) = \frac{g_2(x,u)}{\sqrt{1-\frac{x}{\rho(u)}}} + h_2(x,u),$$

then we can rewrite this to

$$f(x,u) = \frac{\tilde{g}_2(x,u)}{\sqrt{1-\frac{u}{R(x)}}} + \tilde{h}_2(x,u).$$

If we combine Lemma 2.27 and Lemma 2.28 we thus get the following result.

Theorem 2.30. *Suppose that $f(x,u)$ has a singular expansion of the form (2.28) such that $|\rho(u)|$ is the radius of convergence of the function $x \mapsto f(x,u)$, if u is sufficiently close to u_0 and $\rho(u)$ satisfies $\rho(u_0) \neq 0$ and $\rho'(u_0) \neq 0$. Then the partial derivative $f_u(x,u)$ and the integral $\int_0^u f(x,t)\,dt$ have local singular expansions of the form*

$$f_u(x,u) = \frac{g_2(x,u)}{\sqrt{1-\frac{x}{\rho(u)}}} + h_2(x,u) \tag{2.34}$$

and

$$\int_0^u f(x,t)\,dt = g_3(x,u) + h_3(x,u)\left(1-\frac{x}{\rho(u)}\right)^{\frac{3}{2}}, \tag{2.35}$$

where $g_2(x,u)$, $g_3(x,u)$, $h_2(x,u)$, and $h_3(x,u)$ are analytic at (x_0,u_0).

Proof. For the proof of (2.34) and (2.35), we first apply Lemma 2.28 and switch to a singular expansion in terms of $\sqrt{1-u/R(x)}$. Then we apply Lemma 2.27 in order to get an expansion for the derivative or the integral, and finally we apply Lemma 2.28 again in order to get back to an expansion in terms of $\sqrt{1-x/\rho(u)}$.

We terminate this section by proving a variant of Theorem 2.21 where the right hand side of the equation $y = F(x,y,u)$ is not regular but has a square-root like singular behaviour of specific type. We also provide a variant of Lemma 2.26. These properties will be quite useful for enumerating planar graphs asymptotically (see Chapter 9).

Theorem 2.31. *Suppose that $F(x,y,u)$ has a local representation of the form*

$$F(x,y,u) = g(x,y,u) + h(x,y,u)\left(1 - \frac{y}{r(x,u)}\right)^{3/2} \qquad (2.36)$$

with functions $g(x,y,u)$, $h(x,y,u)$, $r(x,u)$ that are analytic around (x_0,y_0,u_0) and satisfy $g_y(x_0,y_0,u_0) \ne 1$, $h(x_0,y_0,u_0) \ne 0$, $r(x_0,u_0) \ne 0$ and $r_x(x_0,u_0) \ne g_x(x_0,y_0,u_0)$. Furthermore, suppose that $y = y(x,u)$ is a solution of the functional equation

$$y = F(x,y,u)$$

with $y(x_0,u_0) = y_0$. Then $y(x,u)$ has a local representation of the form

$$y(x,u) = g_1(x,u) + h_1(x,u)\left(1 - \frac{x}{\rho(u)}\right)^{3/2}, \qquad (2.37)$$

where $g_1(x,u)$, $h_1(x,u)$ and $\rho(u)$ are analytic at (x_0,u_0) and satisfy $h_1(x_0,u_0) \ne 0$ and $\rho(u_0) = x_0$.

Proof. Set $Y = (1 - y/r(x,u))^{1/2}$. Then $F(x,y,u)$ can be represented as

$$F(x,y,u) = A_0(x,u) + A_2(x,u)Y^2 + A_3(x,u)Y^3 + A_4(x,u)Y^4 + \cdots,$$

where $A_k(x,u)$ are analytic functions (compare with (2.29)). If we now consider the equation $y = F(x,y,u)$ and replace the left hand side by $y = r(x,u)(1-Y^2)$, we get

$$r(x,u) - A_0(x,u) = (r(x,u) + A_2(x,u))Y^2 + A_3(x,u)Y^3 + A_4(x,u)Y^4 + \cdots.$$

Since $r(x_0,u_0) = A_0(x_0,u_0) = g(x_0,y_0,u_0)$ and $r_x(x_0,u_0) \ne A_{0,x}(x_0,u_0) = g_x(x_0,y_0,u_0)$, we can again apply the Weierstrass preparation theorem. There exist analytic functions $K(x,u)$ and $\rho(u)$ with $K(x_0,u_0) \ne 0$ and $\rho(u_0) = x_0$, such that locally around (x_0,u_0)

$$r(x,u) - A_0(x,u) = K(x,u)(x - \rho(u)).$$

Hence, if we set $X = (1 - x/\rho(u))^{1/2}$ and $L(x,u) = (-K(x,u)\rho(u))^{1/2}$, we get

$$L(x,u)^2 X^2 = Y^2 \left(r(x,u) + A_2(x,u) + A_3(x,u)Y + A_4(x,u)Y^2 \cdots \right)$$

or

$$L(x,u)X = B_1(x,u)Y + B_2(x,u)Y^2 + B_3(x,u)Y^3 + \cdots ,$$

where $B_1(x,u) = \left(r(x,u) + A_2(x,u)\right)^{1/2}$ and $B_\ell(x,u)$ are suitably chosen analytic functions. Since $L(x,u) \neq 0$ and $B_1(x,u) \neq 0$ are in a neighbourhood of (x_0, u_0), we can invert this relation and get

$$Y = \frac{L(x,u)}{B_1(x,u)} X + C_2(x,u)X^2 + C_3(x,u)X^3 + \cdots .$$

By squaring this equation and substituting $Y^2 = 1 - y/r(x,u)$, we finally obtain the representation

$$1 - \frac{y}{r(x,u)} = \frac{L(x,u)^2}{B_1(x,u)^2} X^2 + D_3(x,u)X^3 + D_4(x,u)X^4 + \cdots$$

which can be rewritten in the form (2.37). Since $h_1(x_0,u_0) = r(x_0,u_0) \cdot D_3(x_0,u_0)$, we only need to check that $D_3(x_0,u_0) \neq 0$. But $D_3 = 2B_2L^2/B_1^2$ and $B_2 = A_3/(2\sqrt{r + A_2})$, and thus, it is sufficient to recall that $L(x_0,u_0) \neq 0$ and $A_3(x_0,u_0) \neq 0$.

Lemma 2.32. *Let $f(x) = \sum_{n\ge 0} a_n x^n$ denote the generating function of a sequence a_n of non-negative real numbers and suppose that $f(x)$ has exactly one dominant singularity at $x = \rho$ of the form*

$$f(x) = f_0 + f_2X^2 + f_3X^3 + \mathcal{O}(X^4),$$

where $X = \sqrt{1 - x/\rho}$ has an analytic continuation to the region $\{x \in \mathbb{C} : |x| < \rho + \epsilon\} \setminus \{x \in \mathbb{R} : x \ge \rho\}$ for some $\epsilon > 0$. Let $H(x,z,w)$ denote a function that has a dominant singularity at $z = f(\rho) > 0$ of the form

$$H(x,z,w) = h_0(x,w) + h_2(x,w)Z^2 + h_3(x,w)Z^3 + \mathcal{O}(Z^4),$$

where w is considered as a parameter, $Z = \sqrt{1 - z/f(\rho)}$, the functions $h_j(x,w)$ are analytic in x and $H(x,z,w)$ has an analytic continuation in a suitable region.

Then the function

$$f_H(x) = H(x, f(x), w)$$

has a power series expansion $f_H(x) = \sum_{n\ge 0} b_n(w)x^n$ and the coefficients b_n satisfy

$$\lim_{n\to\infty} \frac{b_n(w)}{a_n} = -\frac{h_2(\rho,w)}{f_0} + \frac{h_3(\rho,w)}{f_3}\left(-\frac{f_2}{f_0}\right)^{3/2} . \tag{2.38}$$

Proof. The proof is similar to that of Lemma 2.26 and is based on composing the singular expansion of $H(x,z,w)$ with that of $f(x)$. Indeed, taking into account that $f(\rho)=f_0$, we have near $x=\rho$

$$f_H(x) = h_0(x,w) + h_2(x,w)\left(-\frac{f_2X^2+f_3X^3}{f_0}\right) + h_3(x,w)\left(-\frac{f_2X^2+f_3X^3}{f_0}\right)^{3/2} + \cdots$$

Now note that $x=\rho-\rho X^2$. Thus, if we expand and extract the coefficient of X^3 and apply the transfer lemma, we have

$$a_n \sim \frac{f_3}{\Gamma(-3/2)} n^{-5/2}\rho^{-n}$$

and

$$b_n(w) \sim \frac{1}{\Gamma(-3/2)}\left(-\frac{h_2(\rho,w)f_3}{f_0} + h_3(\rho,w)\left(-\frac{f_2}{f_0}\right)^{3/2}\right) n^{-5/2}\rho^{-n},$$

so the result follows.

2.2.5 Systems of Functional Equations

In Section 2.1 we have discussed analytic solutions $y=y(x,u)$ of equations of the form $y=F(x,y,u)$. The main observation was that (under suitable assumptions on F) there is a local singular representation of $y(x,u)$ of the form

$$y(x,u) = g(x,u) - h(x,u)\sqrt{1-\frac{x}{x_0(u)}}. \qquad (2.39)$$

The purpose of this section is to present a generalisation to a system of equations. This concept will be applied in Chapter 3 for the analysis of pattern occurrences in trees (see Section 3.3) and in Chapter 9 for the analysis of number of vertices of given degree in some classed of planar graphs.

Let $\mathbf{F}(x,\mathbf{y},\mathbf{u}) = (F_1(x,\mathbf{y},\mathbf{u}),\ldots,F_N(x,\mathbf{y},\mathbf{u}))$ be a vector of functions $F_j(x,\mathbf{y},\mathbf{u})$, $1\le j\le N$ with complex variables x, $\mathbf{y}=(y_1,\ldots,y_N)$, $\mathbf{u}=(u_1,\ldots,u_k)$ which are analytic around 0 and satisfy $F_j(0,\mathbf{0},\mathbf{0})=0$ for $1\le j\le N$. We are interested in the analytic solution $\mathbf{y}=\mathbf{y}(x,\mathbf{u})=(y_1(x,\mathbf{u}),\ldots,y_N(x,\mathbf{u}))$ of the functional equation

$$\mathbf{y} = \mathbf{F}(x,\mathbf{y},\mathbf{u}) \qquad (2.40)$$

with $\mathbf{y}(0,\mathbf{0})=\mathbf{0}$, i.e., we demand that the (unknown) functions $y_j=y_j(x,\mathbf{u})$, $1\le j\le N$, satisfy the system of functional equations

$$\begin{aligned} y_1 &= F_1(x, y_1, y_2, \ldots, y_N, \mathbf{u}), \\ y_2 &= F_2(x, y_1, y_2, \ldots, y_N, \mathbf{u}), \\ &\vdots \\ y_N &= F_N(x, y_1, y_2, \ldots, y_N, \mathbf{u}). \end{aligned}$$

If the functions $F_j(x, \mathbf{y}, \mathbf{u})$ have non-negative Taylor coefficients then it is easy to see that the solutions $y_j(x, \mathbf{u})$ have the same property. (One only has to solve the system iteratively by setting $\mathbf{y}_0(x, \mathbf{u}) = 0$ and $\mathbf{y}_{i+1}(x, \mathbf{u}) = \mathbf{F}(x, \mathbf{y}_i(x, \mathbf{u}), \mathbf{u})$ for $i \ge 0$. The limit $\mathbf{y}(x, \mathbf{u}) = \lim_{i\to\infty} \mathbf{y}_i(x, \mathbf{u})$ is the (unique) solution of the system above.)

It is convenient to define the notion of a dependency (di)graph $G_{\mathbf{F}} = (V, E)$ for such a system of functional equations $\mathbf{y} = \mathbf{F}(x, \mathbf{y}, \mathbf{u})$. The vertices $V = \{y_1, y_2, \ldots, y_N\}$ are just the unknown functions and an ordered pair (y_i, y_j) is contained in the edge set E, if and only if $F_i(x, \mathbf{y}, \mathbf{u})$ really depends on y_j, that is, if the partial derivative $\frac{\partial F_i}{\partial y_j} \ne 0$.

In the following Theorem 2.33 we will use the assumption that the dependency graph is strongly connected, which means that every pair of vertices can be linked by a directed path in the graph. Informally this says that no subsystem can be solved prior to the whole system of equations. An equivalent condition is that the corresponding adjacency matrix and, thus, the Jacobian matrix $\mathbf{A} = \left(\frac{\partial F_i}{\partial y_j}\right)$ is irreducible, that is, there is no common reordering of the columns and rows of $\mathbf{A}$ such that the resulting matrix has the form (see [154])

$$\begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}.$$

The most important property of irreducible matrices $\mathbf{A}$ with non-negative entries is the Perron-Frobenius theorem (see [154]) saying that there is a unique positive and simple eigenvalue $r = r(\mathbf{A})$ with the property that all other eigenvalues λ satisfy $|\lambda| \le r$. This unique positive eigenvalue is a strictly increasing function of the entries of the non-negative matrix. More precisely if $\mathbf{A} = (a_{ij})$ and $\mathbf{A}' = (a'_{ij})$ are different irreducible non-negative matrices with $a_{ij} \le a'_{ij}$ (for all i, j) then $r(\mathbf{A}) < r(\mathbf{A}')$. Moreover, every principal submatrix has a smaller dominant eigenvalue.

Theorem 2.33. *Let* $\mathbf{F}(x, \mathbf{y}, \mathbf{u}) = (F_1(x, \mathbf{y}, \mathbf{u}), \ldots, F_N(x, \mathbf{y}, \mathbf{u}))$ *be a non-linear system of functions analytic around* $x = 0$, $\mathbf{y} = (y_1, \ldots, y_N) = \mathbf{0}$, $\mathbf{u} = (u_1, \ldots, u_k) = \mathbf{0}$, *whose Taylor coefficients are all non-negative, such that* $\mathbf{F}(0, \mathbf{y}, \mathbf{u}) = \mathbf{0}$, $\mathbf{F}(x, \mathbf{0}, \mathbf{u}) \ne \mathbf{0}$, $\mathbf{F}_x(x, \mathbf{y}, \mathbf{u}) \ne \mathbf{0}$. *Furthermore assume that the dependency graph of* $\mathbf{F}$ *is strongly connected and that the region of convergence of* $\mathbf{F}$ *is large enough that there exists a complex neighbourhood* U *of* $\mathbf{u} = \mathbf{1} = (1, \ldots, 1)$, *where the system*

$$\mathbf{y} = \mathbf{F}(x, \mathbf{y}, \mathbf{u}), \tag{2.41}$$

$$0 = \det(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x, \mathbf{y}, \mathbf{u})), \tag{2.42}$$

has solutions $x = x_0(\mathbf{u})$ *and* $\mathbf{y} = \mathbf{y}_0(\mathbf{u})$ *that are real, positive and minimal for positive real* $\mathbf{u} \in U$.

Let

$$\mathbf{y} = \mathbf{y}(x, \mathbf{u}) = (y_1(x, \mathbf{u}), \ldots, y_N(x, \mathbf{u}))$$

denote the analytic solutions of the system

$$\mathbf{y} = \mathbf{F}(x, \mathbf{y}, \mathbf{u}) \tag{2.43}$$

with $\mathbf{y}(0, \mathbf{u}) = \mathbf{0}$.

Then there exists $\epsilon > 0$ *such that* $y_j(x, \mathbf{u})$ *admit a representation of the form*

$$y_j(x, \mathbf{u}) = g_j(x, \mathbf{u}) - h_j(x, \mathbf{u})\sqrt{1 - \frac{x}{x_0(\mathbf{u})}} \tag{2.44}$$

for $\mathbf{u} \in U$, $|x - x_0(\mathbf{u})| < \epsilon$ *and* $|\arg(x - x_0(\mathbf{u})| \neq 0$, *where* $g_j(x, \mathbf{u}) \neq 0$ *and* $h_j(x, \mathbf{u}) \neq 0$ *are analytic functions with* $(g_j(x_0(\mathbf{u}), \mathbf{u}))_j = (y_j(x_0(\mathbf{u}), \mathbf{u}))_j = \mathbf{y}_0(\mathbf{u})$.

Furthermore, if $[x^n]\, y_j(x, \mathbf{1}) > 0$ *for* $1 \leq j \leq N$ *and for sufficiently large* $n \geq n_1$, *then there exists* $0 < \delta < \epsilon$ *such that* $y_j(x, \mathbf{u})$ *is analytic in* $(x, \mathbf{u})$ *for* $\mathbf{u} \in U$ *and* $|x - x_0(\mathbf{u})| \geq \epsilon$ *but* $|x| \leq |x_0(\mathbf{u})| + \delta$ *(this condition guarantees that* $\mathbf{y}(x, \mathbf{u})$ *has a unique smallest singularity with* $|x| = |x_0(\mathbf{u})|$*).*

Proof. In order to simplify the proof we first assume that $\mathbf{u}$ is real, positive and fixed, and we also suppress the dependency on $\mathbf{u}$ in the notation. Thus, we will work with x_0 and $\mathbf{y}_0$ that satisfy $\mathbf{y}_0 = \mathbf{F}(x_0, \mathbf{y}_0)$ and $\det(\mathbf{I} - \mathbf{A}) = 0$, where $\mathbf{A} = \mathbf{F}_\mathbf{y}(x_0, \mathbf{y}_0)$ abbreviates the Jacobian matrix of $\mathbf{F}$ with respect to $\mathbf{y}$. Note that the condition $\det(\mathbf{I} - \mathbf{A}) = 0$ says that 1 is eigenvalue of $\mathbf{A}$. If the Jacobian matrix $\mathbf{A}$ had no eigenvalue 1, then the implicit function theorem could be applied and it would follow that the system of functional equations $\mathbf{y} = \mathbf{F}(x, \mathbf{y})$ has locally an analytic solution $\mathbf{y} = \mathbf{y}(x)$ with $\mathbf{y}(x_0) = \mathbf{y}_0$.

Recall that the assumption that the dependency graph is strongly connected can be rephrased in $\mathbf{A}$ is a non-negative irreducible matrix. Hence, there is a unique (dominant) positive real simple eigenvalue. It is easy to observe that this dominating eigenvalue is actually 1. Set $\mathbf{A}(x) = \mathbf{F}_\mathbf{y}(x, \mathbf{y}(x))$ for $0 \leq x \leq x_0$. Then by assumption $\mathbf{A}(0) = \mathbf{0}$ and $\mathbf{A}(x)$ is (again) an irreducible positive matrix for $x > 0$ with the property that all non-zero entries of $\mathbf{A}(x)$ are strictly increasing in x. Thus, by the properties of irreducible non-negative matrices the dominant positive eigenvalue of $\mathbf{A}(x)$ is an increasing function, too. Since $(x_0, \mathbf{y}_0)$ is the minimal solution of (2.41) and (2.42), the eigenvalue 1 appears first for $x = x_0$.

Next we divide the system into the first equation and into the system of the remaining equations:

$$y_1 = F_1(x, y_1, \overline{\mathbf{y}}), \tag{2.45}$$

$$\overline{\mathbf{y}} = \overline{\mathbf{F}}(x, y_1, \overline{\mathbf{y}}), \tag{2.46}$$

where $\overline{\mathbf{y}} = (y_2, \ldots, y_N)$ and $\overline{\mathbf{F}} = (F_2, \ldots, F_N)$. Observe that the Jacobian matrix $\overline{\mathbf{F}}_{\overline{\mathbf{y}}}(x_0, \mathbf{y}_0)$ of $\overline{F}$ can be also obtained by deleting the first row and column of $\mathbf{A}$. Hence, all eigenvalues of $\mathbf{B}$ are strictly smaller than 1. Consequently, the matrix $\mathbf{I} - \mathbf{B}$ is invertible and, thus, by the implicit function theorem, there is a local solution $\overline{\mathbf{y}} = \overline{\mathbf{y}}(x, y_1)$ of (2.46), where y_1 is considered as an additional variable. As in the proof of Theorem 2.19 it follows that this solution is precisely the unique solution of (2.46) that has a power series expansion in x and y_1 at 0. Note, too, that due to the positivity assumptions on the coefficients of $\mathbf{F}$ this solution has non-negative coefficients as a power series in x and y_1.

We now insert this function into the first equation (2.45) and obtain a single equation

$$y_1 = F_1(x, y_1, \overline{\mathbf{y}}(x, y_1)) \tag{2.47}$$

for $y_1 = y_1(x)$. Again, the coefficients of

$$G(x, y_1) = F_1(x, y_1, \overline{\mathbf{y}}(x, y_1))$$

as a power series in x and y_1 are non-negative. At this point we can safely apply the methods of Theorem 2.19 (and 2.21). Actually, equation (2.47) is an equation of the form $y_1 = G(x, y_1)$ with the property that $G_{y_1}(x_0, y_{1,0}) = 1$. Thus, it follows from Theorem 2.19 that $y_1(x)$ has a local representation of the form $y_1(x) = g(x) - h(x)\sqrt{1 - x/x_0}$. Consequently there are corresponding local representations for $(y_1(x), \ldots, y_N(x)) = \mathbf{y}_2(x, y_1(x))$. Of course, x_0 is the common radius of convergence of $y_1(x), \ldots, y_N(x)$.

This proves (2.44) for fixed real and positive $\mathbf{u}$. However, it is easy to adapt the above proof for $\mathbf{u}$ close to the real axis. In particular, by continuity it follows that 1 will stay a simple eigenvalue of the Jacobian matrix.

Furthermore, if $[x^n]y_1(x, \mathbf{1}) > 0$ for sufficiently large $n \ge n_1$, then as in the proof of Theorem 2.19 it follows that x_0 is the only singularity on the circle of convergence of $y_1(x)$, and there is an analytic continuation to $|x| \le x_0 + \delta$ and $|x - x_0| \ge \epsilon$. This also applies for y_j for $2 \le j \le N$, and finally by continuity this is also true for $\mathbf{u}$ sufficiently close to the positive real axis.

With the help of a variation of Lemma 2.26 we derive the following

Corollary 2.34 *Let $\mathbf{y} = \mathbf{y}(x, \mathbf{u}) = (y_1(x, \mathbf{u}), \ldots, y_N(x, \mathbf{u}))$ be the solution of the system of equations (2.43) and assume that all assumptions of Theorem 2.33 are satisfied. Suppose that $G(x, \mathbf{y}, \mathbf{u})$ is a power series such that the point $(x_0(\mathbf{1}), \mathbf{y}_0(x_0(\mathbf{1}), \mathbf{1}), \mathbf{1})$ is contained in the interior of the region of convergence of $G(x, \mathbf{y}, \mathbf{u})$ and that $G_{\mathbf{y}}(x_0(\mathbf{1}), \mathbf{y}_0(\mathbf{1}), \mathbf{1}) \ne \mathbf{0}$.*

Then $G(x, \mathbf{y}(x, \mathbf{u}), \mathbf{u})$ has a representation of the form

$$G(x, \mathbf{y}(x, \mathbf{u}), \mathbf{u}) = g(x, \mathbf{u}) - h(x, \mathbf{u})\sqrt{1 - \frac{x}{x_0(\mathbf{u})}} \tag{2.48}$$

for $\mathbf{u} \in U$ and $|x - x_0(\mathbf{u})| < \epsilon$, where $g(x, \mathbf{u}) \ne 0$ and $h(x, \mathbf{u}) \ne 0$ are analytic functions. Moreover, $G(x, \mathbf{y}(x, \mathbf{u}), \mathbf{u})$ is analytic in $(x, \mathbf{u})$ for $\mathbf{u} \in U$ and $|x - x_0(\mathbf{u})| \ge \epsilon$, but $|x| \le |x_0(\mathbf{u})| + \delta$.

Let $G(x, \mathbf{y}, \mathbf{u})$ be also a power series with non-negative Taylor coefficients at $(0, \mathbf{0}, \mathbf{0})$ such that $(x_0(\mathbf{1}), \mathbf{y}_0(\mathbf{1}), \mathbf{1})$ is an inner point of the region of convergence of $G(x, \mathbf{y}, \mathbf{u})$. Then (with $\mathbf{y}(x, \mathbf{u})$, the solution of (2.43) from Theorem 2.33)

$$G(x, \mathbf{y}(x, \mathbf{u}), \mathbf{u}) = \sum_{n,\mathbf{m}} c_{n,\mathbf{m}} x^n \mathbf{u}^{\mathbf{m}}$$

has non-negative coefficients $c_{n,\mathbf{m}}$, too. In fact, for sufficiently large $n \ge n_0$ there exists $\mathbf{m}$ with $c_{n,\mathbf{m}} > 0$. In particular it follows that

$$c_n(\mathbf{u}) = \sum_{\mathbf{m}} c_{n,\mathbf{m}} \mathbf{u}^{\mathbf{m}}$$

is non-zero for $n \ge n_0$.

Let $\mathbf{X}_n = (X_n^{(1)}, \ldots, X_n^{(N)})$, $(n \ge n_0)$ denote an N-dimensional discrete random vector with

$$\mathbb{P}\{\mathbf{X}_n = \mathbf{m}\} = \frac{c_{n,\mathbf{m}}}{c_n}. \tag{2.49}$$

Then the expectation $\mathbb{E}\, \mathbf{u}^{\mathbf{X}_n} = \mathbb{E}\, u_1^{X_n^{(1)}} \cdots u_N^{X_n^{(N)}}$ is given by

$$\mathbb{E}\, \mathbf{u}^{\mathbf{X}_n} = \frac{c_n(\mathbf{u})}{c_n(\mathbf{1})}.$$

Putting these preliminaries together we finally get a central limit theorem for random variables that are related to systems of functional equations. For convenience we set

$$\boldsymbol{\mu} = -\frac{x_{0,\mathbf{u}}(\mathbf{1})}{x_0(\mathbf{1})},$$

and define a matrix $\boldsymbol{\Sigma}$ by

$$\boldsymbol{\Sigma} = -\frac{x_{0,\mathbf{uu}}(\mathbf{1})}{x_0(\mathbf{1})} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} + \mathrm{diag}(\boldsymbol{\mu}), \tag{2.50}$$

where $x = x_0(\mathbf{u})$ and $\mathbf{y} = \mathbf{y}_0(\mathbf{u})$ are the solutions of the system (2.41) and (2.42).

Theorem 2.35. *Suppose that $\mathbf{X}_n$ is a sequence of N-dimensional random vectors that are defined by (2.49), where $\sum_{n,\mathbf{m}} c_{n,\mathbf{m}} x^n \mathbf{u}^{\mathbf{m}} = G(x, \mathbf{y}(x, \mathbf{u}), \mathbf{u})$ and the generating functions $\mathbf{y}(x, \mathbf{u}) = (y_j(x, \mathbf{u}))_{1 \le j \le N}$ satisfy a system of functional equations of the form (2.43), in which $\mathbf{F}$ satisfies the assumptions of Theorem 2.33.*

Then $\mathbf{X}_n$ satisfies a central limit theorem of the form

$$\frac{1}{\sqrt{n}} \left(\mathbf{X}_n - \mathbb{E}\, \mathbf{X}_n\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right). \tag{2.51}$$

Furthermore there exist positive constants c_1, c_2, c_3 such that

$$\mathbb{P}\left\{\|\mathbf{X}_n - \mathbb{E}\, \mathbf{X}_n\| \ge \epsilon\sqrt{n}\right\} \le c_1 e^{-c_2 \epsilon^2} \tag{2.52}$$

uniformly for $\epsilon \le c_3\sqrt{n}$.

Finally we comment the evaluation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The problem is to extract the derivatives of $x_0(\mathbf{u})$. The function $x_0(\mathbf{u})$ is the solution of the system (2.41)–(2.42) and is exactly the location of the singularity of the mapping $x \mapsto \mathbf{y}(x, \mathbf{u})$ when $\mathbf{u}$ is fixed (and close to $\mathbf{1}$).

Let $x_0(\mathbf{u})$ and $\mathbf{y}_0(\mathbf{u}) = \mathbf{y}(x_0(\mathbf{u}), \mathbf{u})$ denote the solutions of (2.41–2.42). Then we have

$$\mathbf{y}_0(\mathbf{u}) = \mathbf{F}(x_0(\mathbf{u}), \mathbf{y}_0(\mathbf{u}), \mathbf{u}). \tag{2.53}$$

Taking derivatives with respect to $\mathbf{u}$ we get

$$\begin{aligned}\mathbf{y}_{0,\mathbf{u}}(\mathbf{u}) = {} & \mathbf{F}_x(x_0(\mathbf{u}), \mathbf{y}_0(\mathbf{u}), \mathbf{u}) x_{0,\mathbf{u}}(\mathbf{u}) + \mathbf{F}_{\mathbf{y}}(x_0(\mathbf{u}), \mathbf{y}_0(\mathbf{u}), \mathbf{u}) \mathbf{y}_{0,\mathbf{u}}(\mathbf{u}) \\ & + \mathbf{F}_{\mathbf{u}}(x_0(\mathbf{u}), \mathbf{y}_0(\mathbf{u}), \mathbf{u}),\end{aligned} \tag{2.54}$$

where the three terms in $\mathbf{F}$ denote evaluations at $(x_0(\mathbf{u}), \mathbf{y}_0(\mathbf{u}), \mathbf{u})$ of the partial derivatives of $\mathbf{F}$, and not the (total) derivative of the composite function $\mathbf{F}(x_0(\mathbf{u}), \mathbf{y}_0(\mathbf{u}), \mathbf{u})$, and where $x_{0,\mathbf{u}}$ and $\mathbf{y}_{0,\mathbf{u}}$ denote the Jacobian matrices of x_0 and $\mathbf{y}_0$ with respect to $\mathbf{u}$. In particular, for $\mathbf{u} = \mathbf{1}$ we have $x_0(\mathbf{1}) = x_0$ and $\mathbf{y}_0(\mathbf{1}) = \mathbf{y}_0$ and

$$\det(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x_0, \mathbf{y}_0, \mathbf{1})) = 0.$$

Since $\mathbf{F}_{\mathbf{y}}$ is a non-negative matrix and the dependency graph is strongly connected, there is a unique Perron-Frobenius eigenvalue of multiplicity 1. Here this eigenvalue equals 1. Thus, $\mathbf{I} - \mathbf{F}_{\mathbf{y}}$ has rank $N - 1$ and (up to scaling) has a unique positive left eigenvector $\mathbf{b}$:

$$\mathbf{b}^{\mathrm{T}}(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x_0, \mathbf{y}_0, \mathbf{1})) = \mathbf{0}.$$

From (2.54) we obtain

$$(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x_0, \mathbf{y}_0, \mathbf{1}))\mathbf{y}_{\mathbf{u}}(\mathbf{1}) = \mathbf{F}_x(x_0, \mathbf{y}_0, \mathbf{1}) x_{\mathbf{u}}(\mathbf{1}) + \mathbf{F}_{\mathbf{u}}(x_0, \mathbf{y}_0, \mathbf{1}).$$

By multiplying $\mathbf{b}$ from the left we thus get

$$\mathbf{b}^{\mathrm{T}} \mathbf{F}_x(x_0, \mathbf{y}_0, \mathbf{1}) x_{\mathbf{u}} + \mathbf{b}^{\mathrm{T}} \mathbf{F}_{\mathbf{u}}(x_0, \mathbf{y}_0, \mathbf{1}) = 0, \tag{2.55}$$

and consequently

$$\boldsymbol{\mu} = \frac{1}{x_0} \frac{\mathbf{b}^{\mathrm{T}} \mathbf{F}_{\mathbf{u}}(x_0, \mathbf{y}_0, \mathbf{1})}{\mathbf{b}^{\mathrm{T}} \mathbf{F}_x(x_0, \mathbf{y}_0, \mathbf{1})}. \tag{2.56}$$

The derivation of $\boldsymbol{\Sigma}$ is more involved. We first define $\mathbf{b}(x, \mathbf{y}, \mathbf{u})$ as the (generalised) vector product[6] of the $N - 1$ last columns of the matrix $\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x, \mathbf{y}, \mathbf{u})$. We define $D(x, \mathbf{y}, \mathbf{u})$ as

$$D(x, \mathbf{y}, \mathbf{u}) = \left(\mathbf{b}^{\mathrm{T}}(x, \mathbf{y}, \mathbf{u}) \left(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x, \mathbf{y}, \mathbf{u})\right)\right)_1 = \det\left(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(x, \mathbf{y}, \mathbf{u})\right),$$

where the subindex denotes the first coordinate. In particular we have

[6] More precisely, this is the wedge product combined with the Hodge duality.

$$D(x(\mathbf{u}), \mathbf{y}(\mathbf{u}), \mathbf{u}) = 0.$$

Then from

$$\begin{aligned}(\mathbf{I} - \mathbf{F}_\mathbf{y})\mathbf{y}_\mathbf{u} &= \mathbf{F}_x x_\mathbf{u} + \mathbf{F}_\mathbf{u}, \\ -D_\mathbf{y}\mathbf{y}_\mathbf{u} &= D_x x_\mathbf{u} + D_\mathbf{u}\end{aligned} \tag{2.57}$$

we can calculate $\mathbf{y}_\mathbf{u}$ (the first system has rank $N-1$; this means that we can skip the first equation. This reduced system is then completed to a regular system by appending the second equation (2.57)).

We now set

$$\begin{aligned}d_1(\mathbf{u}) &= d_1(x(\mathbf{u}), \mathbf{y}(\mathbf{u}), \mathbf{u}) = \mathbf{b}(x(\mathbf{u}), \mathbf{y}(\mathbf{u}), \mathbf{u})\mathbf{F}_x(x(\mathbf{u}), \mathbf{y}(\mathbf{u}), \mathbf{u}) \\ \mathbf{d}_2(\mathbf{u}) &= \mathbf{d}_2(x(\mathbf{u}), \mathbf{y}(\mathbf{u}), \mathbf{u}) = \mathbf{b}(x(\mathbf{u}), \mathbf{y}(\mathbf{u}), \mathbf{u})\mathbf{F}_\mathbf{u}(x(\mathbf{u}), \mathbf{y}(\mathbf{u}), \mathbf{u}).\end{aligned}$$

By differentiating equation (2.55) we get

$$x_{\mathbf{uu}}(\mathbf{u}) = -\frac{(d_{1x}x_\mathbf{u} + d_{1\mathbf{y}}\mathbf{y}_\mathbf{u} + d_{1\mathbf{u}})x_\mathbf{u} + (\mathbf{d}_{2x}x_\mathbf{u} + \mathbf{d}_{2\mathbf{y}}\mathbf{y}_\mathbf{u} + \mathbf{d}_{2\mathbf{u}})}{d_1}, \tag{2.58}$$

where $d_{1x}, d_{1\mathbf{y}}, d_{1\mathbf{u}}, \mathbf{d}_{2x}, \mathbf{d}_{2\mathbf{y}}, \mathbf{d}_{2\mathbf{u}}$ denote the respective partial derivatives, and where we have omitted the dependence on $\mathbf{u}$. With the knowledge of $x_0, \mathbf{y}_0$ and $\mathbf{y}_\mathbf{u}(\mathbf{1})$ we can now evaluate $x_{\mathbf{uu}}$ at $\mathbf{u} = \mathbf{1}$ and compute $\mathbf{\Sigma}$ from (2.50).

3

Advanced Tree Counting

In this third chapter we will present methods for counting trees that are based on the concept of generating functions. First we derive explicit formulas for basic tree classes and asymptotic formulas for simply generated trees and Pólya trees. However, the main goal is to show that certain tree parameters that behave *additively* (in a proper sense) satisfy a central limit theorem in a natural probabilistic setting.

For instance, if we consider certain trees $\mathcal{T}_n$ of size n it is natural to assume that every tree in $\mathcal{T}_n$ is equally likely. Every tree parameter then induces a sequence of random variables X_n (depending on the size n). A prominent example of tree parameters is the number of leaves or the number of nodes of degree k. We call a parameter additive if it is also obtained (up to some small correction term) by splitting the tree into subtrees and adding up. The number of leaves is additive in this sense. Another additive parameter is the number of occurrences of a certain pattern that we will discuss in Section 3.3.

Intuitively we can expect a central limit theorem for additive tree parameters, since they can be seen as the approximate sum of random variables if we split the tree into small subtrees. Nevertheless, it seems not to be possible to apply a direct probabilistic approach. The dependence structure is not easy to cover and there is no direct evolution process that recovers the combinatorial probability model.

Instead we use the method of generating functions that directly extends to the counting problem of additive parameters. Depending on the complexity of the parameter the recursive structure of trees leads to a functional equation or to a system of functional equations for the corresponding generating functions, and the result of Chapter 2 can be applied.

3.1 Generating Functions and Combinatorial Trees

Generating functions are quite natural in the context of tree counting, since (rooted) trees have a recursive structure that usually translates to recurrence relations in corresponding counting problems. And generating functions are a proper tool for solving recurrence equations.

3.1.1 Binary and m-ary Trees

We recall the explicit formula for the number b_n of binary trees with n internal nodes that was obtained by a proper use of generating functions (see Theorem 2.1):

$$b_n = C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Another kind of rooted trees where we can solve the counting problem by using generating functions is the class of m-ary rooted trees, where $m \ge 2$ is a fixed integer. As in the binary case ($m = 2$) we just take into account the number n of internal nodes. The number of leaves is then given by $(m-1)n+1$ and the total number of nodes by $mn + 1$.

Theorem 3.1. *The number $b_n^{(m)}$ of m-ary trees with n internal nodes is given by*

$$b_n^{(m)} = \frac{1}{(m-1)n+1}\binom{mn}{n}.$$

Proof. As in the binary case, m-ary trees $\mathcal{B}_m$ can be described formally by

$$\mathcal{B}_m = \Box + \circ \times \mathcal{B}_m^m,$$

compare also with the schematic Figure 3.1.

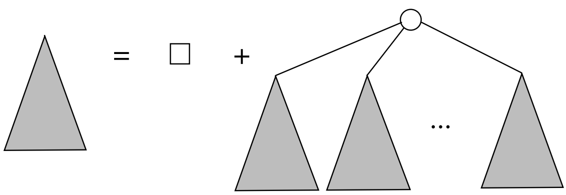


Fig. 3.1. Recursive structure of an m-ary tree

Thus, the generating function

$$b_m(x) = \sum_{n\ge 0} b_n^{(m)} x^n$$

satisfies the relation

$$b_m(x) = 1 + x\, b_m(x)^m.$$

Setting $\tilde{b}_m(x) = b_m(x) - 1$ we get

$$\tilde{b}_m(x) = x(1 + \tilde{b}_m(x))^m$$

and by Lagrange's inversion formula (for $n \ge 1$)

$$\begin{aligned} b_n^{(m)} = [x^n]\tilde{b}_m(x) &= \frac{1}{n}[u^{n-1}](1+u)^{mn} \\ &= \frac{1}{n}\binom{mn}{n-1} = \frac{1}{(m-1)n+1}\binom{mn}{n}. \end{aligned}$$

3.1.2 Planted Plane Trees

We recall that planted plane trees (or Catalan trees) are also rooted trees, where each node has an arbitrary number of successors with a natural left-to-right-order (similarly as for binary trees). There is a similar formula for the number p_n of planted plane trees of size n.

Theorem 3.2. *The number p_n of planted plane trees with $n \ge 1$ nodes is given by*

$$p_n = \frac{1}{n}\binom{2n-2}{n-1}.$$

Proof. We directly proceed in a formal way. Let $\mathcal{P}$ denote the set of planted plane trees. Then from the above description we obtain the recursive relation

$$\mathcal{P} = \circ + \circ \times \mathcal{P} + \circ \times \mathcal{P}^2 + \circ \times \mathcal{P}^3 + \cdots,$$

see also the schematic Figure 3.2.

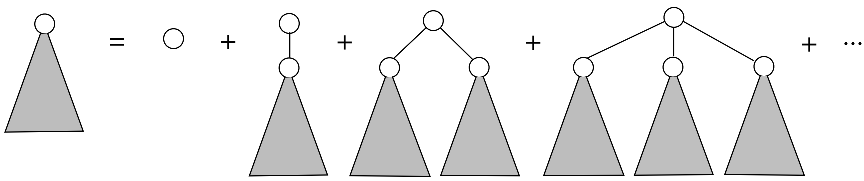


Fig. 3.2. Recursive structure of a planted plane tree

With

$$p(x) = \sum_{n \ge 1} p_n x^n$$

this translates to

$$p(x) = x + x\,p(x) + x\,p(x)^2 + x\,p(x)^3 + \cdots = \frac{x}{1-p(x)}.$$

Hence

$$p(x) = \frac{1-\sqrt{1-4x}}{2} = x\,b(x), \tag{3.1}$$

where $b(x)$ is the generating function of binary trees (compare with Theorem 2.1). Consequently

$$p_n = b_{n-1} = \frac{1}{n}\binom{2n-2}{n-1}.$$

Remark 3.3 *As in the case of binary trees we can also use Lagrange's inversion formula (with $\Phi(x) = 1/(1-x)$ to obtain p_n explicitly:*

$$p_n = \frac{1}{n}[u^{n-1}](1-u)^{-n} = \frac{1}{n}\binom{-n}{n-1}(-1)^{n-1} = \frac{1}{n}\binom{2n-2}{n-1}.$$

The relation $p_n = b_{n-1}$ has a deeper meaning. First there is a *natural bijection* between planted plane trees with n nodes and binary trees with $n-1$ internal nodes, the so-called *rotation correspondence.* Let us start with a planted plane tree with n nodes and apply the following procedure:

1. Delete the root and all edges going to the root.
2. If a node has successors delete all edges to these successors despite one edge to the most left one.
3. Join all these (previous) successors with a path (by *horizontal edges*).
4. Rotate all these new (horizontal) edges by the angle $\pi/4$ below.
5. The remaining $n-1$ nodes are now considered as internal nodes of a binary tree. Append the (missing) $n+1$ external leaves.

The result is a binary tree with $n-1$ internal nodes. It is easy to verify that this procedure is bijective (compare with the example given in Figure 3.3).

A second interpretation of the relation $p_n = b_{n-1}$ comes from an alternate recursive description of planted plane trees. If a planted plane tree has more than one node then we can delete the left-most edge of the root and obtain two planted plane trees, the original one minus the left-most subtree of the root and the left-most subtree of the root (see Figure 3.4). Obviously this description leads to the recursive description

$$\mathcal{P} = \circ + \mathcal{P}^2,$$

which gives

$$p(x) = x + p(x)^2.$$

This is equivalent to $p(x) = x/(1-p(x))$. Setting $p(x) = xb(x)$, this is equivalent to $b(x) = 1 + xb(x)^2$, and consequently we have $p_n = b_{n-1}$. Note that the recursive description (depicted in Figure 3.4) also leads to the rotation correspondence.

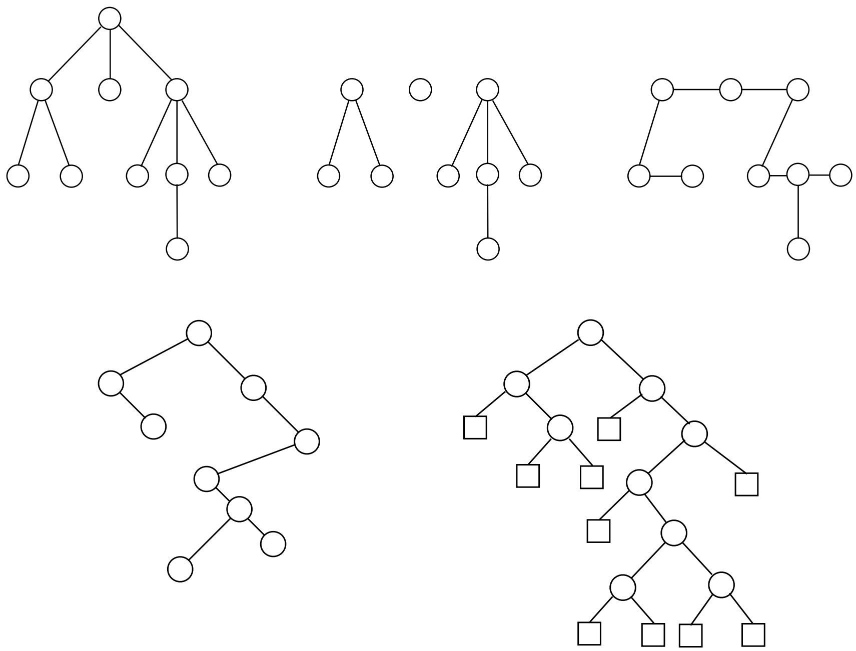


Fig. 3.3. Rotation correspondence

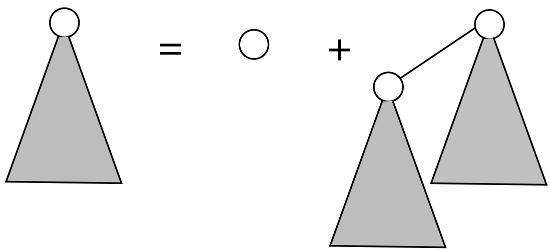


Fig. 3.4. Alternative recurrence for planted plane trees

3.1.3 Labelled Trees

A similar counting procedure applies to labelled (rooted and unrooted) trees, also called *Cayley trees*, and lead to Cayley's formula.

Theorem 3.4. *The number l_n of rooted labelled trees of size n is given by*

$$l_n = n^{n-1}.$$

Consequently the number of unrooted labelled trees of size n equals n^{n-2}.

Proof. Let $\mathcal{L}$ denote the set of labelled rooted trees. Then $\mathcal{L}$ can be recursively described as a root followed by an unordered k-tuple of labelled rooted trees for some $k \ge 0$:

$$\mathcal{L} = \circ + \circ * \mathcal{L} + \frac{1}{2!} \circ * \mathcal{L} * \mathcal{L} + \frac{1}{3!} \circ * \mathcal{L} * \mathcal{L} * \mathcal{L} + \cdots .$$

Thus, it is appropriate to use the exponential generating function

$$l(x) = \sum_{n \ge 0} \frac{l_n}{n!} x^n$$

of l_n. The above recursive description is then translated into

$$l(x) = x + x\, l(x) + x \frac{l(x)^2}{2!} + x \frac{l(x)^3}{3!} + \cdots = x e^{l(x)}.$$

Finally, by Lagrange's inversion formula

$$l_n = n! \frac{1}{n} [u^{n-1}] e^{un} = n^{n-1}.$$

There is also an explicit formula for the number of different planar embeddings.

Theorem 3.5. *The number $\hat{l}_n$ of different planar embeddings of rooted labelled trees of size n is given by*

$$\hat{l}_1 = 1 \quad \textit{and} \quad \hat{l}_n = n \frac{(2n-3)!}{(n-1)!} \quad (n \ge 2).$$

Consequently the number of different planar embeddings of unrooted labelled trees of size n equals $(2n-3)!/(n-1)!$ (for $n \ge 2$).

Proof. Let $\hat{p}(x)$ denote the exponential generating function of labelled rooted plane trees. Then due to the recursive structure of these kinds of trees we have (compare with the proof of Theorem 3.2)

$$\hat{p}(x) = \frac{x}{1 - \hat{p}(x)}.$$

Consequently, the exponential generating function $\hat{l}(x)$ for the numbers $\hat{l}_n$ of different embeddings for labelled rooted trees is given by

$$\hat{l}(x) = x + x \sum_{k \ge 1} \frac{1}{k} \hat{p}(x)^k = x + x \log \frac{1}{1 - \hat{p}(x)}.$$

Hence, by Lagrange's inversion formula we obtain (for $n \ge 2$)

$$\begin{aligned}\hat{l}_n &= n![x^{n-1}]\log\frac{1}{1-\hat{p}(x)}\\ &= n!\frac{1}{n-1}[u^{n-2}]\frac{1}{1-u}\,\frac{1}{(1-u)^{n-1}}\\ &= n\frac{(2n-3)!}{(n-1)!}.\end{aligned}$$

3.1.4 Simply Generated Trees – Galton-Watson Trees

We recall that simply generated trees can be considered as weighted planted plane trees (introduced by Meir and Moon [151]) and are proper generalisations of several types of rooted trees (compare with Section 1.2.7). Let

$$\Phi(x) = \phi_0 + \phi_1 x + \phi_2 x^2 + \cdots$$

be the generating function of the weight sequence ϕ_j, $j \ge 0$. Furthermore, we introduce the generating function $y(x) = \sum_{n\ge 1} y_n x^n$ of the weighted numbers of trees of size n:

$$y_n = \sum_{|T|=n} \omega(T).$$

Recall also that the weight of a rooted tree T is given by $\omega(T) = \prod_{j\ge 0}\phi_j^{D_j(T)}$, where $D_j(T)$ denotes the number of nodes in T with j successors. Due to the recursive structure of planted plane trees and the multiplicative structure of the weights $\omega(T)$ the generating function $y(x)$ satisfies the functional equation

$$y(x) = x\phi_0 + x\phi_1 y(x) + x\phi_2 y(x)^2 + \cdots = x\,\Phi(y(x)).$$

In view of this observation it is convenient to think of simply generated trees $\mathcal{T}$ as a weighted recursive structure of the form

$$\mathcal{T} = \phi_0\cdot\circ + \phi_1\cdot\circ\times\mathcal{T} + \phi_2\cdot\circ\times\mathcal{T}^2 + \cdots .$$

Equivalently we can consider the resulting trees of a Galton-Watson branching process (see Section 1.2.7). The numbers y_n are then the probabilities that a Galton-Watson trees has size n.

By Lagrange's inversion formula we get for all simply generated trees (and for all Galton-Watson trees)

$$y_n = \frac{1}{n}[u^{n-1}]\,\Phi(u)^n. \tag{3.2}$$

But there are only few cases where we can use this formula to obtain *nice* explicit expressions for y_n. Nevertheless, there is a quite general asymptotic result which relies on the fact that (under certain conditions) the generating function $y(x)$ has a finite radius of convergence r and that $y(x)$ has a singularity of square root type at $x_0 = r$, that is, $y(x)$ has a representation of the form

$$
\begin{aligned}
y(x) &= g(x) - h(x)\sqrt{1 - \frac{x}{x_0}} \\
&= c_0 + c_1\sqrt{x - x_0} + c_2(x - x_0) + O\left(|x - x_0|^{3/2}\right),
\end{aligned}
$$

where $g(x)$ and $h(x)$ are analytic at x_0 (compare with Theorem 2.19). For binary and planted plane trees this has been made explicit, see (2.3) and (3.1). Such representations can be used to derive asymptotic expansions for the coefficients y_n.

Theorem 3.6. *Let R denote the radius of convergence of $\Phi(x)$ and suppose that there exists τ with $0 < \tau < R$ that satisfies $\tau\Phi'(\tau) = \Phi(\tau)$. Set $d = \gcd\{j > 0 : \phi_j > 0\}$. Then*

$$
y_n = d\sqrt{\frac{\Phi(\tau)}{2\pi\Phi''(\tau)}}\frac{\Phi'(\tau)^n}{n^{3/2}}\left(1 + O(n^{-1})\right) \qquad (n \equiv 1 \bmod d) \tag{3.3}
$$

and $y_n = 0$ if $n \not\equiv 1 \bmod d$.

Proof. We apply Theorem 2.19 for $F(x, y) = x\Phi(y)$ and assume first for simplicity that $d = 1$. Then all assumptions are satisfied. In particular we have $x_0 = 1/\Phi'(\tau)$ and $y_0 = \tau$.

If $d > 1$ then it is easy to see that $y_n = 0$ if $n \not\equiv 1 \bmod d$. Consequently we have $y(x) = \tilde{y}(x^d)/x^{d-1}$ and (of course) $\Phi(x) = \tilde{\Phi}(x^d)$ for analytic functions $\tilde{y}(x)$ and $\tilde{\Phi}(x)$. They satisfy $\tilde{y}(x) = x\tilde{\Phi}(\tilde{y}(x))$ and the corresponding gcd $\tilde{d} = 1$. Thus, Theorem 2.19 can be applied to this equation and we obtain (3.3) in general.

Note that for m-ary trees and for planted plane trees this asymptotic formula also follows from the explicit formula for $b_n^{(m)}$ and p_n via Stirling's formula.

Remark 3.7 *Theorem 3.6 remains true in a slightly more general situation. Suppose that $\tau > 0$ is the radius of convergence of $\Phi(x)$ for which we have $\tau\Phi'(\tau) = \Phi(\tau)$ and $\Phi''(\tau) < \infty$. Then y_n is asymptotically given by*

$$
y_n \sim d\sqrt{\frac{\Phi(\tau)}{2\pi\Phi''(\tau)}}\frac{\Phi'(\tau)^n}{n^{3/2}} \qquad (n \equiv 1 \bmod d). \tag{3.4}
$$

The proof of (3.4) is in principle very similar to the proof of (3.3). One also observes that the generating function $y(x)$ analytically extends to a Δ-region such that $y(x) = c_0 + c_1\sqrt{x - x_0}\,(1 + o(1))$ as $x \to x_0$ (in Δ). Thus, the Transfer Lemma 2.12, or its variant mentioned in Remark 2.13, leads to (3.4) (for detail see [114]).

3.1.5 Unrooted Trees

Let $\tilde{\mathcal{T}}$ denote the set of unlabelled unrooted trees and $\mathcal{T}$ the set of unlabelled rooted trees (we do not distinguish between different embeddings in the plane). Sometimes these kinds of trees are called Pólya trees. The corresponding cardinalities of these trees (of size n) are denoted by $\tilde{t}_n$ and t_n, and the generating functions by

$$\tilde{t}(x) = \sum_{n\geq 1} \tilde{t}_n x^n \quad \text{and} \quad t(x) = \sum_{n\geq 1} t_n x^n.$$

The structure of these trees is much more complex than that of rooted trees, where the successors have a left-to-right-order. We have to apply Pólya's theory of counting and an amazing observation (3.6) by Otter [168].

Theorem 3.8. *The generating functions $t(x)$ and $\tilde{t}(x)$ satisfy the functional equations*

$$t(x) = x \exp\left(t(x) + \frac{1}{2}t(x^2) + \frac{1}{3}t(x^3) + \cdots \right) \tag{3.5}$$

and

$$\tilde{t}(x) = t(x) - \frac{1}{2}t(x)^2 + \frac{1}{2}t(x^2). \tag{3.6}$$

They have a common radius of convergence $\rho \approx 0.338219$ which is given by $t(\rho) = 1$, that is, $t(x)$ is convergent at $x = \rho$. They have a local expansion of the form

$$t(x) = 1 - b(\rho - x)^{1/2} + c(\rho - x) + d(\rho - x)^{3/2} + O((\rho - x)^2)) \tag{3.7}$$

and

$$\tilde{t}(x) = \frac{1 + t(\rho^2)}{2} + \frac{b^2 - \rho t'(\rho^2)}{2}(\rho - x) + bc(\rho - x)^{3/2} + O((\rho - x)^2), \tag{3.8}$$

where $b \approx 2.6811266$ and $c = b^2/3 \approx 2.3961466$ and $x = \rho$ is the only singularity on the circle of convergence $|x| = \rho$. Finally, t_n and $\tilde{t}_n$ are asymptotically given by

$$t_n = \frac{b\sqrt{\rho}}{2\sqrt{\pi}} n^{-3/2} \rho^{-n} \left(1 + O(n^{-1})\right) \tag{3.9}$$

and

$$\tilde{t}_n = \frac{b^3 \rho^{3/2}}{4\sqrt{\pi}} n^{-5/2} \rho^{-n} \left(1 + O(n^{-1})\right). \tag{3.10}$$

Remark 3.9 *In 1937, Pólya [176] already discussed the generating function $t(x)$ and showed that the radius of convergence ρ satisfies $0 < \rho < 1$ and that $x = \rho$ is the only singularity on the circle of convergence $|x| = \rho$. Later Otter [168] showed that $t(\rho) = 1$ and used the representation (3.7) to deduce the asymptotics for t_n. He also calculated $\rho \approx 0.338219$ and $b \approx 2.6811266$. However, his main contribution was to show (3.6). Consequently he derived (3.8) and (3.10).*

Proof. Firt we derive (3.5). As in the previous cases we can think of rooted trees in a recursive way, that is, $\mathcal{T}$ is a root followed by an unordered sequence of rooted trees:

$$\mathcal{T} = \circ \times \mathcal{M}_{\text{fin}}(\mathcal{T}).$$

Thus, we obtain (3.5).

The radius of convergence ρ of $t(x)$ surely satisfies $\frac{1}{4} \le \rho \le 1$ (this follows from $t_n \le p_n$ and $t_n \to \infty$). Next we show that $t(\rho)$ is finite (although $x = \rho$ is a singularity of $t(x)$) and that $\rho < 1$. From (3.5) it follows that $\log(t(x)/x) \ge t(x)$ for $0 < x < \rho$. Hence,

$$\frac{t(x)/x}{\log(t(x)/x)} \le \frac{1}{x}$$

and consequently $t(\rho)$ has to be finite. If $\rho = 1$ then $t(\rho^k) = t(\rho)$ for all $k \ge 1$ and it would follow that

$$\lim_{x \to \rho-} e^{t(x)+\frac{1}{2}t(x^2)+\frac{1}{3}t(x^3)+\cdots} = \infty,$$

which is impossible. Thus, $\rho < 1$ and consequently the functions $t(x^2), t(x^3), \ldots$ are regular at $x = \rho$. Moreover, they are analytic for $|x| \le \rho + \epsilon$ (for some sufficiently small $\epsilon > 0$) and are also bounded by $|t(x^k)| \le C|x^k|$ in this range. Hence, $t(x)$ may be considered as the solution of the functional equation $y = F(x, y)$, where

$$F(x, y) = x \exp\left(y + \frac{1}{2}t(x^2) + \frac{1}{3}t(x^3) + \cdots\right).$$

This function satisfies the assumptions of Theorem 2.19. In particular, the singularity $x = \rho$ and $\eta = t(\rho)$ satisfy the system of equations

$$\eta = \rho \exp\left(\eta + \frac{1}{2}t(\rho^2) + \frac{1}{3}t(\rho^3) + \cdots\right),$$
$$1 = \rho \exp\left(\eta + \frac{1}{2}t(\rho^2) + \frac{1}{3}t(\rho^3) + \cdots\right)$$

that directly gives $\eta = t(\rho) = 1$. Now, by using the expansion (3.7) and (3.5) we also get $c = b^2/3$ by comparing coefficients. In this way we get (3.8) and (3.9). Note also that Theorem 2.19 implies that $x = \rho$ is the only singularity on the circle of convergence of $t(x)$.

Next, observe that (3.7) and (3.6) imply (3.8) and (with the help of the Transfer Lemma 2.12 or its corollary) (3.10). Therefore it remains to prove (3.6).

We consider three sets of trees, the set $\mathcal{T}$ of rooted trees, the set $\tilde{\mathcal{T}}$ of unrooted trees and the set $\mathcal{T}^{(p)}$ of (unordered) pairs $\{T_1, T_2\}$ of rooted trees of $\mathcal{T}$ with $T_1 \ne T_2$. (It will be convenient to consider the pair $\{T_1, T_2\}$ as a tree that is *rooted* by an additional edge joining the roots of T_1 and T_2.) Let $t_n^{(p)}$ denote the number of pairs of that kind with a total number of n nodes, and let $t^{(p)}(x)$ denote the generating function of $t_n^{(p)}$. Then we have

$$t^{(p)}(x) = \frac{1}{2}t(x)^2 - \frac{1}{2}t(x^2). \tag{3.11}$$

We will now show that there is a bijection between $\mathcal{T}$ and $\tilde{\mathcal{T}} \cup \mathcal{T}^{(p)}$. In view of (3.11) such a bijection implies (3.6).

Recall that an arbitrary (finite) tree has either a *central node* or a *central edge*. The central edge $e = (v, w)$ is called *symmetry line*, if the two subtrees rooted at the endpoints v and w are equal.

We first divide the set $\mathcal{T}$ into 6 subsets:

1. Let $\mathcal{T}_1$ denote those rooted trees that are rooted at the central node.
2. Let $\mathcal{T}_2$ denote those rooted trees that have a central node that is different from the root.
3. Let $\mathcal{T}_3$ denote those rooted trees that have a central edge which is not a symmetry line and where one of the two endpoints of the central edge is the root.
4. Let $\mathcal{T}_4$ denote those rooted trees that have a central edge which is not a symmetry line and where the root is not one of the two endpoints of the central edge.
5. Let $\mathcal{T}_5$ denote those rooted trees that have a central edge which is a symmetry line and where one of the two endpoints of the central edge is the root.
6. Let $\mathcal{T}_6$ denote those rooted trees that have a central edge which is a symmetry line and where the root is not one of the two endpoints of the central edge.

In a similar way we divide the unrooted trees $\tilde{\mathcal{T}}$:

1. Let $\tilde{\mathcal{T}}_1$ denote those unrooted trees that have a central node.
2. Let $\tilde{\mathcal{T}}_2$ denote those unrooted trees that have a central edge, that is not a symmetry line.
3. Let $\tilde{\mathcal{T}}_3$ denote those unrooted trees that have a symmetry line as a central edge.

Finally we partition $\mathcal{T}^{(p)}$ that we consider as trees *rooted* at an edge.

1. Let $\mathcal{T}_1^{(p)}$ be the set of pairs $\{T_1, T_2\}$ with $T_1 \neq T_2$ with the property that, if we join the roots of T_1 and T_2 by an edge then the resulting tree has a central node.
2. Let $\mathcal{T}_2^{(p)}$ be the set of pairs $\{T_1, T_2\}$ with $T_1 \neq T_2$, such that the tree that results from T_1 and T_2 by joining the roots by an edge has a central edge that is not a symmetry line and that is different from the edge joining T_1 and T_2.

3. Let $\mathcal{T}_3^{(p)}$ be the set of pairs $\{T_1, T_2\}$ with $T_1 \neq T_2$, such that the tree that results from T_1 and T_2 by joining the roots by an edge has this edge as the central one, but it is not a symmetry line.
4. Let $\mathcal{T}_4^{(p)}$ be the set of pairs $\{T_1, T_2\}$ with $T_1 \neq T_2$, such that the tree that results from T_1 and T_2 by joining the roots by an edge has a symmetry line as a central edge that is different from the edge joining T_1 and T_2.

There is a natural bijection between $\mathcal{T}_1$ and $\tilde{\mathcal{T}}_1$. We only have to take the central node as the root.

Next, there is a bijection between $\mathcal{T}_2$ and $\mathcal{T}_1^{(p)}$. We identify the first edge from the path connecting the root and the central node with the edge joining T_1 and T_2.

Next, there is a trivial bijection between the sets $\tilde{\mathcal{T}}_2$ and $\mathcal{T}_2^{(p)}$. By marking one of the two endpoints of the central edge in the trees of $\tilde{\mathcal{T}}_2$ we obtain $\mathcal{T}_3$. This can be rewritten as a bijection between $\mathcal{T}_3$ and $\tilde{\mathcal{T}}_2 \cup \mathcal{T}_2^{(p)}$.

Next, there is a bijection between $\mathcal{T}_4$ and $\mathcal{T}_3^{(p)}$. We identify the first edge from the path connecting the root and the central edge with the edge joining T_1 and T_2. Similarly there is a bijection between $\mathcal{T}_6$ and $\mathcal{T}_4^{(p)}$. Finally, there is a natural bijection between $\mathcal{T}_5$ and $\tilde{\mathcal{T}}_3$.

Putting these parts together provides the proposed bijection between $\mathcal{T}$ and $\tilde{\mathcal{T}} \cup \mathcal{T}^{(p)}$.

In a similar way we can deal with unlabelled binary trees $\mathcal{B}$, where we do not care about the embedding in the plane. In particular this means that with the only exception of the root (that has degree 2) all nodes have either degree 1 or 3. The corresponding unrooted version $\tilde{\mathcal{B}}$ is the set of unlabelled trees, where every node has either degree 1 or 3. Let the corresponding cardinalities of these trees (of size n) be denoted by $\tilde{b}_n$ and b_n, and the generating functions by

$$\tilde{b}(x) = \sum_{n\ge 1} \tilde{b}_n x^n \quad \text{and} \quad b(x) = \sum_{n\ge 1} b_n x^n.$$

Then we have (similarly to the above, compare also with [28]):

Theorem 3.10. *The generating functions $b(x)$ and $\tilde{b}(x)$ satisfy the functional equations*

$$b(x) = x + \frac{x}{2}\left(b(x)^2 + b(x^2)\right) \tag{3.12}$$

and

$$\tilde{b}(x) = \frac{x}{6}\left(b(x)^3 + 3b(x)b(x^2) + 2b(x^3)\right) - \frac{1}{2}b(x)^2 + \frac{1}{2}b(x^2). \tag{3.13}$$

They have a common radius of convergence $\rho_2 \approx 0.6345845127$ which is given by $\rho_2\, b(\rho_2) = 1$ and singular expansions corresponding to (3.7) and (3.8). Furthermore, b_n and $\tilde{b}_n$ are asymptotically given by

$$b_n = c_1 n^{-3/2} \rho_2^{-n}\left(1 + O(n^{-1})\right) \quad (n \equiv 1 \bmod 2) \tag{3.14}$$

and

$$\tilde{b}_n = c_2 n^{-5/2} \rho_2^{-n} \left(1 + O(n^{-1})\right) \tag{3.15}$$

with certain positive constants c_1, c_2.

Proof. We just comment on the combinatorial part, that is, on (3.12) and (3.13). The first equation can be rewritten as

$$b(x) = x + x P_{\mathfrak{S}_2}(b(x), b(x^2))$$

and is just a rewriting of the definition. Recall that the root vertex is the only exceptional node that has degree 2. Nevertheless with the help of $b(x)$ we obtain the generating function $r(x)$ of rooted (unlabelled) trees, where all vertices (including the root) have either degree 1 or 3:

$$r(x) = x P_{\mathfrak{S}_3}(b(x), b(x^2), b(x^3)).$$

Finally, we can adapt the above bijection between rooted trees and unrooted trees plus pairs of rooted trees to the *binary case* which leads to

$$r(x) = \tilde{b}(x) + \frac{1}{2} b(x)^2 - \frac{1}{2} b(x^2).$$

This completes the proof of (3.13).

The proof of the analytic part runs along the same lines as in the proof of Theorem 3.8. It is even easier.

3.1.6 Trees Embedded in the Plane

In a similar (but easier) way one can also consider all possible embeddings $\tilde{\mathcal{P}}$ of trees in the plane. We already discussed planted plane trees $\mathcal{P}$ and their generating function $p(x)$ which satisfies $p(x) = x/(1 - p(x))$. Let $\tilde{p}(x)$ denote the generating function of the numbers $\tilde{p}_n$ of different embeddings of (unrooted) trees of size n. Then the following relations hold.

Theorem 3.11. *The generating function* $\tilde{p}(x)$ *is given by*

$$\tilde{p}(x) = x \sum_{k \ge 0} P_{\mathfrak{C}_k}(p(x), p(x^2), \ldots, p(x^k)) - \frac{1}{2} p(x)^2 + \frac{1}{2} p(x^2), \tag{3.16}$$

where $P_{\mathfrak{C}_k}(x_1, x_2, \ldots, x_k) = \frac{1}{k} \sum_{d|k} \varphi(d) x_d^{k/d}$ *denotes the cycle index of the cyclic group* $\mathfrak{C}_k$ *of* k *elements. The numbers* $\tilde{p}_n$ *of different embedding of (unrooted) trees of size* n *are asymptotically given by*

$$\tilde{p}_n = \frac{1}{8\sqrt{\pi}} 4^n n^{-5/2} \left(1 + O(n^{-1})\right). \tag{3.17}$$

Proof. First, the generating function $r(x)$ of different embeddings of rooted trees is given by

$$r(x) = x \sum_{k \ge 0} P_{\mathfrak{C}_k}(p(x), p(x^2), \ldots, p(x^k)).$$

This is due to the fact that the subtrees of the root in planted plane trees have a left-to-right-order, but rotations around the root are not allowed. Second, as in the proof of Theorem 3.8 one has

$$\tilde{p}(x) = r(x) - \frac{1}{2} p(x)^2 + \frac{1}{2} p(x^2).$$

Consequently $\tilde{p}(x)$ has a local expansion of the form

$$\tilde{p}(x) = \frac{1}{6}(1-4x)^{3/2} + \cdots$$

which gives (3.17) by applying Corollary 2.15 of the transfer lemma (Lemma 2.12).

3.2 Additive Parameters in Trees

In this section we will treat more involved enumeration problems. As an introductory example we consider the numbers $p_{n,k}$ of planted plane trees of size n with exactly k leaves. Again the concept of generating functions is a valuable tool for deriving explicit and asymptotic results.

Theorem 3.12. *The numbers $p_{n,k}$ of planted plane trees of size n with exactly k leaves are given by*

$$p_{n,k} = \frac{1}{n}\binom{n}{k}\binom{n-1}{k}.$$

Proof. Let $p(x,u) = \sum_{n,k} p_{n,k} x^n u^k$ denote the bivariate generating function of the numbers $p_{n,k}$. Then, following the recursive description of planted plane trees one gets

$$p(x,u) = xu + x \sum_{k \ge 1} p(x,u)^k = xu + \frac{x\, p(x,u)}{1 - p(x,u)}.$$

For an instance, let x be considered as a parameter. Then we have

$$p(x,u) = \frac{ux}{\left(1 - \frac{x}{1-p(x,u)}\right)}, \tag{3.18}$$

and consequently

$$[u^k]p(x,u) = \frac{1}{k}[v^{k-1}]\left(\frac{x}{1-\frac{x}{1-v}}\right)^k.$$

Finally this implies

$$\begin{aligned} p_{n,k} &= [x^n u^k] p(x,u) \\ &= \frac{1}{k}[x^n v^{k-1}] \left(\frac{x}{1 - \frac{x}{1-v}} \right)^k \\ &= \frac{1}{k}\binom{n-1}{k-1}[v^{k-1}](1-v)^{-n+k} \\ &= \frac{1}{k}\binom{n-1}{k-1}\binom{n-1}{k} \\ &= \frac{1}{n}\binom{n}{k}\binom{n-1}{k}. \end{aligned}$$

By using Stirling's formula we directly obtain bivariate asymptotic expansions for $p_{n,k}$ of the form

$$\begin{aligned} p_{n,k} &= \frac{1}{2\pi k n}\left(\frac{n}{k}\right)^{2k}\left(\frac{n}{n-k}\right)^{2(n-k)}\left(1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right)\right) \\ &= \frac{1}{2\pi n^2}\,\frac{n}{k}\left(\frac{1-\frac{k}{n}}{\frac{k}{n}}\right)^{2k}\left(\frac{1}{1-\frac{k}{n}}\right)^{2n}\left(1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right)\right). \end{aligned} \tag{3.19}$$

In particular, if we fix n then $p_{n,k}$ is maximal if $k \approx n/2$ and we locally get a behaviour of the kind

$$p_{n,k} \sim \frac{4^n}{\pi n^2} \exp\left(-\frac{(n-2k)^2}{n}\right). \tag{3.20}$$

This approximation has several implications. First, it shows that it is most likely that a typical tree of size n has approximately $n/2$ leaves and the distribution of the number of leaves around $n/2$ looks like a Gaussian distribution.

We can make this observation more precise. Let n be given and assume that each of the p_n planted plane trees of size n is equally likely. Then the number of leaves is a random variable on this set of trees which we will denote by X_n. More precisely, we have

$$\mathbb{P}\{X_n = k\} = \frac{p_{n,k}}{p_n}.$$

Then $\mathbb{E}\, X_n \sim n/2$ and $\mathbb{V}\mathrm{ar}\, X_n \sim n/8$, and (3.20) can be restated as a weak limit theorem:

$$\frac{X_n - \mathbb{E}\, X_n}{\sqrt{\mathbb{V}\mathrm{ar}\, X_n}} \xrightarrow{d} N(0,1).$$

In view of Theorem 2.23 this result is not unexpected. The generating function $p(x,u)$ is the solution of the functional equation (3.18) that satisfies all assumptions of Theorem 2.23. Thus, we directly obtain a central limit theorem.

3.2.1 Simply Generated Trees – Galton-Watson Trees

In what follows we show that several parameters on trees that have a certain additive structure satisfy a central limit theorem. The general philosophy is to find a functional equation for the corresponding bivariate generating function and then to apply the *combinatorial central limit theorem* (Theorem 2.23).

We first consider the number of leaves of simply generated trees. This covers m-ary trees, planted plane trees and labelled rooted trees.

Theorem 3.13. *Let R denote the radius of convergence of $\Phi(t)$ and suppose that there exists τ with $0 < \tau < R$ that satisfies $\tau\Phi'(\tau) = \Phi(\tau)$. Let X_n be the random variable describing the number of leaves in trees of size n, that is*

$$\mathbb{P}\{X_n = k\} = \frac{y_{n,k}}{y_n},$$

where $y_{n,k} = \sum_{|T|=n, D_0(T)=k} \omega(T)$. Then $\mathbb{E}\, X_n = \mu\, n + O(1)$ and $\mathbb{V}\mathrm{ar}\, X_n = \sigma^2\, n + O(1)$, where $\mu = \phi_0/\Phi(\tau)$ and

$$\sigma^2 = \frac{\phi_0}{\Phi(\tau)} - \frac{\phi_0^2}{\Phi(\tau)^2} - \frac{\phi_0^2}{\tau^2\Phi(\tau)\Phi''(\tau)}.$$

Furthermore, X_n satisfies a central limit theorem of the form

$$\frac{X_n - \mathbb{E}\, X_n}{\sqrt{\mathbb{V}\mathrm{ar}\, X_n}} \xrightarrow{d} N(0,1).$$

Proof. Set

$$y(x,u) = \sum_{n,k} y_{n,k} x^n u^k.$$

Then $y(x,u)$ satisfies the functional equation

$$y(x,u) = \phi_0 x(u-1) + x\Phi(y(x,u)).$$

Thus, we just have to apply Theorem 2.21 and 2.22.

It is easy to extend the above Theorem 3.13 to the numbers of nodes of given out-degree. Let $X_n^{(k)}$ denote the number of nodes of out-degree k in a random simply generated tree of size n. Then the corresponding generating function $y_k(x,u)$ satisfies the function equation

$$y_k(x,u) = x(u-1)\phi_k\, y_k(x,u)^k + x\Phi(y_k(x,u)).$$

Hence, it follows that $X_n^{(k)}$ satisfies a central limit theorem with

$$\mathbb{E}\, X_n^{(k)} = \frac{\phi_k \tau^k}{\Phi(\tau)} n + O(1).$$

In particular the probability $d_{k;n}$ that a random node in a random simply generated tree of size n has out-degree d satisfies

$$d_{k;n} = \frac{\mathbb{E}\, X_n^{(k)}}{n} \to d_k = \frac{\phi_k \tau^k}{\Phi(\tau)} \qquad (n \to \infty).$$

Recall that $d_k = \phi_k \tau^k / \Phi(\tau)$ is exactly the offspring distribution of the critical Galton-Watson branching process that is associated to simply generated trees. Thus, we can recover this process just by looking at certain tree statistics.

The limiting probabilities $d_k, k \ge 0$ constitute a probability distribution that we will call the *out-degree distribution* of the simply generated tree family. The generating function

$$p(w) = \sum_{k \ge 0} d_k w^k = \frac{\Phi(\tau w)}{\Phi(\tau)},$$

where $\tau > 0$ satisfies $\tau \Phi'(\tau) = \Phi(\tau)$, characterises this distribution.

For example, for planted plane trees and planar embeddings of (labelled or unlabelled) trees we have

$$p(w) = \sum_{k \ge 0} \frac{1}{2^{k+1}} w^k = \frac{1}{2-w}$$

whereas for labelled trees (where we do not distinguish between different planar embeddings) we have

$$p(w) = \sum_{k \ge 0} \frac{1}{e\, k!} w^k = e^{w-1}.$$

Remark 3.14 *It is also interesting to consider the maximum degree Δ_n. As a (heuristic) rule one can say that Δ_n is concentrated around a value $k_0 = k_0(n)$ for which*

$$\mathbb{E}\left(\sum_{k \ge k_0} X_n^{(k)} \right) \approx 1 \tag{3.21}$$

(compare also with Section 6.2.4, in particular the discussion of Theorem 6.12). Since $\mathbb{E}\, X_n^{(k)} \sim d_k n$, where $d_k = \phi_k \tau^k / \Phi(\tau)$, one is led to consider the value

$$\delta(n) = \max\Big\{ k \ge 0 : \sum_{\ell \ge k} d_\ell \ge \frac{1}{n} \Big\}.$$

Actually, if $\phi_{k+1}/\phi_k \to 0$ as $k \to \infty$ then we have very strong concentration (see Meir and Moon [153]):

$$\mathbb{P}\{ |\Delta_n - \delta(n)| \le 1 \} = 1 + o(1).$$

In the remaining cases there is probably no such general theorem. Only some special cases are known. For example, for planted plane trees we have (see Carr, Goh and Schmutz [32])

$$\mathbb{P}\{\Delta_n \le k\} = \exp\left(-2^{-(k-\log_2 n+1)}\right) + o(1).$$

All these results are in accordance with the above mentioned heuristics (3.21).

Next we generalise the method used above to so-called *additive parameters*. Let $v(T)$ denote the value of a parameter of a rooted tree T. We call it additive, if

$$v(T) = v(\circ \times T_1 \times T_2 \times \cdots \times T_k) = c_k + v(T_1) + v(T_2) + \cdots + v(T_k),$$

where $T_1, \ldots, T_k$ denote the subtrees of the root of T that are rooted at the successors of the root and c_k is a given sequence of real numbers. Equivalently

$$v(T) = \sum_{j\ge 0} c_j D_j(T).$$

For example, if $c_k = 1$ and $c_j = 0$ for $j \ne k$ then $v(T)$ is just the number of nodes of out-degree k. For $n \ge 1$ we now set

$$y_n(u) = \sum_{|T|=n} \omega(T) u^{v(T)}$$

and

$$y(x,u) = \sum_{n\ge 1} y_n(u) x^n.$$

The definition of $v(T)$ and the recursive structure of simply generated trees implies that $y(x,u)$ satisfies the functional equation

$$y(x,u) = x \sum_{k\ge 0} \phi_k u^{c_k} y(x,u)^k.$$

If c_k are non-negative integers then $y_k(u)$ may be interpreted as

$$y_n(u) = \sum_{k\ge 0} y_{n,k} u^k,$$

where $y_{n,k}$ denotes the (weighted) number of trees T of size n with $v(T) = k$.

Let X_n denote the random parameter v, assuming the usual probability model on trees of size n, that is, X_n describes the distribution of $v(T)$ on the set of trees of size n, where these trees are distributed according to their weights $\omega(T)$. In particular we have

$$\mathbb{E}\, u^{X_n} = \frac{y_n(u)}{y_n}. \tag{3.22}$$

As above, the distribution of X_n is (usually) Gaussian with mean value and variance of order n.

Theorem 3.15. *Let R denote the radius of convergence of $\Phi(t)$ and suppose that there exists τ with $0 < \tau < R$ that satisfies $\tau\Phi'(\tau) = \Phi(\tau)$. Furthermore, let c_k $(k \ge 0)$ be a sequence of real numbers such that the function*

$$F(x,y,u) = x \sum_{k\ge 0} \phi_k u^{c_k} y^k$$

is analytic at $x = x_0 = 1/\Phi'(\tau)$, $y = y_0 = \tau$, $u = 1$. Then the random variable X_n defined by (3.22) has expected value $\mathbb{E}\, X_n = \mu\, n + O(1)$ and variance $\mathbb{V}\mathrm{ar}\, X_n = \sigma^2\, n + O(1)$, where $\mu = \sum_{k\ge 0} c_k \phi_k \tau^k/\Phi(\tau)$ and $\sigma^2 \ge 0$. If $\sigma^2 > 0$ then X_n satisfies a (weak) central limit theorem of the form

$$\frac{X_n - \mathbb{E}\, X_n}{\sqrt{\mathbb{V}\mathrm{ar}\, X_n}} \xrightarrow{d} N(0,1).$$

Proof. We just have to apply Theorems 2.21 and 2.22.

Remark 3.16 *By using Theorem 2.23 it would have been possible to provide an explicit formula for σ^2 that is not really elegant. Note that there are also cases with $\sigma^2 = 0$. For example, if $c_k = 1$ for all $k \ge 0$ then $v(T) = |T|$ and consequently X_n is concentrated at n.*

3.2.2 Unrooted Trees

If we do not have a recursive structure, for example unrooted trees, it is still possible to define parameters that are additive with respect to the degree distribution. Let us consider the class $\tilde{\mathcal{T}}$ of unrooted trees and define a *additive parameter* v by

$$v(\tilde{T}) = \sum_{j\ge 1} c_j \tilde{D}_j(\tilde{T}), \tag{3.23}$$

where $\tilde{D}_j(\tilde{T})$ denotes the number of nodes in $\tilde{T}$ of degree j. For example, if $c_k = 1$ for some $k \ge 1$ and $c_j = 0$ for $j \ne k$ then $v(\tilde{T})$ is just the number of nodes of degree k.

In order to analyse $v(\tilde{T})$ we also have to consider the class $\mathcal{T}$ of planted rooted trees and use the two generating functions

$$t(x,u) = \sum_{T\in\mathcal{T}} x^{|T|} u^{v'(T)} = \sum_{n\ge 1} \left(\sum_{|T|=n} u^{v'(T)} \right) x^n$$

and

$$\tilde{t}(x,u) = \sum_{\tilde{T}\in\tilde{\mathcal{T}}} x^{|\tilde{T}|} u^{v(\tilde{T})} = \sum_{n\ge 1} \left(\sum_{|\tilde{T}|=n} u^{v(\tilde{T})} \right) x^n,$$

where v' is the proper version of v for rooted trees T:

$$v'(T) = \sum_{j\ge 0} c_{j+1} D_j(T).$$

Note that both definitions are consistent since the degree of any node in a planted rooted tree equals the out-degree plus 1 (see Section 1.2.2).

Following the combinatorial constructions of Section 3.1.5 we obtain the following system of functional equations. In order to make the equations more transparent we include, too, a function $r(x,u)$ that corresponds to rooted trees, where the degree of the root equals the out-degree.

$$t(x,u) = x \sum_{k\ge 0} u^{c_{k+1}} P_{\mathfrak{S}_k}(t(x,u), t(x^2,u^2), \ldots, t(x^k,u^k)), \tag{3.24}$$

$$r(x,u) = x + x \sum_{k\ge 1} u^{c_k} P_{\mathfrak{S}_k}(t(x,u), t(x^2,u^2), \ldots, t(x^k,u^k)), \tag{3.25}$$

$$\tilde{t}(x,u) = r(x,u) - \frac{1}{2} t(x,u)^2 + \frac{1}{2} t(x^2,u^2). \tag{3.26}$$

Finally, we introduce the random variable X_n (describing the distribution of v on trees of size n) in the usual way:

$$\mathbb{E}\, u^{X_n} = \frac{1}{\tilde{t}_n} \sum_{|\tilde{T}|=n} u^{v(\tilde{T})}. \tag{3.27}$$

Theorem 3.17. *Let $(c_k)_{k\ge 1}$ be a bounded sequence of real numbers, and let $v(T)$ and X_n be defined by (3.23) and (3.27). Then there exist μ and $\sigma^2 \ge 0$ with $\mathbb{E}\, X_n = \mu\, n + O(1)$ and $\mathbb{V}\mathrm{ar}\, X_n = \sigma^2\, n + O(1)$. If $\sigma^2 > 0$ then X_n satisfies a (weak) central limit theorem of the form*

$$\frac{X_n - \mathbb{E}\, X_n}{\sqrt{\mathbb{V}\mathrm{ar}\, X_n}} \xrightarrow{d} N(0,1).$$

Proof. The proof runs as follows. First, we apply Theorem 2.21 to (3.24) which implies that $t(x,u)$ has a square root singularity of the kind (2.21). Second, we use this representation and (3.26) to get an expansion for $\tilde{t}(x,u)$ of the form

$$\tilde{t}(x,u) = \overline{g}(x,u) - \overline{h}(x,u) \left(1 - \frac{x}{f(u)}\right)^{3/2}. \tag{3.28}$$

Then we apply Theorem 2.25 to obtain a central limit theorem. The last step is a direct application. So we just have to look at the first two steps.

In order to apply Theorem 2.21 we just have to ensure that the functions $t(x^2,u^2), t(x^3,u^3), \ldots$ are analytic if x is close to ρ and u is close to 1. Since the sequence c_k is bounded we have $|c_k| \le M$ for some $M > 0$ and thus $|v'(T)| \le M\,|T|$. Hence, if $|u| > 1$ and $|xu^M| < \rho$ then we have

$$|t(x,u)| \le \sum_{n\ge 1} t_n |u|^{Mn} |x|^n = t(|xu^M|, 1).$$

In particular if $|x| \le \rho + \eta$ and $|u| \le (\sqrt{\rho}/(\rho+\eta))^{1/M}$ (where $\eta > 0$ is small enough that $(\sqrt{\rho}/(\rho+\eta))^{1/M} > 1$) we get for $k \ge 2$

$$|t(x^k, u^k)| \le t(|xu^M|^k, 1) \le t(\rho^{k/2}, 1) \le C\rho^{k/2}.$$

Thus, we can apply Theorem 2.21 with

$$F(x, y, u) = x \sum_{k\ge 0} u^{c_{k+1}} P_{\mathfrak{S}_k}(y, t(x^2, u^2), \ldots, t(x^k, u^k))$$

and obtain a representation of the form

$$t(x, u) = g(x, u) - h(x, u)\sqrt{1 - \frac{x}{f(u)}}, \tag{3.29}$$

where $g_1 = g(f(u), u)$ satisfies the relation

$$g_1 = f(u) \sum_{k\ge 0} u^{c_{k+1}} P_{\mathfrak{S}_k}(g_1, t(f(u)^2, u^2), \ldots, t(f(u)^k, u^k)).$$

Consequently, from (3.26) and (3.29) we obtain a representation for $\tilde{t}(x, u)$ of the form

$$\tilde{t}(x, u) = g_2(x, u) - h_2(x, u)\sqrt{1 - \frac{x}{f(u)}}, \tag{3.30}$$

where

$$h_2(x, u) = h(x, u)\Bigg(x \sum_{k\ge 1} u^{c_k} \frac{\partial}{\partial x_1} P_{\mathfrak{S}_k}(g(x, u), t(x^2, u^2), \ldots, t(x^k, u^k)) \\ - g(x, u) + (x - f(u))H(x, u)\Bigg)$$

in which $H(x, u)$ denotes an analytic function in x and u. Note that

$$\frac{\partial}{\partial x_1} P_{\mathfrak{S}_k}(x_1, \ldots, x_k) = P_{\mathfrak{S}_{k-1}}(x_1, \ldots, x_{k-1}).$$

This implies that

$$h_2(f(u), u) = h(f(u), u) f(u) \left(\sum_{k\ge 1} u^{c_k} \frac{\partial}{\partial x_1} P_{\mathfrak{S}_k}(g_1, \ldots) - \sum_{k\ge 0} u^{c_{k+1}} P_{\mathfrak{S}_k}(g_1, \ldots) \right) \\ = 0.$$

Hence, $h_2(x, u)$ can be represented as

$$h_2(x, u) = \overline{h}(x, u)\left(1 - \frac{x}{f(u)}\right).$$

This implies (3.28) and completes the proof of Theorem 3.17.

3.3 Patterns in Trees

In the Section 3.2.1 we have shown that a typical tree in the set of unrooted labelled trees of size n has about $\mu_k n$ nodes of degree k, where

$$\mu_k = \frac{1}{e(k-1)!}.$$

(Note that we are now considering the degree and not the out-degree.) Moreover, for any fixed k the total number of nodes of degree k over all trees in $\mathcal{T}_n$ satisfies a central limit theorem with mean and variance asymptotically equivalent to $\mu_k n$ and $\sigma_k^2 n$, where

$$\sigma_k^2 = \mu_k - \mu_k^2 - \frac{(k-2)^2}{e^2(k-1)!(k-1)!}.$$

A node of degree k can be interpreted as an occurrence of a star with k edges. Thus, the number of nodes of degree k is exactly the number of occurrences of a star pattern of fixed size. Our aim is to generalise this problem to general patterns. More precisely we consider a given finite tree, a pattern $\mathcal{M}$, and want to compute the limiting distribution of the number of occurrences of $\mathcal{M}$ in $\mathcal{T}_n$ as $n \to \infty$ (we follow [40]).

In order to make our considerations more transparent, let us consider the example pattern in Figure 3.5 that can be seen as a chain of three nodes of degree 3.

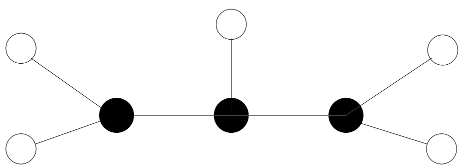


Fig. 3.5. Example pattern

We distinguish between internal (filled) nodes and external (empty) ones and say that a specific pattern $\mathcal{M}$ occurs in a tree T if $\mathcal{M}$ occurs in T as a substructure in the sense that the node degrees for the internal (filled) nodes in the pattern match the degrees of the corresponding nodes in T, while the external (empty) nodes match nodes of arbitrary degree. For instance, the example pattern occurs exactly three times in the tree depicted in Figure 3.6. Note also that there can be overlaps of two or more copies of $\mathcal{M}$, which we intend to count as separate occurrences.

The goal of this section is to prove the following general theorem that extends the result of nodes of given degree to pattern occurrences in labelled (unrooted) trees (see [40]).

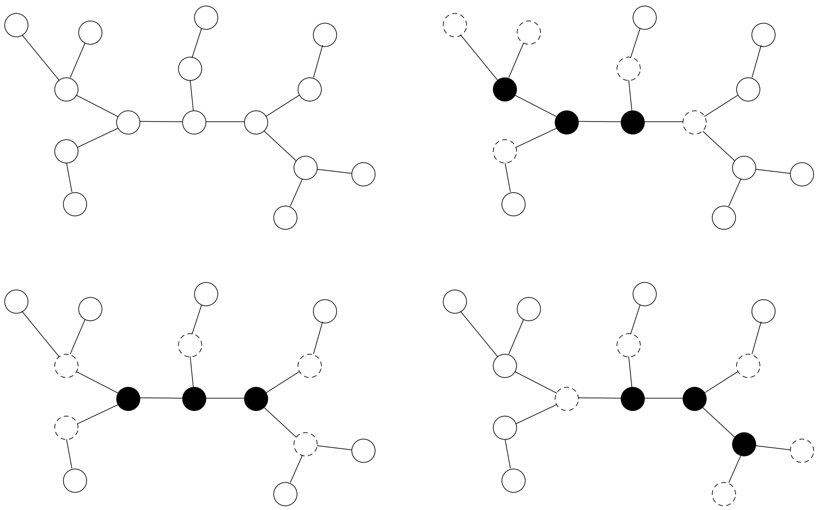


Fig. 3.6. Pattern occurrence

Theorem 3.18. *Let $\mathcal{M}$ be a given finite tree. Then the limiting distribution of the number of occurrences of $\mathcal{M}$ in a labelled tree of size n is asymptotically normal with mean and variance asymptotically equivalent to μn and $\sigma^2 n$, respectively, where $\mu > 0$ and $\sigma^2 \geq 0$ depend on the pattern $\mathcal{M}$, are polynomials in $1/e$ (with rational coefficients) and can be computed explicitly and algorithmically for each given $\mathcal{M}$.*

3.3.1 Planted, Rooted and Unrooted Trees

The counting procedure we use is recursive and based on rooted trees. However, we have to take care of several node degrees. Therefore it is more transparent to use the following three-step-program that we have already used, for example, for the proof of Theorem 3.17. We first consider planted rooted trees, then rooted trees and finally unrooted ones. We recall that a rooted tree is planted if the root is connected (or planted) to an additional phantom node that is not taken into account (compare with Section 1.2.2). This has the advantage that in this case all nodes have the property $d(v) = d^+(v) + 1$. The only but essential difference between planted rooted trees and rooted trees is that the degree of the root node is different.

In order to demonstrate the usefulness of the three-step-procedure above we consider (again) the problem of counting nodes of given degree k (which is equivalent to count stars with k edges). Let $a_{n,m}$ denote the number of planted rooted (and labelled) trees of size n with exactly m nodes of degree k. Furthermore, let $l_{n,m}$ and $t_{n,m}$ be the corresponding numbers for rooted and unrooted (labelled) trees and set

$$a(x,u) = \sum_{n,m=0}^{\infty} a_{n,m} \frac{x^n u^m}{n!}, \quad l(x,u) = \sum_{n,m=0}^{\infty} l_{n,m} \frac{x^n u^m}{n!}, \quad \text{and}$$
$$t(x,u) = \sum_{n,m=0}^{\infty} t_{n,m} \frac{x^n u^m}{n!}.$$

Then we have (compare also with (3.24)–(3.26)):

1. **Planted Rooted Trees:**

$$a(x,u) = \sum_{\substack{n=0 \\ n\neq k-1}}^{\infty} \frac{xa(x,u)^n}{n!} + \frac{xu\,a(x,u)^{k-1}}{(k-1)!} = xe^{a(x,u)} + \frac{x(u-1)a(x,u)^{k-1}}{(k-1)!}.$$

2. **Rooted Trees:**

$$l(x,u) = \sum_{\substack{n=0 \\ n\neq k}}^{\infty} \frac{xa(x,u)^n}{n!} + \frac{xu\,a(x,u)^{k}}{k!} = xe^{a(x,u)} + \frac{x(u-1)a(x,u)^{k}}{k!}.$$

3. **Unrooted Trees:** Since we are considering labelled trees, we simply have $t_{n,m} = l_{n,m}/n$. Thus, the statistics on the number of nodes of degree k are precisely the same. However, it is also possible to express $t(x,u)$ by

$$t(x,u) = l(x,u) - \frac{1}{2}a(x,u)^2.$$

   This follows from a natural bijection between rooted trees on the one hand and unrooted trees and pairs of planted rooted trees (that are joined by identifying the additional edges at their planted roots and discarding the phantom nodes) on the other hand.[1]

We now generalise the counting procedure of Section 3.3.1 to more complicated patterns. For our purpose, a pattern is a given (finite unrooted unlabelled) tree $\mathcal{M}$. For $\mathcal{M}$ we will use the example graph in Figure 3.5 in order to make the arguments more transparent.

3.3.2 Generating Functions for Planted Rooted Trees

Let $a_{n,m}$ denote the number of planted rooted (and labelled) trees with n nodes and exactly m occurrences of the pattern $\mathcal{M}$, where the additional (phantom) node is taken into account in the matching procedure. Furthermore, let

---

[1] Consider the class of rooted (labelled) trees. If the root is labelled by 1 then consider the tree as an unrooted tree. If the root is not labelled by 1 then consider the first edge of the path between the root and 1 and cut the tree into two planted rooted trees at this edge.

$$a = a(x,u) = \sum_{n,m=0}^{\infty} a_{n,m} \frac{x^n u^m}{n!}$$

be the corresponding generating function.

Lemma 3.19. *Let $\mathcal{M}$ be a pattern. Then there exists a certain number of auxiliary functions $a_j(x,u)$ $(0 \le j \le L)$ with*

$$a(x,u) = \sum_{j=0}^{L} a_j(x,u)$$

and polynomials $P_j(y_0,\ldots,y_L,u)$ $(1 \le j \le L)$ with non-negative coefficients such that

$$\begin{aligned} a_0(x,u) &= x e^{a_0(x,u)+\cdots+a_L(x,u)} - x \sum_{j=1}^{L} P_j(a_0(x,u),\ldots,a_L(x,u),1) \\ a_1(x,u) &= x\, P_1(a_0(x,u),\ldots,a_L(x,u),u) \\ &\ \vdots \\ a_L(x,u) &= x\, P_L(a_0(x,u),\ldots,a_L(x,u),u). \end{aligned} \tag{3.31}$$

Furthermore,

$$\sum_{j=1}^{L} P_j(y_0,\ldots,y_L,1) \le_c e^{y_0+\cdots+y_L},$$

where $f \le_c g$ means that all Taylor coefficients of the left-hand side are smaller than or equal to the corresponding coefficients of the right-hand-side. Moreover, the dependency graph of the system (3.31) is strongly connected.

The proof of this lemma is in fact the core of the proof of Theorem 3.18. As already mentioned we will demonstrate all steps of the proof using the example pattern in Figure 3.5. At each step we will also indicate how to make all constructions explicit so that it is possible to generate System (3.31) effectively.

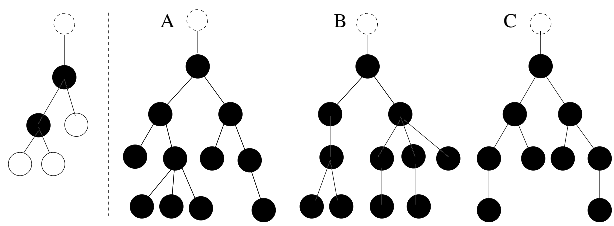


Fig. 3.7. Planted pattern matching

We introduce the notion of a *planted pattern* . A planted pattern $\mathcal{M}_p$ is just a planted rooted tree where we again distinguish between internal (filled) and external (empty) nodes. It matches a planted rooted tree from $\mathcal{T}_n$ if $\mathcal{M}_p$ occurs starting from the (planted) root, that is, the branch structure and node degrees of the filled nodes match. Two occurrences may overlap. For example, in Figure 3.7 the planted pattern $\mathcal{M}_p$ on the left matches the planted tree A twice (following the left or the right edge from the root), but B not at all. Also remark that, notwithstanding the symmetry of C, the pattern $\mathcal{M}_p$ really matches C twice.

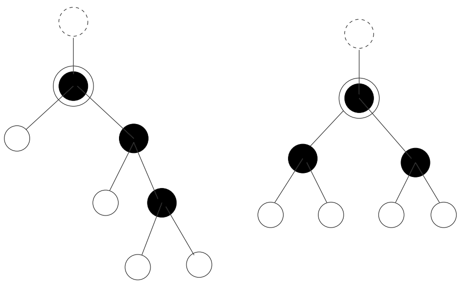


Fig. 3.8. Planted patterns for the pattern in Figure 3.5

We now construct a planted pattern for each internal (filled) node of our pattern $\mathcal{M}$ which is adjacent to an external (empty) node. The internal (filled) node is considered as the planted root and one of the free attached leaves as the plant. In our example we obtain the two graphs in Figure 3.8.

The next step is to partition all planted trees according to their degree distribution up to some adequate level. To this end, let D denote the set of out-degrees of the internal nodes in the planted patterns introduced above and h be the maximal height of these patterns. In our example we have $D = \{2\}$ and $h = 3$. For obtaining a partition, we more precisely consider all trees of height less than or equal to h with out-degrees in D. We distinguish two types of leaves in these trees, depending on the depth at which they appear: leaves in level h, denoted "$\circ$", and leaves at levels less than h, denoted "$\Box$". For our example we get 11 different trees $a_0, a_1, \ldots, a_{10}$, depicted on Figure 3.9.

These trees induce a natural partition of all planted trees for the following interpretation of the two types of leaves: We say that a tree T is contained in class[2] a_j if it matches the finite tree (or pattern) a_j in such a way that a node of type $\Box$ has degree not in D, while a node of type $\circ$ has any degree. For example, a_0 corresponds to those planted trees where the out-degree of the root is not in D.

[2] By abuse of notation the tree class corresponding to the finite tree a_j is denoted by the same symbol a_j.

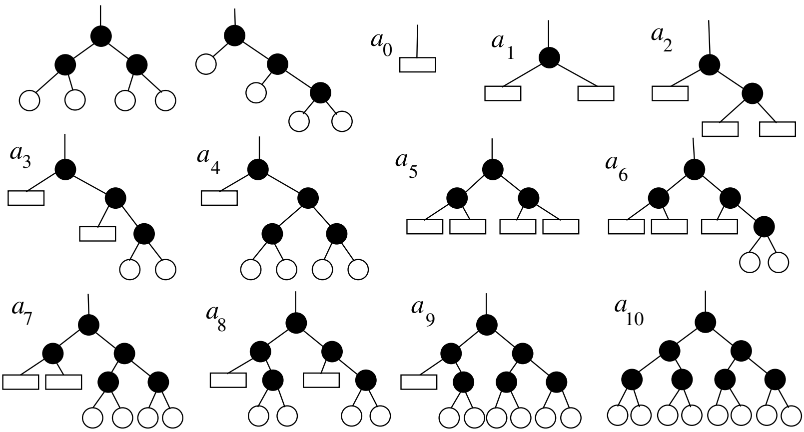


Fig. 3.9. Tree partition

It is easy to observe that these (obviously disjoint) classes of trees form a partition. Indeed, take any rooted tree. For any path from the root to a leaf, consider the first node with out-degree not in D, and replace the whole subtree at it with $\square$. Then replace any node at depth h with $\circ$. The tree obtained in this way is one in the list.

The above classes can be described recursively. Here it seems to be convenient to introduce a formal notation to describe operations between classes of trees: $\oplus$ denotes the disjoint union of classes; $\setminus$ denotes set difference; recursive descriptions of tree classes are given in the form $a_i = x a_{j_1}^{e_1} \cdots a_{j_\ell}^{e_\ell}$, to express that the class a_i is constructed by attaching e_1 subtrees from the class a_{j_1}, e_2 subtrees from the class a_{j_2}, etc, to a root node that we denote by x.

In our example we get the following relations:

$$
\begin{aligned}
a_0 &= x \oplus x \bigoplus_{i=0}^{10} a_i \oplus x \bigoplus_{n=3}^{\infty} \left(\bigoplus_{i=0}^{10} a_i \right)^n, \\
a_1 &= x a_0^2, \\
a_2 &= x a_0 a_1, \\
a_3 &= x a_0 (a_2 \oplus a_3 \oplus a_4), \\
a_4 &= x a_0 (a_5 \oplus a_6 \oplus a_7 \oplus a_8 \oplus a_9 \oplus a_{10}), \\
a_5 &= x a_1^2, \\
a_6 &= x a_1 (a_2 \oplus a_3 \oplus a_4), \\
a_7 &= x a_1 (a_5 \oplus a_6 \oplus a_7 \oplus a_8 \oplus a_9 \oplus a_{10}),
\end{aligned}
$$

$$a_8 = x(a_2 \oplus a_3 \oplus a_4)^2,$$
$$a_9 = x(a_2 \oplus a_3 \oplus a_4)(a_5 \oplus a_6 \oplus a_7 \oplus a_8 \oplus a_9 \oplus a_{10}),$$
$$a_{10} = x(a_5 \oplus a_6 \oplus a_7 \oplus a_8 \oplus a_9 \oplus a_{10})^2.$$

This is to be interpreted as follows. First, trees in a_0 consist of a (planted) root that is denoted by x and has out-degree j which is not contained in $D = \{2\}$. The successors of the root are arbitrary trees. Next, trees in a_1 consist of a root that has out-degree 2, and two successors that are of out-degree distinct from 2, that is, in a_0. Similarly, trees in a_3 consist of a root x with out-degree 2 and subject to the following additional constraints: one subtree at the root is exactly of type a_0; the other subtree, call it T, is of out-degree 2, either with both subtrees of degree other than 2 (leading to T in a_2), or with one subtree of degree 2 and the other of degree other than 2 (leading to T in a_3), or with both of its subtrees of degree 2 (leading to T in class a_4). Summarising: $a_3 = xa_0(a_2 \oplus a_3 \oplus a_4)$. Of course this can also be interpreted as $a_3 = xa_0a_2 \oplus xa_0a_3 \oplus xa_0a_4$. Another more involved example corresponds to a_8; here both subtrees are of the form $a_2 \oplus a_3 \oplus a_4$.

To show that the recursive description can be obtained in general, consider a tree a_j obtained from some planted pattern $\mathcal{M}_p$. Let $s_1, \ldots, s_d$ denote its subtrees at the root. Then, in each s_i, leaves of type $\circ$ can appear only at level $h-1$. Substitute for all such $\circ$ either $\square$ or a node of out-degree chosen from D and having $\circ$ for all its subtrees. Do this substitution in all possible ways. The collection of trees obtained are some of the a_k's, say $a_{k_1^{(j)}}$, $a_{k_2^{(j)}}$, etc. Thus, we obtain the recursive relation $a_j = x(a_{k_1^{(1)}} \oplus a_{k_2^{(1)}} \oplus \cdots) \cdots (a_{k_1^{(d)}} \oplus a_{k_2^{(d)}} \oplus \cdots)$ for a_j.

In general, we obtain a partition of $L+1$ classes $a_0, \ldots, a_L$ and corresponding recursive descriptions, where each tree type a_j can be expressed as a disjoint union of tree classes of the kind

$$xa_{j_1} \cdots a_{j_r} = xa_0^{l_0} \cdots a_L^{l_L}, \tag{3.32}$$

where r denotes the degree of the root of a_j and the non-negative integer l_i is the number of repetitions of the tree type a_i.

We proceed to show that this directly leads to a system of equations of the form (3.31), where each polynomial relation stems from a recursive equation between combinatorial classes.

Let Λ_j be the set of tuples $(l_0, \ldots, l_L)$ with the property that $(l_0, \ldots, l_L) \in \Lambda_j$, if and only if the term of type (3.32) is involved in the recursive description of a_j (in expanded form). Further, let $k = K(l_0, \ldots, l_L)$ denote the number of *additional occurrences* of the pattern $\mathcal{M}$ in (3.32) in the following sense: if $b = xa_{j_1} \cdots a_{j_r}$ and T is a labelled tree of b with subtrees $T_1 \in a_{j_1}$, $T_2 \in a_{j_2}$, etc, and $\mathcal{M}$ occurs m_1 times in T_1, m_2 times in T_2, etc, then T contains $\mathcal{M}$ exactly $m_1+m_2+\cdots+m_d+k$ times. The number k corresponds to the number of occurrences of $\mathcal{M}$ in T in which the root of T occurs as internal node of the pattern. By construction of the classes a_i this number only depends on

b and not on the particular tree $T \in b$. Let us clarify the calculation of $k = K(l_0, \ldots, l_L)$ with an example. Consider the class a_9 of the partition for the example pattern. Now we have to match the planted patterns of Figure 3.8 at the root of an arbitrary tree of class a_9. The left planted pattern of Figure 3.8 matches three times, the right one matches once. Thus we find that in this case $k = 4$. For the other classes we find the following values of $k = K(l_0, \ldots, l_L)$:

Terms of class	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
Value of k	0	0	0	1	2	1	2	3	3	4	5

Now define series P_j by

$$P_j(y_0, \ldots, y_L, u) = \sum_{(l_0,\ldots,l_L)\in\Lambda_j} \frac{1}{l_0! \cdots l_L!} y_0^{l_0} \cdots y_L^{l_L} u^{K(l_0,\ldots,l_L)}.$$

These are in fact polynomials for $1 \le j \le L$ by the finiteness of the corresponding Λ_j. All matches of the planted patterns are handled in the P_j, $1 \le j \le L$, thus

$$P_0(y_0, \ldots, y_L, u) = e^{y_0+\cdots+y_L} - \sum_{j=1}^{L} P_j(y_0, \ldots, y_L, 1)$$

does not depend on u.

In our pattern we get for example

$$\begin{aligned} P_8(y_0, \ldots, y_{10}, u) &= \frac{1}{2}xy_2^2u^3 + xy_2y_3u^3 + xy_2y_4u^3 + \frac{1}{2}xy_3^2u^3 + xy_3y_4u^3 + \frac{1}{2}xy_4^2u^3 \\ &= \frac{1}{2}x(y_2 + y_3 + y_4)^2u^3. \end{aligned}$$

Finally, let $a_{j;n,m}$ denote the number of planted rooted trees of type a_j with n nodes and m occurrences of the pattern $\mathcal{M}$ and set

$$a_j(x, u) = \sum_{n,m=0}^{\infty} a_{j;n,m} \frac{x^n u^m}{n!}.$$

By this definition it is clear that

$$a_j(x, u) = x \cdot P_j\big(a_0(x, u), \ldots, a_L(x, u), u\big),$$

because the number of labelled trees is counted by x (exponential generating function) and the number of patterns is additive and counted by u. Hence, we explicitly obtain the proposed structure of the system of functional equations (3.31).

Concerning the example pattern we reach the following system of equations, where we denote the generating function of the class a_i by the same symbol a_i:

$$
\begin{aligned}
a_0 &= a_0(x,u) = x + x\sum_{i=0}^{10} a_i + x\sum_{n=3}^{\infty}\frac{1}{n!}\left(\sum_{i=0}^{10} a_i\right)^n,\\
a_1 &= a_1(x,u) = \frac{1}{2}xa_0^2,\\
a_2 &= a_2(x,u) = xa_0a_1,\\
a_3 &= a_3(x,u) = xa_0(a_2+a_3+a_4)u,\\
a_4 &= a_4(x,u) = xa_0(a_5+a_6+a_7+a_8+a_9+a_{10})u^2,\\
a_5 &= a_5(x,u) = \frac{1}{2}xa_1^2u,\\
a_6 &= a_6(x,u) = xa_1(a_2+a_3+a_4)u^2,\\
a_7 &= a_7(x,u) = xa_1(a_5+a_6+a_7+a_8+a_9+a_{10})u^3,\\
a_8 &= a_8(x,u) = \frac{1}{2}x(a_2+a_3+a_4)^2u^3,\\
a_9 &= a_9(x,u) = x(a_2+a_3+a_4)(a_5+a_6+a_7+a_8+a_9+a_{10})u^4,\\
a_{10} &= a_{10}(x,u) = \frac{1}{2}x(a_5+a_6+a_7+a_8+a_9+a_{10})^2u^5.
\end{aligned}
$$

In order to complete the proof of Lemma 3.19 we just have to show that the dependency graph is strongly connected. By construction, $a_0 = a_0(x,u)$ depends on all functions $a_i = a_i(x,u)$. Thus, it is sufficient to prove that every a_i $(1 \le i \le L)$ eventually depends on a_0 in the sense that there exists a sequence of indices $i_1, i_2, \ldots, i_r$ such that $a_i = xP_i(a_0,\ldots,a_K)$ directly depends on a_{i_1}, a_{i_1} directly depends on a_{i_2} etc, and finally a_{i_r} directly depends on a_0. For this purpose consider the subtree of $\mathcal{M}$ that was labelled by a_i and consider a path from its root to an empty node. Each edge of this path corresponds to another subtree of $\mathcal{M}$ corresponding to $a_{i_1}, a_{i_2}, \ldots, a_{i_r}$. Then, by construction of the system of functional equations above, a_i depends on a_{i_1}, a_{i_1} depends on a_{i_2}, etc. Finally the root of a_{i_r} is adjacent to an empty node and thus (the corresponding generating function) depends on a_0. This completes the proof of Lemma 3.19.

Note that we obtain a relatively more compact form of this system by introducing

$$
\begin{aligned}
b_0 &= b_0(x,u) = a_0(x,u)\\
b_1 &= b_1(x,u) = a_1(x,u)\\
b_2 &= b_2(x,u) = a_2(x,u) + a_3(x,u) + a_4(x,u)\\
b_3 &= b_3(x,u) = a_5(x,u) + a_6(x,u) + a_7(x,u) + a_8(x,u) + a_9(x,u) + a_{10}(x,u),
\end{aligned}
\tag{3.33}
$$

together with the recursive relations

$$b_0 = xe^{b_0+b_1+b_2+b_3} - \frac{1}{2}x(b_0+b_1+b_2+b_3)^2$$
$$b_1 = \frac{1}{2}xb_0^2$$
$$b_2 = xb_0b_1 + xb_0b_2u + xb_0b_3u^2$$
$$b_3 = \frac{1}{2}xb_1^2u + xb_1b_2u^2 + xb_1b_3u^3 + \frac{1}{2}xb_2^2u^3 + +xb_2b_3u^4 + \frac{1}{2}xb_3^2u^5.$$

The combinatorial classes corresponding to the b_i (which we will also denote by b_i) have the interpretation as shown in Figure 3.10. We could have obtained the classes b_i directly by restraining the construction to a maximal depth $h-1$ instead of h. In principle, we could then apply the analytic treatment of Section 3.3.4 to the system of the b_i. However we feel that the existence of a recursive structure of the system of the b_i with a well-defined $K(l_0,..,l_L)$ for each term in the recursive description is slightly less clear. Therefore we preferred to work with the a_i which has a well-defined $K(a_i)$.

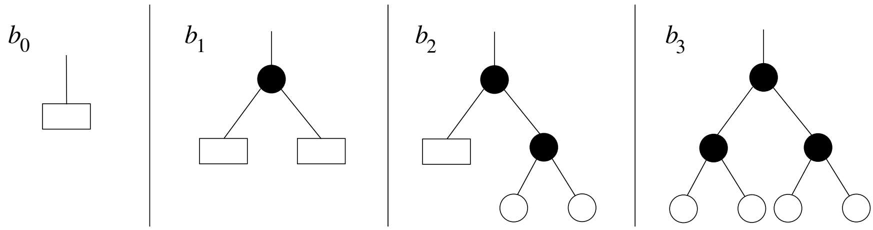


Fig. 3.10. The classes corresponding to the b_i of the equations (3.33)

3.3.3 Rooted and Unrooted Trees

The next step is to find equations for the exponential generating function of rooted trees (where occurrences of the pattern are marked with u). As above we set

$$l(x,u) = \sum_{n,m=0}^{\infty} l_{n,m}\frac{x^n u^m}{n!},$$

where $l_{n,m}$ denotes the number of rooted trees of size n with exactly m occurrences of the pattern $\mathcal{M}$. (That is, occurrences of the rooted patterns $\mathcal{M}_r$ deducible from $\mathcal{M}$. Here, a rooted pattern is defined in a very similar way as a planted pattern.)

Lemma 3.20. *Let $\mathcal{M}$ be a pattern and let*

$$a_0(x,u), \ldots, a_L(x,u)$$

denote the auxiliary functions introduced in Lemma 3.19. Then there exists a polynomial $Q(y_0,\ldots,y_L,u)$ with non-negative coefficients satisfying $Q(y_0,\ldots,y_L,1) \le_c e^{y_0+\cdots+y_L}$, such that

$$l(x,u) = G(x,u,a_0(x,u),\ldots,a_L(x,u)) \tag{3.34}$$

for

$$G(x,u,y_0,\ldots,y_L) = x\left(e^{y_0+\cdots+y_L} - Q(y_0,\ldots,y_L,1) + Q(y_0,\ldots,y_L,u)\right). \tag{3.35}$$

Proof. In principle the proof is a direct continuation of the proof of Lemma 3.19. We recall that a rooted tree is just a root with zero, one, two, ... planted subtrees, i.e., the class of rooted trees can be described as a disjoint union of classes c of rooted trees of the form $xa_{j_1}\cdots a_{j_d}$. Let l_i denote the number of classes a_i in this term such that $c = xa_0^{l_0}\cdots a_L^{l_L}$, and set $\bar{K}(l_0,\ldots,l_L)$ to be the number of additional occurrences of the pattern $\mathcal{M}$. This number again corresponds to the number of occurrences of $\mathcal{M}$ in a tree $T \in c$ in which the root of T occurs as internal node of the pattern. Set

$$Q_d(y_0,\ldots,y_L,u) = \sum_{l_0+\cdots+l_L=d} \frac{1}{l_0!\cdots l_L!} y_0^{l_0}\cdots y_L^{l_L} u^{\bar{K}(l_0,\ldots,l_L)}.$$

Then by construction

$$l(x,u) = x\sum_{d\ge 0} Q_d(a_0(x,u),\ldots,a_L(x,u),u).$$

Note that $\sum_{d\ge 0} Q_d(y_0,\ldots,y_L,1) = e^{y_0+\cdots+y_L}$. Let $\bar{D}$ denote the set of degrees of the internal (filled) nodes of the pattern, that is, $\bar{D} = \{d+1 : d \in D\}$; then $Q_d(y_0,\ldots,y_L,u)$ does not depend on u if $d \notin \bar{D}$. With

$$Q(y_0,\ldots,y_L,u) := \sum_{d\in\bar{D}} Q_d(y_0,\ldots,y_L,u)$$

we obtain (3.34) and (3.35). The number $\bar{K}(l_0,\ldots,l_L)$ is well-defined for a similar reason as was $K(l_0,\ldots,l_L)$, and can be calculated similarly.

We again illustrate the proof with our example. In Figure 3.11 the corresponding rooted patterns are shown. For convenience let $l_0 = l_0(x,u)$ denote the function

$$l_0 = xe^a - \frac{xa^3}{3!},$$

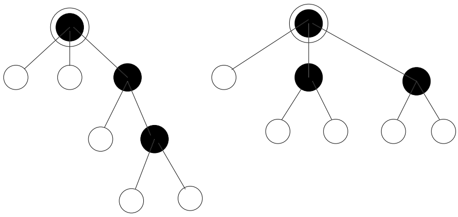


Fig. 3.11. Rooted patterns for the pattern in Figure 3.5

where $a = a_0 + \cdots + a_{10}$. The function r_0 might also be interpreted as a catch-all function for the "uninteresting" subtrees – just a root x with an unspecified number of planted trees attached, except the ones we handle differently, namely the cases $d \in \bar{D} = \{3\}$. The generating function $r = r(x, u)$ for rooted trees is then given by

$$l = l_0 + \frac{1}{6}xb_0^3 + \frac{1}{2}x \sum_{1\le i\le 3} b_0^2 b_i u^{i-1} + \frac{1}{2}x \sum_{1\le i,j\le 3} b_0 b_i b_j u^{i+j-1} + \frac{1}{6}x \sum_{1\le i,j,k\le 3} b_i b_j, b_k u^{i+j+k}$$

where the b_i are defined in (3.33).

As above we have $t_{n,m} = l_{n,m}/n$, where $t_{n,m}$ denotes the number of unrooted trees with n nodes and exactly m occurrences of the pattern $\mathcal{M}$. This relation is sufficient for our purposes. It is also possible to express the corresponding generating function $t(x, u)$. In a way similar as before, we can define the number of additional occurrences $\hat{K}(i, j)$ of the pattern $\mathcal{M}$ that appear by constructing an unrooted tree from two planted trees of the class a_i and a_j by identifying the additional edges at their planted roots and discarding the phantom nodes. For our example we get

$$t(x, u) = l(x, u) - \frac{1}{2}a(x, u)^2 - \frac{1}{2} \sum_{1\le i,j\le 3} b_i(x, u)b_j(x, u)(u^{i+j-2} - 1).$$

3.3.4 Asymptotic Behaviour

In order to complete the proof of Theorem 3.18 we want to apply Theorem 2.35. Several assumptions have already been checked. The only missing point is the existence of a non-negative solution $(x_0, \mathbf{a}_0)$ of the system

$$\mathbf{a} = \mathbf{F}(x, \mathbf{a}, 1), \tag{3.36}$$

$$0 = \det(\mathbf{I} - \mathbf{F_a}(x, \mathbf{a}, 1)), \tag{3.37}$$

where (3.36) is the system of functional equations of Lemma 3.19 and $\mathbf{F_a}$ is the Jacobian matrix of $\mathbf{F}$. Since the sum of all unknown functions $a(x, u)$ is known for $u = 1$:

$$a(x,1) = a(x) = \sum_{n\ge 1} n^{n-1}\frac{x^n}{n!} = 1 - \sqrt{2}\sqrt{1-ex} + \cdots,$$

it is not unexpected that $x_0 = 1/e$.

Lemma 3.21. *There exists a unique non-negative solution* $(x_0, \mathbf{a}_0)$ *of System (3.36)–(3.37),* $x_0 = 1/e$ *and the components of* $\mathbf{a}_0$ *are polynomials (with rational coefficients) in* $1/e$.

Proof. For a proof, set $u = 1$ and consider the solution $\mathbf{a}(x,1) = (a_0(x,1), \ldots, a_{L-1}(x,1))$. Since the dependency graph is strongly connected, it follows that all functions $a_j(x,1)$ have the same radius of convergence which has to be $x_0 = 1/e$, and all functions are singular at $x = x_0$. Since $0 \le a_j(x,1) \le p(x,1) < \infty$ for $0 \le x \le x_0$, it also follows that $a_j(x_0,1)$ is finite, and we have $\mathbf{a}(x_0,1) = \mathbf{F}(x_0, \mathbf{a}(x_0,1), 1)$. If we had the inequality $\det(\mathbf{I} - \mathbf{F_a}(x_0, \mathbf{a}(x_0,1),1)) \ne 0$ then the implicit function theorem would imply the existence of an analytic continuation for $a_j(x,1)$ around $x = x_0$, which is, of course, a contradiction. Thus, the system above has a unique solution.

To see that the components $\bar{a}_0, \ldots \bar{a}_L$ (with $\bar{a}_i = a_i(1/e,1)$) of $\mathbf{a}_0$ are polynomials in $1/e$ we will construct the partition $\mathcal{A} = \{a_0, a_1, \ldots, a_L\}$ on which the system of equations (3.36)–(3.37) is based by refining the trivial partition consisting of only one class p step by step. The recursive description of this trivial partition is given by the formal equation $p = x\sum_{i\ge 0} p^i$ and the solution of the corresponding equation $p = x\exp(p)$ for the generating function p (denoted by the same symbol p) is given by $(x_0, \bar{p}) = (1/e, 1)$, with $\bar{p}$ clearly a (constant) polynomial in $1/e$. Now let $D = \{d_1, \ldots, d_s\}$ ($s \in \mathbb{N}$) again denote the set of out-degrees that occur in the planted patterns. We will refine p by introducing for each $d_i \in D$ a class a_i consisting of all trees of root out-degree d_i together with the class a_0 of trees with root out-degree not in D. The partition $\{a_0, a_1, \ldots, a_s\}$ has the recursive description

$$\begin{aligned} a_0 &= x \sum_{j\in\mathbb{N}\setminus D} (a_0 \oplus a_1 \oplus \cdots \oplus a_s)^j \\ a_i &= x(a_0 \oplus a_1 \oplus \cdots \oplus a_s)^{d_i} \qquad (i = 1, \ldots, s) \end{aligned} \tag{3.38}$$

and the solution of the corresponding system of equations is given by

$$\begin{aligned} x_0 &= \frac{1}{e}, \\ \bar{a}_i &= \frac{1}{d_i! e} \quad (i = 1, \ldots, s), \\ \bar{a}_0 &= 1 - (\bar{a}_1 + \cdots + \bar{a}_s), \end{aligned} \tag{3.39}$$

thus again polynomials in $1/e$. We continue by refining this last partition by introducing classes $c_1, \ldots, c_u$ ($u \in \mathbb{N}$) for each term at the right hand side of (3.38) after expanding. Such a class c_j is of the form $c_j = x a_0^{l_0^{(j)}} a_1^{l_1^{(j)}} \cdots a_s^{l_s^{(j)}}$

with natural numbers $l_i^{(j)}$, $i = 0, \ldots, s$. We get a new partition $\{a_0, c_1, \ldots, c_u\}$ which has a recursive description by construction (because we can replace the a_i by disjoint unions of certain c_j). The solution of the corresponding system of equations for the generating functions is given by

$$\begin{aligned} x_0 &= \frac{1}{e}, \\ \bar{c}_j &= \frac{1}{e} \frac{1}{l_0^{(j)}! l_1^{(j)}! \cdots l_s^{(j)}!} \bar{a}_0^{l_0^{(j)}} \bar{a}_1^{l_1^{(j)}} \cdots \bar{a}_s^{l_s^{(j)}} \qquad (j = 1, \ldots, u), \\ \bar{a}_0 &= 1 - (\bar{a}_1 + \cdots + \bar{a}_s) \end{aligned}$$

with the $\bar{a}_i$ of (3.39). Thus the $\bar{c}_j$ are again polynomials in $1/e$. By continuing this procedure until level h (i.e. performing the refinement step h times) we end up with the partition $\mathcal{A}$ and we see that the solution for the corresponding system of equations consists of polynomials in $1/e$, which ends the proof of Lemma 3.21.

Note that there is a close link to Galton-Watson branching processes. Let $p_k = \frac{1}{k!e}$ denote the offspring distribution. Now we interpret a class a_i as the class of branching processes for which the (non-plane) branching structure at the start of the process is the same as the tree structure at the root of the trees in a_i. By arguments of probabilistic independence it follows that the solution $\bar{a}_i$, $i = 0, \ldots, L$ from the system of equations (3.36) corresponds to the probabilities q_i, $i = 0, \ldots, L$ that a branching process is in the class a_i, $i = 0, \ldots, L$. To see this, let the equation of class a_i be given by

$$a_i = x \frac{1}{l_0! \ldots l_L!} a_0^{l_0} \cdots a_L^{l_L} = \frac{x}{n!} \frac{n!}{l_0! \ldots l_L!} a_0^{l_0} \cdots a_L^{l_L}, \qquad (n = l_0 + \cdots + l_L).$$

Now the probability that a branching process is in class a_i is the probability $\frac{x_0}{n!} = \frac{1}{en!}$ that the first node has degree n, and that l_i of the branching processes starting at the children of this first node are in the class a_i $(0 \le i \le L)$. This yields a factor $a_0^{l_0} \cdots a_L^{l_L}$. Since the left-to-right-order of this child-processes is not relevant, we also get a factor $\frac{n!}{l_0! \ldots l_L!}$ to discount different embeddings into the plane. However we can also calculate the q_i by elementary probabilistic considerations. Therefore we consider the probabilities $p_{d_1}, \ldots, p_{d_s}$ of the out-degrees $d_1, \ldots, d_s$ of the pattern nodes and $r = 1 - p_{d_1} - \cdots - p_{d_s}$ for the other out-degrees. Now we get the probabilities $q_i (= \bar{a}_i)$ by multiplication of the p_{d_k} and r corresponding to the branching degrees of the processes in each class a_i and discounting different embeddings into the plane by some multiplicative factors which follow from symmetry considerations. So again we see that $\mathbf{a}_0$ consists of polynomials in $1/e$.

Consequently, it follows from Theorem 2.35 that the numbers $l_{n,m}$ have a Gaussian limiting distribution with mean and variance which are proportional to n. Since $t_{n,m} = l_{n,m}/n$, we get exactly the same rule for unrooted trees.

We now solve the system of equations (3.36)–(3.37) for the example pattern, where we apply the relation to Galton-Watson processes as described above. Recall that $x_0 = 1/e$. We only have to consider the probabilities $p = 1/(2e)$ for nodes with out-degree 2 and $q = 1 - 1/(2e)$ for the other nodes. For example we get $\bar{a}_4 = a_4(1/e, 1) = 2qp^3 = \frac{2e-1}{16e^5}$. The factor 2 comes from the fact that the two subtrees of the root may be exchanged (see Figure 3.9). The other classes can be treated similarly and we find:

$$\begin{aligned} a(1/e,1) &= 1, & a_5(1/e,1) &= (2e-1)^4/(128e^7),\\ a_0(1/e,1) &= (2e-1)/(2e) & a_6(1/e,1) &= (2e-1)^3/(32e^7),\\ a_1(1/e,1) &= (2e-1)^2/(8e^3) & a_7(1/e,1) &= (2e-1)^2/(64e^7),\\ a_2(1/e,1) &= (2e-1)^3/(16e^5) & a_8(1/e,1) &= (2e-1)^2/(32e^7),\\ a_3(1/e,1) &= (2e-1)^2/(8e^5) & a_9(1/e,1) &= (2e-1)/(32e^7),\\ a_4(1/e,1) &= (2e-1)/(16e^5) & a_{10}(1/e,1) &= 1/(128e^7). \end{aligned} \tag{3.40}$$

We now give a formula for the mean value μ.

Lemma 3.22. *Let $x_0 = 1/e$ and $\mathbf{a}_0$ be given by Lemma 3.21 and let $P_j(\mathbf{y}, u)$ $(1 \le j \le L)$ be the polynomials of Lemma 3.19, with $\mathbf{y} = (y_0, \ldots, y_L)$. Then μ (of Theorem 3.18) is a polynomial in $1/e$ with rational coefficients and given by*

$$\mu = \frac{1}{e}\sum_{j=1}^{L} \frac{\partial P_j}{\partial u}(\mathbf{a}_0, 1). \tag{3.41}$$

Proof. Let $\mathbf{a} = \mathbf{F}(x, \mathbf{a}, u)$ be the system of functional equations of Lemma 3.19. We recall that (2.56) provides a formula for the mean constant:

$$\mu = \frac{1}{x_0}\frac{\mathbf{b}^{\mathrm{T}}\mathbf{F}_u(x_0, \mathbf{a}_0, 1)}{\mathbf{b}^{\mathrm{T}}\mathbf{F}_x(x_0, \mathbf{a}_0, 1)}. \tag{3.42}$$

Here $\mathbf{b}^T$ denotes a positive left eigenvector of $\mathbf{I} - \mathbf{F_a}$, which is unique up to scaling.

Since

$$\mathbf{F}(x, \mathbf{a}, u) = \begin{pmatrix} x\left(e^{a_0+\cdots+a_L} - \sum_{j=1}^{L} P_j(\mathbf{a}, 1)\right) \\ xP_1(\mathbf{a}, u) \\ xP_2(\mathbf{a}, u) \\ \vdots \\ xP_L(\mathbf{a}, u) \end{pmatrix},$$

we get, after denoting $\frac{\partial P_i}{\partial a_j}$ with P_{i,a_j},

$$\mathbf{F_a} = x\begin{pmatrix} e^{a_0+\cdots+a_L} - \sum_{j=1}^{L} P_{j,a_0} & \cdots & e^{a_0+\cdots+a_L} - \sum_{j=1}^{L} P_{j,a_L} \\ P_{1,a_0} & \cdots & P_{1,a_L} \\ \vdots & & \vdots \\ P_{L,a_0} & \cdots & P_{L,a_L} \end{pmatrix}.$$

Since $a_0(x_0,1)+\cdots+a_L(x_0,1) = a(x_0,1) = 1$, we have $x_0 e^{a_0(x_0,1)+\cdots a_L(x_0,1)} = 1$. Consequently the sum of all rows of $\mathbf{F_a}$ equals $(1,1,\ldots,1)$ for $x = x_0 = 1/e$. Thus, denoting the transpose of a vector v by v^{T}, the vector $\mathbf{b}^{\mathrm{T}} = (1,1,\ldots,1)$ is the unique positive left eigenvector of $\mathbf{I} - \mathbf{F_a}$, up to scaling.

It is now easy to check that

$$x_0 \mathbf{b}^{\mathrm{T}} \mathbf{F}_x(x_0, \mathbf{a}_0, 1) = \frac{1}{e} e^{a_0(x_0,1)+\cdots a_L(x_0,1)} = 1$$

and that

$$\mathbf{b}^{\mathrm{T}} \mathbf{F}_u(x_0, \mathbf{a}_0, 1) = \frac{1}{e} \sum_{j=1}^{L} P_{j,u}(\mathbf{a}_0, 1).$$

The fact that μ is a polynomial in $1/e$ is now a direct consequence from the observation that $\mathbf{a}_0$ consists of polynomials in $1/e$, and the fact that the coefficients are rational follows from the property that $\mathbf{F}(x, \mathbf{a}, u)$ has rational coefficients.

This completes the proof of Theorem 3.18. We point out here that it is also possible to prove that the constant σ^2 can be represented as a polynomial in $1/e$ with rational coefficients. However, the proof is very involved (see [40]).

4

The Shape of Galton-Watson Trees and Pólya Trees

Galton-Watson trees or simply generated trees are random trees that are obtained by Galton-Watson branching processes conditioned on the total progeny.[1] This is in fact a very natural and general concept of random trees that includes several kinds of combinatorial tree models like binary trees, planted plane trees, labelled rooted trees, etc. (compare with Section 1.2.7).

One important point is that Galton-Watson trees have a continuous limiting object, the so-called *continuum random tree*, by scaling the distances to the root by $1/\sqrt{n}$. This concept has been introduced by Aldous [2, 3, 4] and further developed by Duquesne and Le Gall [72, 73, 74].

First we present an introduction to these limiting objects (for more details we refer to the recent surveys by Le Gall [135, 136]).

The main focus of this chapter is to characterise the so-called profile of Galton-Watson trees, that is, the number of nodes of given distance to the root. It is a very natural measure of the *shape* of a tree. Interestingly, the profile process can be approximated by the total local time of a Brownian excursion of duration 1. Since the profile is not a continuous functional of a random tree (more precisely of its depth-first search process), its limiting behaviour cannot be deduced from the structure of the continuum random tree.

The approach that we present here – that is based on combinatorics with the help of generating functions – is probably not the most elegant one, since it is quite technical. However, it can be generalised to unlabelled rooted trees (Pólya trees) that cannot be represented as conditioned Galton-Watson trees.

We also comment on the height of conditioned Galton-Watson trees and Pólya trees that are of order $\sqrt{n}$.

[1] The notion Galton-Watson tree (or more precisely, conditioned Galton-Watson tree) has become standard in the literature. Therefore we restrict ourselves to this notion throughout this chapter and do not use the equivalent notion of simply generated trees.

4.1 The Continuum Random Tree

4.1.1 Depth-First Search of a Rooted Tree

In order to analyse a plane rooted tree T we consider the so-called *depth-first search*. It can be described as a walk $(v(i), 0 \le i \le 2n)$ around the vertices $V(T)$. For convenience we add a planted node where we start and terminate: $v(0) = v(2n)$. Next let $v(1)$ be the root. Now we proceed inductively. Given $v(i)$ choose (if possible) the first edge at $v(i)$ (in the ordering) leading away from the root which has not already been traversed, and let $(v(i), v(i+1))$ be that edge. If this is not possible, let $(v(i), v(i+1))$ be the edge from $v(i)$ leading towards the root. This walk terminates with $v(2n-1)$, that equals (again) the root, and with $v(2n)$, that is the planted node.

The search depth $x(i)$ is now defined as the distance of $v(i)$ to the planted node which is precisely the distance to the root plus 1, $x(0) = x(2n) = 0$. For non-integer i we use linear interpolation and thus $x(t)$, $0 \le t \le 2n$, can be considered as a continuous *excursion* (see Figure 4.1).

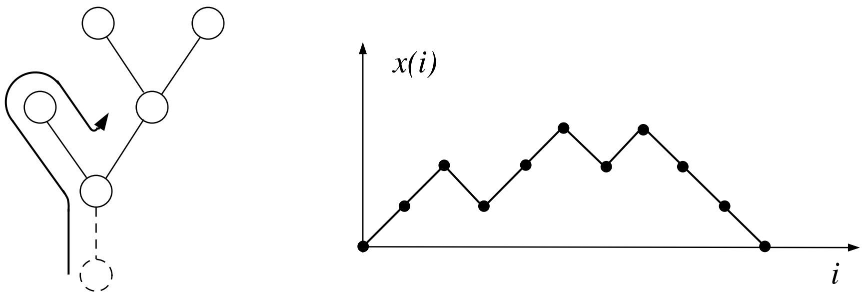


Fig. 4.1. Depth-first search of a rooted tree

Note that the height h and the path length I (that is the sum of all distances to the root) can be expressed with help of $x(i)$:

$$h = \max_{0 \le i \le 2n} x(i) - 1 \quad \text{and} \quad I = \frac{1}{2}\sum_{i=0}^{2n} x(i) - \frac{n}{2}. \tag{4.1}$$

Observe that the above procedure provides also a bijection between planted plane trees and so-called *Dyck paths*, that is, discrete excursions starting and terminating at 0 and being positive in-between, where each step is either up or down. In particular, trees of size n correspond to paths of length $2n$.

In the uniform random model every tree of size n and also every Dyck path of length $2n$ is equally likely. This also corresponds to the *standard random walk* with probability $\frac{1}{2}$ going one step up resp. one step down conditioned on $x(0) = x(2n) = 0$ and $x(j) > 0$ for $1 \le j \le 2n-1$.

4.1.2 Real Trees

Suppose that a (rooted or unrooted) tree T is embedded into the plane such that the edges are straight lines (or rectifiable curves) that only intersect at their incident vertices. Let $\mathcal{T} \subseteq \mathbb{R}^2$ denote the actually embedded set. Then $\mathcal{T}$ can be considered as a (compact) metric space $(\mathcal{T}, d)$, where the distance $d(x, y)$ between two points $x, y \in \mathcal{T}$ is defined as the shortest length of a curve $\gamma \subseteq \mathcal{T}$ that connects x and y. Actually, the range of every curve of length $d(x, y)$ that connects x and y coincides with γ.

These properties motivate the definition of an (abstract) *real tree.*

Definition 4.1. *A metric space $(\mathcal{T}, d)$ is a real tree if the following two properties hold for every $x, y \in \mathcal{T}$.*

1. *There is a unique isometric map $h_{x,y} : [0, d(x, y)] \to \mathcal{T}$ such that $h_{x,y}(0) = x$ and $h_{x,y}(d(x, y)) = y$.*
2. *If q is a continuous injective map from $[0, 1]$ into $\mathcal{T}$ with $q(0) = x$ and $q(1) = y$ then*
$$q([0, 1]) = h_{x,y}([0, d(x, y)]).$$

A rooted real tree $(\mathcal{T}, d)$ is a real tree with a distinguished vertex $r = r(\mathcal{T})$ called the root.

Of course, the embedded tree from above is a real tree in this sense. If we interpret combinatorial trees as real trees it is convenient to choose the embedding in a way that adjacent vertices are connected by lines of length 1 so that the usual distance of nodes is preserved.

Next we deal with the problem of equivalent real trees resp. with the problem of measuring the difference between real trees. For simplicity we assume now that all real trees that we consider are compact and rooted.

We say that two real trees $(\mathcal{T}_1, d_1)$, $(\mathcal{T}_2, d_2)$ are equivalent if there is a root-preserving isometry that maps $\mathcal{T}_1$ onto $\mathcal{T}_2$. We denote by $\mathbb{T}$ the set of all equivalence classes of rooted compact real trees.

It is easy to measure the difference of two real trees (if they are both contained in the same metric space) by using the Hausdorff distance

$$\delta_{\mathrm{Haus}}(X, Y) = \max\left\{ \sup_{x\in X} \inf_{y\in Y} d(x, y),\ \sup_{y\in Y} \inf_{x\in X} d(x, y) \right\}.$$

Thus, it is natural to define the so-called *Gromov-Hausdorff distance* $d_{\mathrm{GH}}(\mathcal{T}_1, \mathcal{T}_2)$ of two real trees $(\mathcal{T}_1, d_1)$, $(\mathcal{T}_2, d_2)$ by

$$d_{\mathrm{GH}}(\mathcal{T}_1, \mathcal{T}_2) = \inf\Big(\max\big\{ \delta_{\mathrm{Haus}}(\phi_1(\mathcal{T}_1), \phi_2(\mathcal{T}_2)),\ \delta(\phi_1(r(\mathcal{T}_1)), \phi_2(r(\mathcal{T}_2))) \big\} \Big),$$

where the infimum is taken over all isometric embeddings $\phi_1 : \mathcal{T}_1 \to E$ and $\phi_2 : \mathcal{T}_2 \to E$ of $\mathcal{T}_1$ and $\mathcal{T}_2$ into a common metric space (E, δ). It is immediately clear that this distance only depends on the equivalence classes of $\mathcal{T}_1$ resp. $\mathcal{T}_2$. Moreover, d_{GH} defines a metric on $\mathbb{T}$ and we have the following important properties (see [76]):

Theorem 4.2. *The metric space* $(\mathbb{T}, d_{\mathrm{GH}})$ *is complete and separable, that is, it is a Polish space.*

Real trees are closely related to continuous functions that represent excursions. Let $g : [0,\infty) \to [0,\infty)$ be a non-zero continuous function with compact support such that $g(0) = 0$. For every $s, t \ge 0$ set

$$d_g(s,t) = g(s) + g(t) - 2 \inf_{\min\{s,t\} \le u \le \max\{s,t\}} g(u)$$

(compare with Figure 4.2). Then d_g is non-negative, symmetric and satisfies the triangle inequality. If we define an equivalence relation by

$$s \sim t \iff d_g(s,t) = 0$$

and if $\mathcal{T}_g = [0,\infty)/\sim$ denotes the system of equivalence classes then the space

$$(\mathcal{T}_g, d_g)$$

is a metric space.

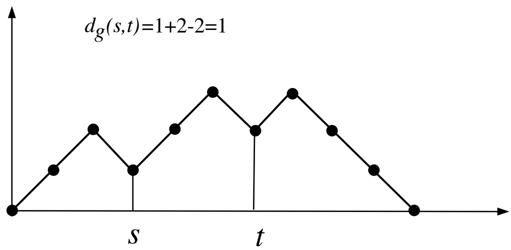


Fig. 4.2. Distance d_g

Actually we have the following property (see [74]):

Theorem 4.3. *The metric space* $(\mathcal{T}_g, d_g)$ *is a compact real tree.*

Usually the equivalence class corresponding to $t = 0$ is considered as the root so that $\mathcal{T}_g$ becomes a rooted real tree.

For example, let g be the excursion $x(t)$ given in Figure 4.1. Then this identification process can be seen as the inverse process of the depth-first search (compare with Figure 4.3).

Intuitively it is clear that the real trees $\mathcal{T}_{g_1}$ and $\mathcal{T}_{g_2}$ are close if g_1 and g_2 are close. This is actually true in a strong sense (see [135]).

Lemma 4.4. *The mapping* $g \mapsto \mathcal{T}_g$ *is continuous. More precisely, if* g_1, g_2 *are two functions from* $[0,\infty) \to [0,\infty)$ *with compact support and* $g_1(0) = g_2(0)$ *then*

$$d_{\mathrm{GH}}(\mathcal{T}_{g_1}, \mathcal{T}_{g_2}) \le \|g_1 - g_2\|_\infty,$$

where $\|\cdot\|_\infty$ *denotes the supremum norm.*

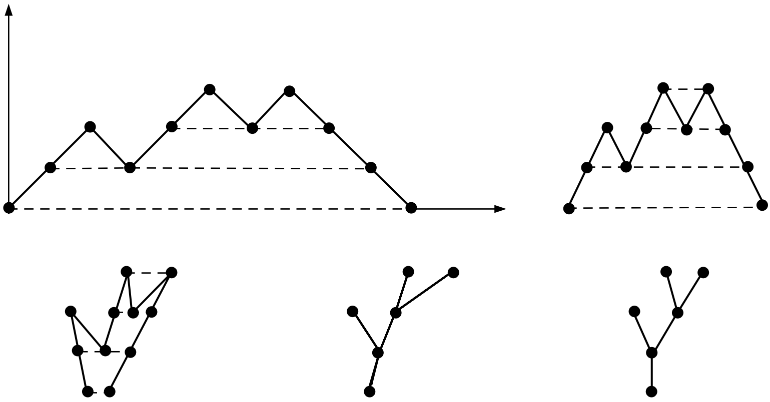


Fig. 4.3. Construction of a real tree $\mathcal{T}_g$

4.1.3 Galton-Watson Trees and the Continuum Random Tree

The space $(\mathcal{T}_g, d_g)$ is a Polish space, that is, a complete separable metric space and therefore perfectly equipped for applying standard probability theory, in particular, weak convergence of measures (see [17, 18]).

Suppose that (S,d) is a Polish space and let $X_n : \Omega \to S$ be a sequence of S-valued random variables. Then X_n converges weakly to a random variable $X : \Omega \to S$, if and only if

$$\lim_{n\to\infty} \mathbb{E}\, F(X_n) = \mathbb{E}\, F(X) \tag{4.2}$$

holds for all continuous bounded functions $F : S \to \mathbb{R}$. As usual we will denote this by $X_n \xrightarrow{d} X$.

Using this definition it is immediately clear that if $G : S \to S'$ is a continuous function between Polish spaces and if we know that $X_n \xrightarrow{d} X$ for S-valued random variables then we automatically obtain $G(X_n) \xrightarrow{d} G(X)$.

We now introduce special random real trees. First we use the standard Brownian excursion $(e(t), 0 \le t \le 1)$ of duration 1 that is a stochastic process on non-negative continuous functions of support $[0,1]$ and zeros at $t = 0$ and $t = 1$ (see Figure 4.4[2]).

There are several possibilities to define the Brownian excursion. For example, it can be seen as a (rescaled) Brownian motion between two zeros. Alternatively it can be defined as a Markov process with $\mathbb{P}\{e(0) = 0\} = \mathbb{P}\{e(1) = 0\} = 1$ that is uniquely given by the densities of the one- and two-dimensional distributions:

[2] This figure was produced by Jean-François Marckert.

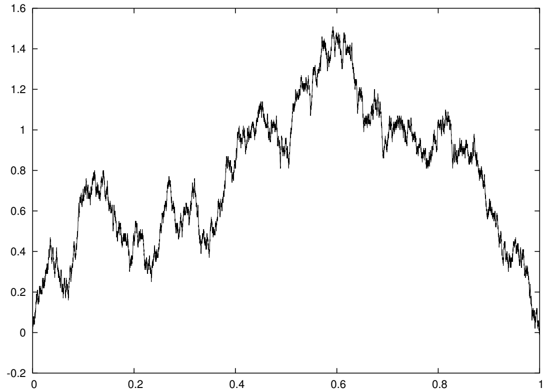


Fig. 4.4. Brownian excursion

$$\mathbb{P}\{e(t) = x\} = \frac{2x^2}{\sqrt{2\pi t^3(1-t)^3}} \exp\left(-\frac{x^2}{2t(1-t)}\right) dx, \tag{4.3}$$

$$\begin{aligned}\mathbb{P}\{e(t_1) = x,\, e(t_2) = y\} &= \frac{xy}{\pi\sqrt{t_1^3(t_2-t_1)t_2^3}} \exp\left(-\frac{x^2}{2t_1} - \frac{y^2}{2(1-t_2)}\right) \\ &\quad\times \left(\exp\left(-\frac{(x-y)^2}{2(t_2-t_1)}\right) - \exp\left(-\frac{(x+y)^2}{2(t_2-t_1)}\right)\right) dx\, dy.\end{aligned} \tag{4.4}$$

Anyway it is a probability distribution on non-negative continuous functions g of compact support. Hence, the mapping $g \mapsto \mathcal{T}_g$ induces a corresponding probability distribution on real trees.

Definition 4.5. *Let* $(e(t), 0 \le t \le 1)$ *denote the standard Brownian excursion of duration* 1. *Then the random real tree* $\mathcal{T}_{2e}$ *is called Continuum Random Tree.*[3]

Next we consider Galton-Watson trees defined by their offspring distribution ξ and let T_n denote a random Galton-Watson tree of size n. For simplicity we assume that the offspring distribution is aperiodic, that is

$$\gcd\{j : \mathbb{P}\{\xi = j\} > 0\} = 1$$

or equivalently, the distribution is not concentrated on $d\mathbb{Z}$ for some $d > 1$. This condition assures that trees of arbitrary size will evolve. By using the depth-first search procedure these random trees induce random excursions $(X_n(t), 0 \le t \le 2n)$ of length $2n$. These excursions are again non-negative continuous functions g with compact support. Thus, the mapping $g \mapsto \mathcal{T}_g$ induces a class of random real trees that we denote by $\mathcal{T}_{X_n}$. If we assume that the variance $\sigma^2 := \mathbb{V}\mathrm{ar}\,\xi$ of the offspring distribution is finite then the

[3] The factor 2 in the $2e$ is just a convenient scaling factor.

expected height of a Galton-Watson tree is of order $\sqrt{n}$ (see Section 4.4). Hence it is natural to scale $X_n(t)$ and consequently the corresponding real trees by the factor $1/\sqrt{n}$ so that the resulting trees are of comparable size.

The main result in this context is an invariance principle by Aldous [4].

Theorem 4.6. *Suppose that the offspring distribution ξ of a Galton-Watson branching process is aperiodic and critical with finite variance $\sigma^2 = \mathbb{V}\mathrm{ar}\,\xi$. Then the scaled real trees*

$$\frac{\sigma}{\sqrt{n}}\,\mathcal{T}_{X_n}$$

corresponding to Galton-Watson trees of size n converge weakly to the continuum random tree $\mathcal{T}_{2\mathrm{e}}$.

There are also several extensions of Theorem 4.6, for example by Duquesne and Le Gall [72, 73, 74] if the variance $\mathbb{V}\mathrm{ar}\,\xi$ is infinite but ξ is in the domain of attraction of a stable law. However, for this purpose one has to introduce random real trees based on a stable Lévy process.

A natural way to prove Theorem 4.6 is to consider the depth-first search process $X_n(t)$ corresponding to Galton-Watson trees and to prove the following property that is also due to Aldous [4].

Theorem 4.7. *Suppose that the offspring distribution ξ of a Galton-Watson branching process is aperiodic and critical with finite variance $\sigma^2 = \mathbb{V}\mathrm{ar}\,\xi$. Then the depth-first search process $X_n(t)$ satisfies*

$$\left(\frac{\sigma}{2\sqrt{n}}X_n(2nt), 0 \le t \le 1\right) \xrightarrow{d} (e(t), 0 \le t \le 1). \tag{4.5}$$

In other words, the normalised depth-first search process is approximated by the Brownian excursion Obviously, Theorem 4.7 implies Theorem 4.6, since the mapping $g \mapsto \mathcal{T}_g$ is continuous.

We do not supply the original proof of this theorem here. There are several recent surveys on this subject (see [135, 136]). Nevertheless, in Section 4.2.8 we comment on an alternative approach that is based on generating functions and analytic methods.

One important issue of Theorems 4.6 and 4.7 is that they can be used to derive the limiting distribution of several parameters of trees. For example, let us consider the height H_n, the maximal distance between a node and the root.

Theorem 4.8. *Suppose that the offspring distribution ξ of a Galton-Watson branching process satisfies the assumptions of Theorem 4.6 and let H_n denote the height of the corresponding Galton-Watson trees. Then we have*

$$\frac{1}{\sqrt{n}}H_n \xrightarrow{d} \frac{2}{\sigma}\max_{0\le t\le 1} e(t).$$

This theorem was first proved by Flajolet and Odlyzko [82]. A local version is due to Flajolet, Gao, Odlyzko, and Richmond [80].

It is well known that the distribution function of $M = \max_{0 \le t \le 1} e(t)$ is given by

$$\mathbb{P}\{M \le x\} = 1 - 2\sum_{k=1}^{\infty}(4x^2k^2 - 1)e^{-2x^2k^2}$$

and the moments by

$$\mathbb{E}(M^r) = 2^{-r/2}r(r-1)\Gamma(r/2)\zeta(r),$$

where $\zeta(s)$ denotes the Riemann ζ-function (and where we make the convention $(r-1)\zeta(r) = 1$ for $r = 1$).

Thus, one might expect that

$$\mathbb{E}(H_n^r) \sim \left(\frac{2}{\sigma}\right)^r \mathbb{E}(M^r)\, n^{r/2}$$

for $r \ge 0$. In fact this is true if we additionally assume that the exponential moment $\mathbb{E}e^{\eta\xi}$ is finite for some $\eta > 0$. One possible way is to apply Theorem 4.15 (the assumptions can be verified by Proposition 4.41) or to use analytic methods (see Theorem 4.29).

Another corollary concerns the path length I_n of Galton-Watson trees, that is the sum of all distances to the root. Here the limiting relation (4.5) and the second part of (4.1) provide the following result.

Theorem 4.9. *Suppose that the offspring distribution ξ of a Galton-Watson branching process satisfies the assumptions of Theorem 4.6 and let I_n denote the path length of the corresponding Galton-Watson trees. Then we have*

$$\frac{1}{n^{3/2}} I_n \xrightarrow{d} \frac{2}{\sigma}\int_0^1 e(t)\,dt.$$

The moments of $I = \int_0^1 e(t)\,dt$ are determined by

$$\mathbb{E}(I^r) = K_r \frac{4\sqrt{\pi}r!}{\Gamma\left(\frac{3r-1}{2}\right)2^{r/2}},$$

where K_r is recursively given by

$$K_r = \frac{3r-4}{4}K_{r-1} + \sum_{j=1}^{r-1} K_j K_{r-j}, \qquad (r \ge 2),$$

with initial values $K_0 = -\frac{1}{2}$ and $K_1 = \frac{1}{8}$. The convergence of moments was first proved by Takács [200]. Louchard [140] identified then the relation of the limiting distribution to the Brownian excursion area.

4.2 The Profile of Galton-Watson Trees

Let T be a rooted tree. We denote by $L_T(k)$ the number of nodes at distance k from the root. The sequence $(L_T(k))_{k\ge 0}$ is called the profile of T. If T is a random tree of size n, for example a conditioned Galton-Watson tree (or equivalently a simply generated tree) then we denote the profile by $(L_n(k))_{k\ge 0}$, which is now a sequence of random variables (see Figure 4.5, where y-axis is used as the *time axis*).

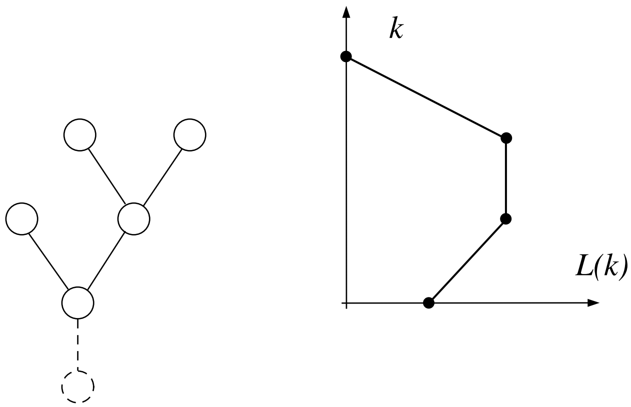


Fig. 4.5. Profile of a rooted tree

It is convenient to view this random sequence as a stochastic process. For non-integer arguments we define the values by linear interpolation, that is, we have

$$L_n(t) = (\lfloor t\rfloor + 1 - t)L_n(\lfloor t\rfloor) + (t - \lfloor t\rfloor)L_n(\lfloor t\rfloor + 1), \quad t \ge 0.$$

$(L_n(t), t \ge 0)$ is then a stochastic process and is called the profile process of a random tree. By definition, the sample paths of the profile are continuous functions on $[0,\infty)$.

There is a close relation between the profile $L_n(t)$ and the depth first search process $X_n(t)$. For integers k we have by definition

$$\frac{1}{2}\int_0^{2n} \mathbf{1}_{[0,k]}(X_n(t))\,dt = \sum_{\ell<k} L_n(\ell).$$

Moreover, suppose that f is (uniformly) continuous and bounded on $[0,\infty)$, then (4.5) implies

$$\frac{1}{n}\sum_{\ell\ge 0} f\left(\frac{\sigma\ell}{2\sqrt{n}}\right) L_n(\ell) = \frac{1}{2n}\int_0^{2n} f\left(\frac{\sigma}{2\sqrt{n}}X_n(t)\right)dt + o(1)$$
$$= \int_0^1 f\left(\frac{\sigma}{2\sqrt{n}}X_n(2nt)\right)dt + o(1) \qquad (4.6)$$
$$\xrightarrow{d} \int_0^1 f(e(t))\,dt.$$

In other words, the occupation measure of the scaled depth-first search converges to the occupation measure of $e(t)$.

It is natural to ask whether the occupation measure of $e(t)$ has a (random) density. Actually such a density

$$l(t) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_0^1 \mathbf{1}_{[t,t+\epsilon]}(e(s))\,ds$$

exists (almost surely) and constitutes a continuous stochastic process $(l(t),\, t \ge 0)$ that is also called *local time* local time of a stochastic process of the Brownian excursion (of duration 1).

One also expects that the (normalised) profile process is approximated by $l(t)$. Since the mapping that assigns the local time l to an excursion e is not continuous, we cannot apply (4.5) resp. Theorem 4.6 to obtain a weak limit theorem for the profile process. Nevertheless, such a property still holds (compare also with the discussion following Theorem 4.17).

Theorem 4.10. *Suppose that the offspring distribution ξ of a Galton-Watson branching process is aperiodic and critical with finite variance $\sigma^2 = \mathbb{V}\mathrm{ar}\,\xi$. Then we have (in $\mathrm{C}[0,\infty)$)*

$$(l_n(t),\, t \ge 0) \xrightarrow{d} \left(\frac{\sigma}{2} l\left(\frac{\sigma}{2}t\right),\, t \ge 0\right),$$

where

$$l_n(t) = \frac{1}{\sqrt{n}} L_n(t\sqrt{n}), \quad t \ge 0,$$

denotes the profile process scaled by $\sqrt{n}$.

Theorem 4.10 was conjectured by Aldous [3, Conjecture 4], and then proved by Drmota and Gittenberger [66], by Kersting [124] and Pitman [170]. In what follows we present a detailed proof of Theorem 4.10 that is based on the methods of [66]. The proof is analytically very involved and not as elegant as, for example, Pitman's approach that is based on stochastic differential equations. However, it seems that this analytic approach is, at the moment, the only way to generalise to Pólya trees that we discuss in Section 4.3.

By the way, originally Aldous [3] formulated his conjecture in terms of the step function process $\frac{1}{\sqrt{n}}L_n(\lfloor t\sqrt{n}\rfloor)$ (in the space $\mathrm{D}[0,\infty)$). However, it is not difficult to show that both properties are equivalent. The setting in $C[0,\infty)$ is technically easier. Therefore we will only work with continuous processes.

Historically the first investigations of the profile of random trees seem to go back to [193], who derived explicit formulas for the distribution of the size of one level. Further papers mainly deal with simply generated trees as defined by [151]. Kolchin (see [131, 132]) related the level size distributions to distributions occurring in particle allocation schemes.

Another interesting tree parameter is the *width*

$$W = \max_{k\ge 0} L(k) = \max_{t\ge 0} L(t),$$

that is the maximal number of nodes in a level. Since the maximum is a continuous functional, Theorem 4.10 provides a distributional result for the width W_n of Galton-Watson trees.

Theorem 4.11. *Under the assumption of Theorem 4.10 we have*

$$\frac{1}{\sqrt{n}} W_n \xrightarrow{d} \frac{\sigma}{2} \sup_{0\le t\le 1} l(t).$$

A corresponding result for convergence of moments seems to be more difficult to obtain. Nevertheless first results on the expected width are due Odlyzko and Wilf [166], later Chassaing and Marckert [34, 33] and Drmota and Gittenberger [67] settled the general case.

Note that the distribution of $\sup_{t\ge 0} l(t)$ is the same as that of $2\sup_{0\le t\le 1} e(t)$ (see [15] or [3]).

There are several extensions of Theorem 4.10. We just state one that is related to the leaves of Galton-Watson trees. Let $\hat{L}_T(k)$ denote the number of leaves at distance k from the root and $\hat{L}_n(k)$ the random variable we get, if T is a random Galton-Watson tree. As above $\hat{L}_n(t)$ denotes the linear interpolated function of the sequence $\hat{L}_n(k)$. Then we have in $C[0,\infty)$

$$(\hat{l}_n(t), t\ge 0) \xrightarrow{d} \left(\frac{\phi_0}{\phi(\tau)} \frac{\sigma}{2} l\left(\frac{\sigma}{2} t\right), t \ge 0 \right),$$

where

$$\hat{l}_n(t) = \frac{1}{\sqrt{n}} \hat{L}_n(\sqrt{n}\, t).$$

Interestingly, a limiting result of the form given in Theorem 4.10 seems to be quite universal. The property for $\hat{L}_n(t)$ is only a first indication in the framework of Galton-Watson trees. However, there are also other kinds of trees that have such a limiting shape. In what follows we will in particular discuss unlabelled rooted trees (Pólya trees). The natural approach to unlabelled rooted trees is to use generating functions; and the proof that the profile of unlabelled rooted trees is approximated by $l(t)$ is based on the analysis of properly defined generating functions. Therefore we have decided to present also a generating function proof of Theorem 4.10, although it is not the most general one. In particular, we assume here that the offspring distribution has finite exponential moments (which makes the analysis easier, compare with Remark 3.7). Nevertheless, we need a proper probabilistic framework in order to prove weak convergence of stochastic processes (see Theorem 4.14).

4.2.1 The Distribution of the Local Time

There are several interesting integral representations of densities and distribution functions of the local time $l(t)$ that we collect here. Some of them will be actually used in the proof of Theorem 4.10.

The one dimensional density of $(l(t); t \ge 0)$ at $t = \rho$ is well studied. There are several representations available in the literature: Using the theory of branching processes Kennedy [123] and Kolchin [132, Theorem 2.5.6] obtained

$$f_\rho(x) = \frac{x}{4}\int_0^1 (1-s)^{-\frac{3}{2}} e^{-\frac{x^2\rho^2}{8(1-s)}} g_{2\rho}\left(\frac{x}{2}, s\right)\, ds, \tag{4.7}$$

where $g_r(z,s)$ is the density of a distribution given by its characteristic function:

$$\psi_r(\vartheta_1,\vartheta_2) = \left[\frac{\sinh(r\sqrt{-2i\vartheta_2})}{r\sqrt{-2i\vartheta_2}} - i\vartheta_1\left(\frac{\sinh(r\sqrt{-i\vartheta_2/2})}{r\sqrt{-i\vartheta_2/2}}\right)^2\right]^{-1}.$$

Takács [200] calculated this density by means of a generating function approach

$$f_\rho(x) = 2\sum_{j\ge 1}\sum_{k=1}^{j}\binom{j}{k} e^{-(x+2\rho j)^2/2}\frac{(-x)^k}{(k-1)!}H_{k+2}(x+2\rho j). \tag{4.8}$$

$H_k(z)$ are the Hermite polynomials defined by

$$H_k(z) = (-1)^k e^{z^2/2}\frac{d^k}{dz^k}e^{-z^2/2}.$$

Knight [127] worked with the Brownian excursion and obtained

$$f_\rho(x) = 2^{-1/2}\pi^{5/2}\rho^{-3}\int_0^1 f^*\left(\frac{\pi^2(1-s)}{2\rho^2}\right) h(s,x)\, ds \tag{4.9}$$

with

$$f^*(z) = 4\sqrt{2}\pi^3\sum_{k\ge 1}k^2\frac{d}{dz}\left(\frac{e^{-k^2\pi^2/z}}{\sqrt{2\pi z^3}}\right)$$

and

$$h(s,x) = -\frac{1}{2\rho\sqrt{2\pi s}} \sum_{i\ge 0} \frac{1}{i!} \frac{d^{i-1}}{dx^{i-1}} \left(x^i \frac{d^2}{dx^2} e^{-2\rho^2(x+i)^2/s} \right),$$

where $\frac{d^{-1}}{dx^{-1}} = \left(\frac{d}{dx}\right)^{-1}$.

Getoor and Sharpe [90] also used a direct approach and derived a double Laplace transform of the density:

$$\int_0^\infty e^{-\alpha t} \int_0^t \frac{1}{\sqrt{2\pi s}} \int_{t-s}^\infty \frac{1}{\sqrt{2\pi r^3}} \mathbb{E}\left(e^{-\beta l(\rho/\sqrt{r})\sqrt{r}}\right) dr\, ds\, dt = \varphi_\rho(\alpha,\beta) \qquad (4.10)$$

where

$$\varphi_\rho(\alpha,\beta) = \frac{1}{\alpha} \frac{\sqrt{2\alpha} + \beta(1+e^{-2\rho\sqrt{2\alpha}})}{\sqrt{2\alpha} + \beta(1-e^{-2\rho\sqrt{2\alpha}})} - \frac{\beta\sqrt{2}}{(1+2\rho\beta)\alpha^{3/2}}.$$

This formula was also shown by Louchard [141] who found a considerably shorter proof via Kac's formula for Brownian functionals.

When studying $M/M/1$-queues Cohen and Hooghiemstra [42] got an integral representation for the above density:

$$f_\rho(x) = \frac{1}{i\sqrt{2\pi}} \int_\gamma \frac{-se^{-s}}{\sinh^2(\rho\sqrt{-2s})} \exp\left(-\frac{x}{\sqrt{2}} \frac{\sqrt{-s}e^{\rho\sqrt{-2s}}}{\sinh(\rho\sqrt{-2s})} \right) ds, \qquad (4.11)$$

where γ is the straight line $\{z \in \mathbb{C} : \Re z = -1\}$. They also derived the Laplace transform of the two dimensional densities of local and occupation time (compare also with the general result in Theorem 4.13).

Finally it should be mentioned that Hooghiemstra [103] found a direct proof for the equivalence of (4.9) and (4.11).

Remark 4.12 *Note that the expressions (4.7)–(4.11) are only representations of the continuous part of the local density. Obviously the local time density has a jump of magnitude* $\mathbb{P}\left\{\sup_{0\le t\le 1} e(t) < \rho\right\}$ *at* 0. *This quantity is given by*

$$\mathbb{P}\left\{ \sup_{0\le t\le 1} e(t) \le \rho \right\} = 1 - 2\sum_{k\ge 1} (4\rho^2k^2 - 1)e^{-2\rho^2k^2}. \qquad (4.12)$$

The proof of Theorem 4.10 is based on a precise analysis of finite dimensional projections (and on tightness estimates). For this purpose we need proper representations of the characteristic function of finite dimensional projections of $l(t)$ (compare with [66]).

Theorem 4.13. *Suppose that* $0 < \kappa_1 < \kappa_2 < \cdots < \kappa_d$ *are positive real number and let* $f_{\kappa_1,\ldots,\kappa_d}$ *be defined by*

$$f_{\kappa_1,\ldots,\kappa_d}(s,t_1,\ldots,t_d) = \qquad (4.13)$$
$$\Psi_{\kappa_1}(s, it_1 + \Psi_{\kappa_2-\kappa_1}(\ldots \Psi_{\kappa_{d-1}-\kappa_{d-2}}(s, it_{d-1} + \Psi_{\kappa_d-\kappa_{d-1}}(s, it_d))\cdots)$$

with

$$\Psi_\kappa(s,t) = \frac{t\sqrt{-2s}\, e^{-\kappa\sqrt{-2s}}}{\sqrt{-2s}\, e^{\kappa\sqrt{-2s}} - 2t\sinh\left(\kappa\sqrt{-2s}\right)}. \tag{4.14}$$

Then the characteristic function of the random vector $(l(\kappa_1),\ldots,l(\kappa_d))$ *is given by*

$$\mathbb{E}\, e^{t_1 l(\kappa_1)+\cdots+t_d l(\kappa_d)} = 1 + \frac{\sqrt{2}}{i\sqrt{\pi}} \int_\gamma f_{\kappa_1,\ldots,\kappa_d}(s,t_1,\ldots,t_d)\, e^{-s}\, ds, \tag{4.15}$$

where γ *is the straight line* $\{s \in \mathbb{C} : \Re s = -1\}$.

4.2.2 Weak Convergence of Continuous Stochastic Processes

Let $S = \mathrm{C}[0,1]$ be the Polish space of continuous functions on $[0,1]$. An S-valued random variable X (where the measure on S is defined on the Borel sets of S) is just a stochastic process $(X(t), 0 \le t \le 1)$. It is well known that the distribution of $(X(t), 0 \le t \le 1)$ is characterised by the finite dimensional distributions of the random vectors

$$(X(\kappa_1),\ldots,X(\kappa_d)),$$

where $d \ge 1$ and $0 \le \kappa_1 < \kappa_2 < \cdots < \kappa_d$.

However, weak convergence $X_n \xrightarrow{d} X$ of a sequence of stochastic processes cannot be characterised just by finite dimensional convergence. In particular, if

$$(X_n(\kappa_1),\ldots,X_n(\kappa_d)) \xrightarrow{d} (X(\kappa_1),\ldots,X(\kappa_d))$$

for all $d \ge 1$ and $0 \le \kappa_1 < \kappa_2 < \cdots < \kappa_d$ then this does not guarantee that $X_n \xrightarrow{d} X$. One has to assume additionally that X_n is tight. Recall that a sequence (X_n) of random variables in a metric space S is *tight* if for every $\varepsilon > 0$, there exists a compact subset $K \subseteq S$ such that $\mathbb{P}\{X_n \in K\} > 1 - \epsilon$ for every n. In a Polish space, a complete separable metric space, tightness is equivalent to relative compactness (of the corresponding distributions) by Prohorov's theorem [18, Theorems 6.1 and 6.2], [119, Theorem 16.3].

In our special case of $S = \mathrm{C}[0,1]$ (with the supremum norm) tightness means that

$$\lim_{a\to 0} \limsup_{n\to\infty} \mathbb{P}\{|X_n(0)| > a\} = 0 \tag{4.16}$$

and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{|s-t|\le\delta} |X_n(s) - X_n(t)| \ge \epsilon \right\} = 0 \tag{4.17}$$

for all $\epsilon > 0$. Informally, this says that, with high probability, $X_n(s)$ and $X_n(t)$ do not differ too much if s and t are close.

Property (4.17) cannot be checked easily in a direct way. However, there are sufficient moment conditions (4.19) so that we have the following theorem (compare with [18]).

Theorem 4.14. *Suppose that* $(X_n(t), 0 \le t \le 1)$ *is a sequence of stochastic processes and* $(X(t), 0 \le t \le 1)$ *a stochastic process on* C[0, 1] *that satisfies the following conditions.*

1. *For every* $d \ge 1$ *and* $0 \le \kappa_1 < \kappa_2 < \cdots < \kappa_d$ *we have*
$$(X_n(\kappa_1), \ldots, X_n(\kappa_d)) \xrightarrow{d} (X(\kappa_1), \ldots, X(\kappa_d)).$$
2. *There exists a constant* $C > 0$ *and an exponent* $\beta > 0$ *such that*
$$\mathbb{E}\left(|X_n(0)|^\beta\right) \le C \tag{4.18}$$
for all n.
3. *There exists a constant* $C > 0$ *and exponents* $\alpha > 1$ *and* $\beta > 0$ *such that*
$$\mathbb{E}\left(|X_n(t) - X_n(s)|^\beta\right) \le C|t-s|^\alpha \quad \text{for all } s, t \in [0, 1]. \tag{4.19}$$

Then
$$(X_n(t), 0 \le t \le 1) \xrightarrow{d} (X(t), 0 \le t \le 1).$$

By Markov's inequality it is clear that (4.18) implies (4.16). However, it is a non-trivial task to show that (4.19) is sufficient to prove (4.17). Actually, (4.19) implies
$$\mathbb{P}\left\{\sup_{|t-s|\le\delta} |X_n(t) - X_n(s)| \ge \epsilon\right\} \le K\frac{\delta^{\alpha-1}}{\epsilon^\beta} \tag{4.20}$$
for some constant K (see [18]). Note that the condition $\alpha > 1$ is essential for deducing (4.17).

Theorem 4.14 does not apply directly to processes in $C[0, \infty)$. However, weak convergence in $C[0, \infty)$ holds, if and only if we have weak convergence of the restrictions to the compact interval $[0, T]$ for all $T < \infty$ (see [18]). Hence, we can work with Theorem 4.14 in both cases.

Finally we want to mention that a condition of the kind (4.19) is actually quite strong. It is sufficient (but not necessary) to prove tightness, and it can also be used to prove convergence of moments. The problem is that the function $F(x) = x^r$ is unbounded so that $X_n \xrightarrow{d} X$ not necessarily implies $\mathbb{E}\, X_n^r \to \mathbb{E}\, X^r$ even if the expected values exist.

Theorem 4.15. *Suppose that a sequence of stochastic processes* $X_n = X_n(t)$, *defined on* $C[0, 1]$, *converges weakly to* $X = X(t)$. *Furthermore suppose that there exists* $s_0 \in [0, 1]$ *such that for all* $r \ge 0$
$$\sup_{n\ge 0} \mathbb{E}\left(|X_n(s_0)|^r\right) < \infty, \tag{4.21}$$
and that for every $\alpha > 1$ *there are* $\beta > 0$ *and* $C > 0$ *with (4.19).*

Let $F : \mathrm{C}[0,1] \to \mathbb{R}$ be a continuous functional of polynomial growth, that is, there exists $r \ge 0$ with

$$|F(y)| \le (1 + \|y\|_\infty)^r$$

for all $y \in \mathrm{C}[0,1]$. Then

$$\lim_{n\to\infty} \mathbb{E}\, F(X_n) = \mathbb{E}\, F(X). \tag{4.22}$$

For example, Theorem 4.15 implies that all finite dimensional moments of $X_n(t)$ converge to that of $X(t)$, that is, for all fixed $0 \le \kappa_1 < \kappa_2 < \cdots < \kappa_p$ and $r_1, r_2, \ldots, r_p \ge 0$

$$\lim_{n\to\infty} \mathbb{E}\left(X_n(\kappa_1)^{r_1} X_n(\kappa_2)^{r_2} \cdots X_n(\kappa_p)^{r_p}\right) = \mathbb{E}\left(X(\kappa_1)^{r_1} X(\kappa_2)^{r_2} \cdots X(\kappa_p)^{r_p}\right).$$

Furthermore, we have

$$\lim_{n\to\infty} \mathbb{E}\left(\max_{0\le t\le 1} X_n(t)\right)^r = \mathbb{E}\left(\max_{0\le t\le 1} X(t)\right)^r$$

and similarly for the integral $\int_0^1 X_n(t)\, dt$.

We present a short proof of Theorem 4.15 since it also sheds some light on the tightness condition (4.19).

The key of the proof of Theorem 4.15 is the following observation.

Lemma 4.16. *Suppose that $X_n(t)$ and $X(t)$ are stochastic processes satisfying the assumptions of Theorem 4.15. Then for every $\alpha > 1$ there exists a constant $K > 0$ such that for $\epsilon > 0$ and $0 < \delta < 1$*

$$\mathbb{P}\left\{\sup_{|s-t|\le\delta} |X_n(s) - X_n(t)| \ge \epsilon\right\} \le K \frac{\delta^{\alpha-1}}{\epsilon^\beta} \tag{4.23}$$

and consequently

$$\mathbb{E}\left(\sup_{|s-t|\le\delta} |X_n(s) - X_n(t)|^r\right) = O\left(\delta^{r\frac{\alpha-1}{\beta}}\right) \tag{4.24}$$

for every fixed $r < \beta$.

Note that if (4.19) holds for all $\alpha > 1$, then $\beta = \beta(\alpha)$ is unbounded as a function in α. Thus, the restriction $r < \beta$ is not that serious. Actually we necessarily have $\beta \ge \alpha$, except in the trivial case $X_n(t) = X_n(0)$ for all t and n.

Proof. We already mentioned that (4.20) follows from (4.19) by using the methods of Billingsley [18, pp. 95]. Finally, (4.24) is an easy consequence of (4.20).

Using these preliminaries it is easy to complete the proof of Theorem 4.15.

Proof. By combining (4.21) and (4.19) it follows that for every $s_0 \in [0,1]$

$$\sup_{n\ge 0} \mathbb{E}\left(|X_n(s_0)|^r\right) < \infty.$$

Next, by a direct combination of (4.21) with $s_0 = \frac{1}{2}$ and (4.24) with $\delta = \frac{1}{2}$ we obtain (for all $r \ge 0$)

$$\sup_{n\ge 0} \mathbb{E}\left(\max_{0\le t\le 1} |X_n(t)|^r\right) < \infty$$

and consequently (for any $\epsilon > 0$)

$$\sup_{n\ge 0} \mathbb{E}\left(F(X_n)^{1+\epsilon}\right) < \infty.$$

By weak convergence we also have $F(X_n) \xrightarrow{d} F(X)$. Thus, by [17, p. 348] the expected value $\mathbb{E}\,F(X)$ exists and

$$\lim_{n\to\infty} \mathbb{E}\,F(X_n) = \mathbb{E}\,F(X).$$

Again this theorem does not apply to stochastic processes X_n in $\mathrm{C}[0,\infty)$. However, it is easy to adapt the above theorem to cover this case, too.

For example, suppose that for any choice of fixed positive integers r and α there exist positive constants c_0, c_1, c_2, c_3 such that

$$\sup_{n\ge 0} \mathbb{E}\left(|X_n(t)|^r\right) \le f_1(t) \text{ for all } t \ge 0, \tag{4.25}$$

and

$$\sup_{n\ge 0} \mathbb{E}\left(|X_n(t+s) - X_n(t)|^{2\alpha}\right) \le f_2(t)s^\alpha \text{ for all } s,t \ge 0, \tag{4.26}$$

where $f_1(t)$ and $f_2(t)$ are bounded functions with $f_1(t) \to 0$ and $f_2(t) \to 0$ as $t \to \infty$. Then for all continuous functions $F : \mathrm{C}[0,\infty) \to \mathbb{R}$ of polynomial growth (in $\mathrm{C}_0([0,\infty))$) we have

$$\lim_{n\to\infty} \mathbb{E}\,F(X_n) = \mathbb{E}\,F(X),$$

(compare with [67]).

For example, Theorem 4.15 (and its adaption to $\mathrm{C}[0,\infty)$) can be applied to the height and width of Galton-Watson trees (compare with [67, 69]).

We finally give another useful application of tightness that could lead to a shortcut in the proof of Theorem 4.10 provided that Theorem 4.7 is already known. The basis is the following observation due to Bousquet-Mélou and Janson [24].

Theorem 4.17. *Let S_1 and S_2 be two Polish spaces, and let $\varphi : S_1 \to S_2$ be an injective continuous map. If (W_n) is a tight sequence of random elements of S_1 such that $\varphi(W_n) \xrightarrow{d} Z$ in S_2 for some random $Z \in S_2$ then $W_n \xrightarrow{d} W$ in S_1 for some W with $\varphi(W) \overset{\mathrm{d}}{=} Z$.*

Proof. By Prohorov's theorem, each subsequence of (W_n) has a subsequence that converges in distribution to some limit. Let W' and W'' be limits in distribution of two such subsequences $W_{n_i'}$ and $W_{n_i''}$. Since φ is continuous, $\varphi(W_{n_i'}) \xrightarrow{d} \varphi(W')$ and $\varphi(W_{n_i''}) \xrightarrow{d} \varphi(W'')$. Hence, $\varphi(W') \overset{\mathrm{d}}{=} Z \overset{\mathrm{d}}{=} \varphi(W'')$.

Let A be a (Borel) measurable subset of S_1. By the Souslin–Lusin theorem [45, Theorem III.21, see also III.16–17], $\varphi(A) \subseteq S_2$ is measurable. Thus, using the injectivity of φ,

$$\mathbb{P}\{W' \in A\} = \mathbb{P}\{\varphi(W') \in \varphi(A)\} = \mathbb{P}\{\varphi(W'') \in \varphi(A)\} = \mathbb{P}\{W'' \in A\}.$$

Consequently, $W' \overset{\mathrm{d}}{=} W''$.

In other words, there is a unique distribution of the subsequence limits. Thus, if W is one such limit, then every subsequence of (W_n) has a subsequence that converges in distribution to W; this is equivalent to $W_n \xrightarrow{d} W$ (recall that by (4.2), weak convergence can be reduced to convergence of sequences of real numbers).

This theorem applies in the following situation. Let $S_1 = \{f \in \mathrm{C}_0(\mathbb{R}) : f \ge 0\}$, with the uniform topology inherited from $\mathrm{C}_0(\mathbb{R})$; let S_2 be the space of locally finite measures on $\mathbb{R}$ with the vague topology (see Kallenberg [119, Appendix A2]) and let φ map a function f to the corresponding measure $f\,dx$, that is, $\varphi(f)$ is the measure with density f. Then S_1 is a closed subset of the separable Banach space $\mathrm{C}_0(\mathbb{R})$, and is thus Polish, and so is S_2 by [119, Theorem A2.3]. Further, φ is continuous and injective.

In particular, let

$$l_n(t) = \frac{1}{\sqrt{n}} L_n(t\sqrt{n}) \in S_1$$

be the normalised profile of Galton-Watson trees. Then $\varphi(l_n)$ is precisely the measure

$$\frac{1}{n}\sum_{j\ge 0} L_n(j)\nu_{jn^{-1/2},n^{-1/2}},$$

where $\nu_{y,h}$ denotes the measure with (triangular) density function

$$f(x) = \frac{1}{h}\max\left\{1 - \frac{|x-y|}{h}, 0\right\}$$

that is centred at y and has support $[y-h, y+h]$. Note that $\nu_{y,h}$ approximates the delta distribution δ_y, if $h \to \infty$. It also follows easily that the difference of $\varphi(l_n)$ and the normalised occupation measure

$$\mu_n = \frac{1}{n} \sum_{j \ge 0} L_n(j) \delta_{jn^{-1/2}}$$

converge to 0 in probability.

It follows from Theorem 4.7 that μ_n converges weakly to the occupation measure μ of $\frac{2}{\sigma} e(t)$. Hence, the same is true for $\varphi(l_n)$. Thus, in order to prove Theorem 4.10, that is, $l_n(t) \xrightarrow{d} \frac{\sigma}{2} l\left(\frac{\sigma}{2} t\right)$, it is sufficient to prove Theorem 4.7 and tightness of $l_n(t)$.

4.2.3 Combinatorics on the Profile of Galton-Watson Trees

In Sections 4.2.3–4.2.6 we present a proof of Theorem 4.10 that deals with the limiting behaviour of the profile of Galton-Watson trees. We start with some combinatorics.

Recall that the generating function $y(x) = \sum_{n \ge 1} y_n x^n$ for the weighted numbers $y_n = \sum_{|T|=n} \omega(T)$ of Galton-Watson trees of size n satisfies the functional equation

$$y(x) = x\Phi(y(x)),$$

where $\Phi(x) = \mathbb{E}\, x^{\xi} = \phi_0 + \phi_1 x + \phi_2 x^2 + \cdots$ is the probability generating function of the offspring distribution ξ (or equivalently the generating function of the weights ϕ_j).

In order to compute the distribution of the number of nodes in some given levels in a tree of size n we have to calculate the weighted number

$$y_{k_1,m_1,k_2,m_2,\cdots k_d,m_d;n} = \sum_{|T|=n,\, L_T(k_1)=m_1,\ldots,\, L_T(k_d)=m_d} \omega(T)$$

of trees of size n with m_i nodes in level k_i, $i = 1, \ldots, d$. Then the finite dimensional distribution of the profile of Galton-Watson trees of size n is given by

$$\mathbb{P}\{L_n(k_1) = m_1, \ldots, L_n(k_d) = m_d\} = \frac{y_{k_1,m_1,k_2,m_2,\cdots k_d,m_d;n}}{y_n}.$$

For this purpose we introduce the generating functions

$$\begin{aligned} y_{k_1,\ldots,k_d}(x, u_1, \ldots, u_d) &= \sum_{m_1,\ldots,m_d,n} y_{k_1,m_1,k_2,m_2,\cdots k_d,m_d;n} u_1^{m_1} \cdots u_d^{m_d} x^n \\ &= \sum_{n \ge 1} y_n\, \mathbb{E}\left(u_1^{L_n k_1} \cdots u_d^{L_n k_d}\right) x^n. \end{aligned}$$

These functions can be recursively calculated.

Lemma 4.18. *For $d = 1$ we have*

$$y_0(x, u) = u y(x) \tag{4.27}$$

$$y_{k+1}(x, u) = x\Phi\left(y_k(x, u)\right) \quad k \ge 0 \tag{4.28}$$

and for $d > 1$ with integers $0 \le k_1 < k_2 < k_3 < \cdots < k_d$

$$y_{0,k_2,\ldots,k_d}(x, u_1, \ldots, u_d) = u_1 y_{k_2,\ldots,k_d}(x, u_2, \ldots, u_d) \tag{4.29}$$

$$y_{k_1+1,k_2+1,\ldots,k_d+1}(x, u_1, \ldots, u_d) = x\Phi\left(y_{k_1,k_2,\ldots,k_d}(x, u_1, u_2, \ldots, u_d)\right). \tag{4.30}$$

Proof. The initial equation (4.27) is obvious, since there is always one node at level 0. Next, if we want to count the nodes at level $k+1$ this is equivalent to counting the nodes at level k of all subtrees of the root. This gives (4.28). Similarly we obtain (4.29) and (4.30).

Remark 4.19 *In order to make the structure of the generating functions $y_{k_1,\ldots,k_d}(x, u_1, \ldots, u_d)$ more transparent we introduce the functions $Y_k(x, u)$ by the recurrence*

$$Y_0(x, u) = u,$$
$$Y_{k+1}(x, u) = x\Phi(Y_k(x, u)), \quad k \ge 0. \tag{4.31}$$

For example, the function $Y_k(x, 0)$ is the generating function of Galton-Watson trees of height $< k$. However, their main property is that

$$\begin{aligned} &y_{k_1,\ldots,k_d}(x, u_1, \ldots, u_d) \\ &= Y_{k_1}\Big(x, u_1 Y_{k_2-k_1}\big(x, u_2 Y_{k_3-k_2}(x, \ldots, u_{d-1} Y_{k_d-k_{d-1}}(x, u y(x)) \cdots)\big)\Big). \end{aligned} \tag{4.32}$$

Hence, all generating functions of interest can be expressed in terms of $Y_k(x, u)$ and $y(x)$.

4.2.4 Asymptotic Analysis of the Main Recurrence

In order to get an impression of our problem we first consider a special case, namely planted plane trees that are determined by the generating series $\Phi(x) = 1/(1-x)$. The corresponding generating function that satisfies $y(x) = x\Phi(y(x))$ is explicitly given by

$$y(x) = \frac{1 - \sqrt{1-4x}}{2}.$$

In this special case the recurrence (4.31) can be explicitly solved. Here we have

$$Y_k(x, u) = y(x) + \frac{(u - y(x))\alpha(x)^k}{1 - \frac{u-y(x)}{\sqrt{1-4x}} + \frac{u-y(x)}{\sqrt{1-4x}}\alpha(x)^k},$$

where $\alpha(x)$ abbreviates

$$\alpha(x) = \frac{x}{(1-y(x))^2} = \frac{1 - 2x - \sqrt{1-4x}}{2x}.$$

In fact, this kind of property is generally true, at least from an asymptotic point of view (compare with Lemma 4.20). However, here and in what follows we assume that there exists $\eta > 0$ such that $\mathbb{E}\, e^{\eta\xi}$ is finite, where ξ denotes the offspring distribution of the underlying Galton-Watson branching process. Equivalently there exists $\tau > 0$ that is strictly smaller than the radius of convergence of $\Phi(x)$ with $\tau\Phi'(\tau) = \Phi(\tau)$.[4]

We recall that – under this technical assumption – the generating function $y(x)$ has a local expansion of the form

$$y(x) = \tau - \frac{\tau\sqrt{2}}{\sigma}\sqrt{1 - \frac{x}{x_0}} + O\left(\left|1 - \frac{x}{x_0}\right|\right) \tag{4.33}$$

around its singularity $x_0 = 1/\Phi'(\tau)$. The assumption $d = \gcd\{j > 0 : \phi_j > 0\} = 1$ ensures that $|x\phi'(y(x))| < 1$ for $|x| = x_0$, $x \ne x_0$ (compare with the proof of Theorems 2.19 and 3.6). Hence, by the implicit function theorem $y(x)$ has an analytic continuation to the region $|x| < x_0 + \delta$, $\arg(x - x_0) \ne 0$ for some $\delta > 0$. It also follows that $\alpha = \alpha(x) = x\Phi'(y(x))$ has similar analytic properties, especially it has the local expansion

$$\alpha(x) = 1 - \sigma\sqrt{2}\sqrt{1 - \frac{x}{x_0}} + O\left(\left|1 - \frac{x}{x_0}\right|\right). \tag{4.34}$$

In what follows we also use the abbreviation $\beta = \beta(x) = x\Phi''(y(x))/2$.

Lemma 4.20. *There exists a constant $c > 0$ such that we uniformly have*

$$\begin{aligned} Y_k(x,u) &= y(x) \\ &+ \frac{(u - y(x))\alpha^k}{1 - (u - y(x))\frac{\beta}{\alpha(1-\alpha)} + (u - y(x))\frac{\beta}{\alpha(1-\alpha)}\alpha^k + O\left(|u - y(x)|^2 \frac{1-|\alpha|^{2k}}{1-|\alpha|}\right)} \end{aligned} \tag{4.35}$$

as long as

$$k|u - y(x)| \le c, \quad |x - x_0| \le c, \quad \textit{and} \quad |\arg(x - x_0)| \ge c.$$

Proof. Set $w_k = w_k(x,u) = Y_k(x,u) - y(x)$. In a first step we show that there exists a constant $C > 0$ such that

$$\frac{|w_0|\,|\alpha|^k}{1 + Ck|w_0|} \le |w_k| \le \frac{|w_0|\,|\alpha|^k}{1 - Ck|w_0|} \tag{4.36}$$

as long as $k|w_0| \le 1/(2C)$.

---

[4] This technical assumption makes the analysis easier but it is not necessary. It would be sufficient to assume that the second moment $\mathbb{E}\,\xi^2$ is finite or equivalently that $\Phi''(\tau)$ is finite. We would have $y(x) = \tau - \frac{\tau\sqrt{2}}{\sigma}\sqrt{1 - x/x_0}\,(1 + o(1))$ and $\alpha(x) = 1 - \sigma\sqrt{2}\sqrt{1 - x/x_0}\,(1 + o(1))$ instead of (4.33) and (4.34) and would not get error terms (compare again with Remark 3.7).

We proceed by induction. Note that we always assume that w_k is sufficiently small, say $|w_k| \le c_0$ so that the estimates in the following local expansions are uniform and that we are within the region of convergence of $\Phi(x)$.

By using the local expansions we obtain

$$\begin{aligned} Y_{k+1}(x,u) &= x\Phi(Y_k(x,u)) = x\Phi(y(x)+w_k) \\ &= y(x) + x\Phi'(y(x))w_k + x\Phi''(y(x)+\vartheta_k)w_k^2/2 \\ &= y(x) + \alpha w_k + O(|w_k|^2). \end{aligned}$$

Hence, there exists a constant C such that

$$|\alpha| \frac{|w_k|}{1+C|w_k|} \le |w_{k+1}| \le |\alpha| \frac{|w_k|}{1-C|w_k|}$$

and consequently

$$\frac{1}{|\alpha|\,|w_k|} - \frac{C}{|\alpha|} \le \frac{1}{|w_{k+1}|} \le \frac{1}{|\alpha|\,|w_k|} + \frac{C}{|\alpha|}.$$

Inductively this leads to

$$\frac{1}{|\alpha|^k|w_0|} - \frac{C}{1-|\alpha|}\left(\frac{1}{|\alpha|^k} - 1\right) \le \frac{1}{|w_k|} \le \frac{1}{|\alpha|^k|w_0|} + \frac{C}{1-|\alpha|}\left(\frac{1}{|\alpha|^k} - 1\right)$$

which is equivalent to

$$\frac{|w_0|\,|\alpha|^k}{1+C|w_0|\frac{1-|\alpha|^k}{1-|\alpha|}} \le |w_k| \le \frac{|w_0|\,|\alpha|^k}{1-C|w_0|\frac{1-|\alpha|^k}{1-|\alpha|}}.$$

Observe that $|x-x_0| \le c$ and $|\arg(x-x_0)| \ge c$ imply that $|\alpha| \le 1$. Hence $(1-|\alpha|^k)/(1-|\alpha|) \le k$ and we get (4.36).

Note that the condition $k|w_0| \le 1/(2C)$ ensures that all estimates remain non-negative and also that the a priori assumption $|w_k| \le c_0$ can be verified step by step, where we have to choose c_0 appropriately.

The asymptotic relation (4.35) follows now from a slightly more precise derivation. Similarly to the above we have

$$w_{k+1} = \alpha w_k + \beta w_k^2 + O\left(|w_k|^3\right).$$

This can be rewritten to

$$\begin{aligned} \frac{1}{w_{k+1}} &= \frac{1}{\alpha w_k}\frac{1}{1+\beta w_k/\alpha + O\left(|w_k|^2\right)} \\ &= \frac{1}{\alpha w_k} - \frac{\beta}{\alpha^2} + O\left(\frac{|w_k|}{|\alpha|}\right). \end{aligned}$$

If we set $q_k = \alpha^k / w_k$ then

$$q_{k+1} = q_k - \beta\alpha^{k-1} + O\left(|w_k||\alpha|^k\right)$$

which, with help of the a priori estimate, provides $w_k = O(w_0\alpha^k)$

$$q_k = \frac{1}{w_0} - \frac{\beta}{\alpha}\frac{1-\alpha^k}{1-\alpha} + O\left(|w_0|\left|\frac{1-\alpha^{2k}}{1-\alpha^2}\right|\right)$$

and consequently (4.35).

4.2.5 Finite Dimensional Limiting Distributions

We will use the results of the previous section to prove finite dimensional convergence. For simplicity we only discuss the cases $d = 1$ and $d = 2$ in detail. It will then be clear how to deal with the general case.

Let us start with $d = 1$. Recall that the characteristic function of the local time l at level κ is given by

$$\mathbb{E}\, e^{it\, l(\kappa)} = 1 + \frac{\sqrt{2}}{\sqrt{\pi}} \int_{-1-i\infty}^{-1+i\infty} \frac{t\sqrt{-2s}\exp(-\kappa\sqrt{-2s}\,)}{\sqrt{-2s}\exp(\kappa\sqrt{-2s}\,) - 2it\sinh\left(\kappa\sqrt{-2s}\,\right)} e^{-s}\, ds. \tag{4.37}$$

Proposition 4.21. *Let $\kappa \ge 0$ be given. Then we have for $|t| \le c/(\tau\kappa)$*

$$\lim_{n\to\infty} \mathbb{E} e^{it\frac{1}{\sqrt{n}}L_n(\kappa\sqrt{n})} = \mathbb{E}\, e^{it\frac{\sigma}{2}\, l(\frac{\sigma}{2}\kappa)}. \tag{4.38}$$

Consequently

$$\frac{1}{\sqrt{n}} L_n(\kappa\sqrt{n}) \xrightarrow{d} \frac{\sigma}{2} l\left(\frac{\sigma}{2}\kappa\right).$$

Proof. Recall that the generating function $y_k(x,u) = Y_k(x, uy(x))$ can be interpreted as

$$y_k(x,u) = \sum_{n\ge1} y_n\, \mathbb{E}\, u^{L_n(k)}\, x^n$$

and is, thus, related to the characteristic function of $L_n(k)$. In particular, we have

$$\mathbb{E} e^{itL_n(k)/\sqrt{n}} = \frac{1}{y_n}[x^n] y_k\left(x, e^{it/\sqrt{n}}\right).$$

In order to get asymptotics for this characteristic function we will use the local representation of $Y_k(x,u)$ from Lemma 4.20 and a properly chosen contour integration applied to Cauchy's formula.

Let $\Gamma = \gamma' \cup \Gamma'$ consist of a line

$$\gamma' = \left\{x = x_0\left(1 + \frac{s}{n}\right) : \Re(s) = -1,\, |\Im(s)| \le C\log^2 n\right\}$$

with an arbitrarily chosen fixed constant $C > 0$, and Γ' be a circular arc centred at the origin and making Γ a closed curve. Then

$$\mathbb{E}e^{itL_n(k)/\sqrt{n}} = \frac{1}{y_n}\frac{1}{2\pi i}\int_\Gamma y_k\left(x, e^{it/\sqrt{n}}\right)\frac{dx}{x^{n+1}}. \tag{4.39}$$

The contribution of Γ' is exponentially small, since for $x \in \Gamma'$ we have $|x^{-n-1}| \sim x_0^{-n}e^{-C\log^2 n}$, whereas $\left|y_k\left(x, e^{it/\sqrt{n}}\right)\right|$ is bounded.

If $x \in \gamma'$ then the local expansion (4.35) is valid for a proper range for k. In particular, we replace k by $\lfloor\kappa\sqrt{n}\rfloor$, u by $e^{it/\sqrt{n}}y(x)$ and x by $x_0\left(1+\frac{s}{n}\right)$. If we assume that $\kappa > 0$ is given then the assumptions of Lemma 4.20 are satisfied provided that $|t| \le c/(\kappa\tau)$, where c is the constant from Lemma 4.20:

$$k|(u-1)y(x)| \le \kappa\sqrt{n}\left|e^{it/\sqrt{n}}-1\right||y(x)| \le \kappa|t|\tau \le c.$$

Next we use the local expansions

$$\begin{aligned} y(x) &= \tau - \frac{\tau\sqrt{2}}{\sigma}\sqrt{-\frac{s}{n}} + O(|s|/n), \\ \alpha &= 1 - \sigma\sqrt{2}\sqrt{-\frac{s}{n}} + O(|s|/n), \\ \beta &= \frac{\sigma^2}{2\tau} + O\left(\sqrt{-\frac{s}{n}}\right) \end{aligned}$$

and

$$\alpha^k = \exp\left(-\sigma\kappa\sqrt{-2s}\right)\left(1+O(\kappa s/\sqrt{n})\right)$$

to get asymptotically for $x \in \gamma'$

$$\begin{aligned} &\frac{(u-1)y(x)\alpha^k}{1-(u-1)y(x)\frac{\beta}{\alpha(1-\alpha)} + (u-1)y(x)\frac{\beta}{\alpha(1-\alpha)}\alpha^k + O\left(|u-1|^2\frac{1-|\alpha|^{2k}}{1-|\alpha|}\right)} \\ &= \frac{1}{\sqrt{n}}\frac{it\tau e^{-\sigma\kappa\sqrt{-2s}}(1+O(s/\sqrt{n}))}{1-\frac{it\sigma}{2\sqrt{-2s}}+\frac{it\sigma}{2\sqrt{-2s}}\exp\left(-\sigma\kappa\sqrt{-2s}\right)+O(|s|^{3/2}/\sqrt{n})} \\ &= \frac{1}{\sqrt{n}}\frac{it\sqrt{-2s}\,\tau\exp\left(-\sigma\kappa\sqrt{-2s}\right)(1+O(s/\sqrt{n}))}{\sqrt{-2s}-\frac{it\sigma}{2}+\frac{it\sigma}{2}\exp\left(-\sigma\kappa\sqrt{-2s}\right)+O(\sqrt{|s|/n})} \\ &= \frac{1}{\sqrt{n}}\frac{it\sqrt{-2s}\,\tau\exp\left(-\frac{1}{2}\sigma\kappa\sqrt{-2s}\right)(1+O(s/\sqrt{n}))}{\sqrt{-2s}\exp\left(\frac{1}{2}\sigma\kappa\sqrt{-2s}\right)(1+O(1/\sqrt{n})) - it\sigma\sinh\left(\frac{1}{2}\sigma\kappa\sqrt{-2s}\right)}. \end{aligned}$$

Finally, by applying the substitution $x = x_0(1+s/n)$ and the approximation $x^{-n} = x_0^{-n}e^{-s}(1+O(|s|^2/n))$ we notice that the integral

$$\frac{1}{2\pi i}\int_{\gamma'}\frac{(u-1)y(x)\alpha^k}{1-(u-1)y(x)\frac{\beta}{\alpha(1-\alpha)}(1-\alpha^k)+O\left(|u-1|^2\frac{1-|\alpha|^{2k}}{1-|\alpha|}\right)}\frac{dx}{x^{n+1}}$$

is approximated by

$$\frac{\tau x_0^{-n}}{\sigma\sqrt{2\pi}\, n^{3/2}} \frac{\sqrt{2}}{\sqrt{\pi}} \int\limits_{-1-\log^2 n}^{1+\log^2 n} \frac{\frac{it\sigma}{2}\sqrt{-2s}\exp(-\frac{1}{2}\kappa\sigma\sqrt{-2s}\,)}{\sqrt{-2s}\exp(\frac{1}{2}\kappa\sigma\sqrt{-2s}\,) - 2it\sinh\left(\frac{1}{2}\kappa\sigma\sqrt{-2s}\right)}\, e^{-s}\, ds.$$

Since the integral is absolutely convergent, we can safely replace the integral $\int_{-1-\log^2 n}^{-1+\log^2 n}$ by $\int_{-1-\infty}^{-1+\infty}$. We recall that $y_n \sim (\tau/(\sigma\sqrt{2\pi}))x_0^{-n}n^{-3/2}$. Hence,

$$\lim_{n\to\infty} \mathbb{E}\, e^{it\frac{1}{\sqrt{n}}L_n(\lfloor\kappa\sqrt{n}\rfloor)} = \mathbb{E}\, e^{it\frac{\sigma}{2}\, l(\frac{\sigma}{2}\kappa)}$$

follows. However, in view of (4.41) (that we will prove independently of the present proposition)

$$\mathbb{E}\, e^{it\frac{1}{\sqrt{n}}L_n(\lfloor\kappa\sqrt{n}\rfloor)} \quad \text{and} \quad \mathbb{E}\, e^{it\frac{1}{\sqrt{n}}L_n(\kappa\sqrt{n})}$$

have the same limit. This completes the proof of (4.38) for a small open interval for t (that contains $t = 0$).

Since the characteristic function $\mathbb{E}\, e^{it\frac{\sigma}{2}\, l(\frac{\sigma}{2}\kappa)}$ represents an analytic function in t, we thus get weak convergence of $\frac{1}{\sqrt{n}}L_n(\kappa\sqrt{n}) \xrightarrow{d} \frac{\sigma}{2}l\left(\frac{\sigma}{2}\kappa\right)$ (compare with [139, pp. 224–225]).

With the help of the same techniques it is also possible to obtain asymptotics for moments (compare with [199, 93]). We just demonstrate this for the expected profile $\mathbb{E}\, L_n(k)$.

By definition the partial derivative

$$\gamma_k(x) = \left[\frac{\partial y_k(x,u)}{\partial u}\right]_{u=1} = \sum_{n\ge 1} y_n \mathbb{E}\, L_n(k)\; x^n$$

encodes the first moment. By induction it follows that

$$\gamma_k(x) = y(x)\alpha(x)^k.$$

Thus if $k = \lfloor\kappa\sqrt{n}\rfloor$ (with some $\kappa > 0$) we obtain (by using exactly the same kind of contour integration as above)

$$y_n \mathbb{E}\, L_n(k) \sim \frac{\tau x_0^{-n}}{n} \frac{1}{2\pi i} \int_{-1-i\infty}^{-1+i\infty} \exp\left(-\sigma\kappa\sqrt{-2s}\right) e^{-s}\, ds.$$

In view of Lemma 4.22 we immediately derive

$$\mathbb{E}\, L_n(\lfloor\kappa\sqrt{n}\rfloor) \sim \sigma^2\kappa e^{-\frac{1}{2}\sigma^2\kappa^2}\sqrt{n}.$$

Lemma 4.22. *Let $r > 0$. Then the following formula holds:*

$$\frac{1}{2\pi i} \int_{-1-i\infty}^{-1+i\infty} \exp\left(-r\sqrt{-2s}\right) e^{-s}\, ds = \frac{r}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2}.$$

Proof. We first show that the integral over the line from $-1-i\infty$ to $-1+i\infty$ can be replaced by the contour that surrounds the positive real axis as a so-called Hankel contour H (see Figure 2.5). For this purpose we have to show that the integrals

$$\int_{|s|=R,\, 0<|\arg(s)|\le\pi/2} \exp\left(-r\sqrt{-2s}\right) e^{-s}\, ds$$

are negligible (as $R\to\infty$) which is actually true since

$$\int_0^{\pi/2} \left| e^{-r\sqrt{-2Re^{i\varphi}}-Re^{i\varphi}} \right| R\, d\varphi \le \frac{R\pi}{2} \max\left\{ e^{-r\sqrt{R}}, e^{-R/\sqrt{2}} \right\}.$$

After that we use the substitution $s=\frac{1}{2}t^2$, where t goes from $i-\infty$ to $i+\infty$, and the resulting integral can be finally calculated. Summing up we have

$$\begin{aligned}
\frac{1}{2\pi i}\int_{-1-i\infty}^{-1+i\infty} \exp\left(-r\sqrt{-2s}\right) e^{-s}\, ds &= \frac{1}{2\pi i}\int_H \exp\left(-r\sqrt{-2s}\right) e^{-s}\, ds \\
&= \frac{1}{2\pi i}\int_{i-\infty}^{i+\infty} e^{irt-\frac{1}{2}t^2}\, t\, dt \\
&= \frac{e^{-\frac{1}{2}r^2}}{2\pi i}\int_{i-\infty}^{i+\infty} e^{-\frac{1}{2}(t-ir)^2}(t-ir+ir)\, dt \\
&= \frac{re^{-\frac{1}{2}r^2}}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}w^2}\, dw \\
&= \frac{r}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2}.
\end{aligned}$$

Next consider the two dimensional case. The characteristic function of $(l(\kappa_1), l(\kappa_2))$ with $0<\kappa_1<\kappa_2$ is given by

$$\mathbb{E}\, e^{it_1\, l(\kappa_1)+it_2\, l(\kappa_2)} = 1 + \frac{\sqrt{2}}{i\sqrt{\pi}} \int_{-1-i\infty}^{-1+i\infty} \Psi_{\kappa_1}\left(s, it_1 + \Psi_{\kappa_2-\kappa_1}(s, it_2)\right) e^{-s}\, ds,$$

with

$$\Psi_{\kappa_1}\left(s, it_1+\Psi_{\kappa_2-\kappa_1}(s,it_2)\right) = \frac{\sqrt{-2s}\, e^{-\kappa_1\sqrt{-2s}} \left(it_1 + \frac{it_2\sqrt{-2s}\, e^{-(\kappa_2-\kappa_1)\sqrt{-2s}}}{\sqrt{-2s}e^{(\kappa_2-\kappa_1)\sqrt{-2s}} - 2it_2\sinh\left((\kappa_2-\kappa_1)\sqrt{-2s}\right)} \right)}{\sqrt{-2s}\, e^{\kappa_1\sqrt{-2s}} - \left(it_1 + \frac{it_2\sqrt{-2s}e^{-(\kappa_2-\kappa_1)\sqrt{-2s}}}{\sqrt{-2s}\, e^{(\kappa_2-\kappa_1)\sqrt{-2s}} - 2it_2\sinh\left((\kappa_2-\kappa_1)\sqrt{-2s}\right)} \right) \sinh\left(\kappa_1\sqrt{-2s}\right)}.$$

Proposition 4.23. *Let $0\le\kappa_1<\kappa_2$ be given. Then we have for real t_1 and t_2 with sufficiently small absolute value*

$$\lim_{n\to\infty} \mathbb{E} e^{it_1 \frac{1}{\sqrt{n}} L_n(\kappa_1\sqrt{n}) + it_2 \frac{1}{\sqrt{n}} L_n(\kappa_2\sqrt{n})} = \mathbb{E}\, e^{it_1 \frac{\sigma}{2}\, l(\frac{\sigma}{2}\kappa_1) + it_2 \frac{\sigma}{2}\, l(\frac{\sigma}{2}\kappa_2)}. \qquad (4.40)$$

Consequently

$$\left(\frac{1}{\sqrt{n}} L_n(\kappa_1\sqrt{n}), \frac{1}{\sqrt{n}} L_n(\kappa_2\sqrt{n})\right) \xrightarrow{d} \left(\frac{\sigma}{2} l\left(\frac{\sigma}{2}\kappa_1\right), \frac{\sigma}{2} l\left(\frac{\sigma}{2}\kappa_2\right)\right).$$

Proof. By definition we have for $k_1 < k_2$

$$\begin{aligned}
&\mathbb{E}\, e^{it_1 L_n(k_1)/\sqrt{n} + it_2 L_n(k_2)/\sqrt{n}} \\
&\quad = \frac{1}{y_n} [x^n] Y_{k_1}\left(x, e^{it_1/\sqrt{n}} Y_{k_2-k_1}\left(x, e^{it_2/\sqrt{n}} y(x)\right)\right) \\
&\quad = \frac{1}{y_n} \frac{1}{2\pi i} \int_\Gamma Y_{k_1}\left(x, e^{it_1/\sqrt{n}} Y_{k_2-k_1}\left(x, e^{it_2/\sqrt{n}} y(x)\right)\right) \frac{dx}{x^{n+1}}
\end{aligned}$$

with the same path of integration $\Gamma = \Gamma' \cup \gamma'$ as in the proof of Proposition 4.21.

If $x \in \gamma'$ then substitute $x = x_0\left(1 + \frac{s}{n}\right)$ and set $\alpha = x\phi'(y(x))$. We also use the abbreviations $u_1 = e^{it_1/\sqrt{n}}$, $u_2 = e^{it_2/\sqrt{n}}$, and $h = k_2 - k_1$ resp. $\eta = \kappa_2 - \kappa_1$ where $k_1 = \lfloor \kappa_1\sqrt{n} \rfloor$ and $k_2 = \lfloor \kappa_1\sqrt{n} \rfloor$. Now, applying Lemma 4.20 yields

$$Y_h(x, u_2 y(x)) = y(x) + w_h(u_2, x)$$

where

$$\begin{aligned}
w_h(u_2, x) &\sim \frac{(u_2 - 1)y(x)\alpha^h}{1 - (u_2-1)y(x)\frac{\beta}{\alpha(1-\alpha)} + (u_2-1)y(x)\frac{\beta}{\alpha(1-\alpha)}\alpha^h} \\
&\sim \frac{1}{\sqrt{n}} \frac{it_2\sqrt{-2s}\,\tau \exp\left(-\frac{1}{2}\sigma\eta\sqrt{-2s}\right)}{\sqrt{-2s}\,\exp\left(\frac{1}{2}\sigma\eta\sqrt{-2s}\right) - it_2\sigma \sinh\left(\frac{1}{2}\sigma\eta\sqrt{-2s}\right)} \\
&= \frac{1}{\sqrt{n}} \frac{2\tau}{\sigma} \Psi_{\eta\sigma/2}\left(s, \frac{it_2\sigma}{2}\right)
\end{aligned}$$

and

$$Y_{k_1}(x, u_1 Y_h(x, u_2 y(x))) = y(x) + \tilde{w}_{k_1}(u_1, u_2, x)$$

where $(w_h = w_h(u_2, x))$

$$\begin{aligned}
&\tilde{w}_{k_1}(u_1, u_2, x) \\
&\sim \frac{((u_1-1)y(x) + u_1 w_h)\alpha^{k_1}}{1 - ((u_1-1)y(x) + u_1 w_h)\frac{\beta}{\alpha(1-\alpha)} + ((u_1-1)y(x) + u_1 w_h)\frac{\beta}{\alpha(1-\alpha)}\alpha^{k_1}} \\
&\sim \frac{1}{\sqrt{n}} \frac{\left(it_1 + \frac{2}{\sigma}\Psi_{\eta\sigma/2}\left(s, \frac{it_2\sigma}{2}\right)\right)\sqrt{-2s}\,\tau \exp\left(-\frac{1}{2}\sigma\kappa_1\sqrt{-2s}\right)}{\sqrt{-2s}\,\exp\left(\frac{1}{2}\sigma\kappa_1\sqrt{-2s}\right) - \left(it_1\sigma + 2\Psi_{\eta\sigma/2}\left(s, \frac{it_2\sigma}{2}\right)\right)\sinh\left(\frac{1}{2}\sigma\kappa_1\sqrt{-2s}\right)} \\
&= \frac{1}{\sqrt{n}} \frac{2\tau}{\sigma} \Psi_{\kappa_1\sigma/2}\left(s, \frac{it_1\sigma}{2} + \Psi_{\eta\sigma/2}\left(s, \frac{it_2\sigma}{2}\right)\right).
\end{aligned}$$

Consequently, the integral

$$\frac{1}{2\pi i}\int_{\gamma'} \tilde{w}_{\kappa_1}(u_1,u_2,x)\,\frac{dx}{x^{n+1}}$$

is approximated by

$$\frac{\tau x_0^{-n}}{\sigma\sqrt{2\pi}\,n^{3/2}}\,\frac{\sqrt{2}}{\sqrt{\pi}}\int_{-1-\log^2 n}^{-1+\log^2 n} \Psi_{\kappa_1\sigma/2}\left(s,\frac{it_1\sigma}{2}+\Psi_{\eta\sigma/2}\left(s,\frac{it_2\sigma}{2}\right)\right)e^{-s}\,ds.$$

Since the integral on Γ' is negligible (as in the proof of Proposition 4.21), we obtain (4.40) and, thus, weak convergence of the corresponding random vectors.

By inspecting the proofs of Propositions 4.21 and 4.23 it is clear that finite dimensional convergence follows by iterating the procedure. Thus, for every d-tuple κ_j with $0\le\kappa_1<\kappa_2<\cdots<\kappa_d$ we get

$$\left(\frac{1}{\sqrt{n}}L_n(\kappa_1\sqrt{n}),\ldots,\frac{1}{\sqrt{n}}L_n(\kappa_d\sqrt{n})\right)\xrightarrow{d}\left(\frac{\sigma}{2}l\left(\frac{\sigma}{2}\kappa_1\right),\ldots,\frac{\sigma}{2}l\left(\frac{\sigma}{2}\kappa_d\right)\right).$$

4.2.6 Tightness

In this section we will show that the sequence of random variables $l_n(t)=n^{-1/2}L_n(t\sqrt{n})$, $t\ge 0$, is tight in $C[0,\infty)$ by checking the second and third assumption of Theorem 4.14. Since a sequence of stochastic processes $(X_n(t),t\ge 0)$ is tight in $C[0,\infty)$, if and only if $(X_n(t),\,0\le t\le T)$ is tight in $C[0,T]$ for all $T>0$ (see [120, p. 63]) we may restrict ourselves to finite intervals. Note that Theorem 4.14 is only formulated for the unit interval $[0,1]$. But this extends to arbitrary intervals $[0,T]$ by scaling.

The second condition of Theorem 4.14 is trivially true, since $L(0)=0$. The third condition will be verified for $\alpha=2>1$ and $\beta=4$. In particular we show the following property.

Theorem 4.24. *There exists a constant $C>0$ such that*

$$\mathbb{E}\,(L_n(k)-L_n(k+h))^4\le C\,h^2 n \tag{4.41}$$

holds for all non-negative integers n,k,h.

Obviously (4.41) proves (4.19) (for $\alpha=2$ and $\beta=4$), if $k=\kappa_1\sqrt{n}$ and $h=\eta\sqrt{n}=(\kappa_2-\kappa_2)\sqrt{n}$ are non-negative integers, since it rewrites to

$$\mathbb{E}\left|\frac{1}{\sqrt{n}}L_n(\kappa_1\sqrt{n})-\frac{1}{\sqrt{n}}L_n(\kappa_2\sqrt{n})\right|^4\le C\,|\kappa_2-\kappa_1|^2. \tag{4.42}$$

However, in the case of linear interpolation it is an easy exercise to extend (4.41) to arbitrary $\kappa_1, \kappa_2 \ge 0$ (compare with [92]). Namely if $k \le s < t \le k+1$ for some non-negative integer k then

$$\begin{aligned}\mathbb{E}\left(L_n(t) - L_n(s)\right)^4 &= (t-s)^4\,\mathbb{E}\left(L_n(r) - L_n(r \pm 1)\right)^4 \\ &\le C\,(t-s)^4 n \\ &\le C\,(t-s)^2 n.\end{aligned}$$

Also with help of this upper bound we get for the remaining case, where $s \le k \le t$ (for some non-negative integer k)

$$\begin{aligned}\mathbb{E}\left(L_n(t) - L_n(s)\right)^4 &\le 3^4\mathbb{E}\left(L_n(t) - L_n(\lfloor t \rfloor)\right)^4 \\ &\quad + 3^4\mathbb{E}\left(L_n(\lfloor t \rfloor) - L_n(\lceil s \rceil)\right)^4 \\ &\quad + 3^4\mathbb{E}\left(L_n(\lceil s \rceil) - L_n(s)\right)^4 \\ &\le 3^4 C\,\left((t - \lfloor t \rfloor)^2 + (\lceil s \rceil - \lfloor t \rfloor)^2 + (s - \lceil s \rceil)^2\right) n \\ &\le 3^4 C\,(t-s)^2 n.\end{aligned}$$

This proves (4.42) for general κ_1 and κ_2.

Remark 4.25 *It should be mentioned that it is not sufficient to consider the second moment* $\mathbb{E}\left(L_n(r) - L_n(r+h)\right)^2$. *The optimal upper bound is given by*

$$\mathbb{E}\left(L_n(r) - L_n(r+h)\right)^2 \le C\,h\sqrt{n}$$

which provides (4.19) just for $\alpha = 1$.

By using the combinatorial approach we get

$$\mathbb{P}\{L_n(k) - L_n(k+h) = m\} = \frac{1}{y_n}[x^n u^m] Y_k(x, uY_h(x, u^{-1}y(x)))$$

and consequently

$$\mathbb{E}\left(L_n(k) - L_n(k+h)\right)^4 = \frac{1}{y_n}[x^n] H_{kh}(x), \tag{4.43}$$

in which

$$H_{kh}(x) = \left[\left(\frac{\partial}{\partial u} + 7\frac{\partial^2}{\partial u^2} + 6\frac{\partial^3}{\partial u^3} + \frac{\partial^4}{\partial u^4}\right) Y_k(x, uY_h(x, u^{-1}y(x)))\right]_{u=1}. \tag{4.44}$$

In order to prove Theorem 4.24 we have to estimate $H_{kh}(x)$ in a proper way. For this purpose we use the following representation.

Proposition 4.26. *Set* $\alpha = x\Phi'(y(x))$ *and*

$$\Delta = \{x : |x| < x_0 + \eta,\ |\arg(x - x_0)| > \theta\}, \tag{4.45}$$

in which $\eta > 0$ is sufficiently small and $0 < \theta < \pi/2$. Then $H_{kh}(x)$ can be represented as

$$H_{kh}(x) = G_{1,kh}(x)\frac{(1-\alpha^h)^2}{(1-\alpha)^3} + G_{2,kh}(x)\frac{1-\alpha^h}{(1-\alpha)^2} + G_{3,kh}(x)\frac{1}{1-\alpha} + G_{4,kh}(x), \tag{4.46}$$

in which $G_{j,kh}(x)$, $1 \le j \le 4$, are functions that are uniformly bounded for $x \in \Delta$ and $k, h \ge 0$.

The proof of Proposition 4.26 requires the following formulas.

Lemma 4.27. *Let $\alpha = x\Phi'(y(x))$, $\beta = x\Phi''(y(x))$, $\gamma = x\Phi'''(y(x))$, and $\delta = x\Phi''''(y(x))$. Then we have*

$$\begin{aligned}
\frac{\partial Y_k}{\partial u}(x,1) &= \alpha^k, \\
\frac{\partial^2 Y_k}{\partial u^2}(x,1) &= \frac{\beta}{\alpha}\alpha^k\frac{1-\alpha^k}{1-\alpha}, \\
\frac{\partial^3 Y_k}{\partial u^3}(x,1) &= \frac{\gamma}{\alpha}\alpha^k\frac{1-\alpha^{2k}}{1-\alpha^2} + 3\frac{\beta^2}{\alpha}\alpha^k\frac{(1-\alpha^k)(1-\alpha^{k-1})}{(1-\alpha)(1-\alpha^2)}, \\
\frac{\partial^4 Y_k}{\partial u^4}(x,1) &= \frac{\delta}{\alpha}\alpha^k\frac{1-\alpha^{3k}}{1-\alpha^3} \\
&\quad + \left(2\beta\gamma(2+5\alpha+5\alpha^k+3\alpha^{k+1}) + 3\beta^3/\alpha\right)\alpha^k\frac{(1-\alpha^k)(1-\alpha^{k-1})}{(1-\alpha^2)(1-\alpha^3)} \\
&\quad + 3\beta^3(1+5\alpha)\alpha^k\frac{(1-\alpha^k)(1-\alpha^{k-1})(1-\alpha^{k-2})}{(1-\alpha)(1-\alpha^2)(1-\alpha^3)}.
\end{aligned}$$

Proof. From $Y_{k+1}(x,u) = x\Phi(Y_k(x,u))$ we obtain the recurrence relations

$$\begin{aligned}
\frac{\partial Y_{k+1}}{\partial u} &= x\Phi'(Y_k)\frac{\partial Y_k}{\partial u}, \\
\frac{\partial^2 Y_{k+1}}{\partial u^2} &= x\Phi''(Y_k)\left(\frac{\partial Y_k}{\partial u}\right)^2 + x\Phi'(Y_k)\frac{\partial^2 Y_k}{\partial u^2}, \\
\frac{\partial^3 Y_{k+1}}{\partial u^3} &= x\Phi'''(Y_k)\left(\frac{\partial Y_k}{\partial u}\right)^3 + 3x\Phi''(Y_k)\frac{\partial Y_k}{\partial u}\frac{\partial^2 Y_k}{\partial u^2} + x\Phi'(Y_k)\frac{\partial^3 Y_k}{\partial u^3}, \\
\frac{\partial^4 Y_{k+1}}{\partial u^4} &= x\Phi''''(Y_k)\left(\frac{\partial Y_k}{\partial u}\right)^4 + 6x\Phi'''(Y_k)\left(\frac{\partial Y_k}{\partial u}\right)^2\frac{\partial^2 Y_k}{\partial u^2} \\
&\quad + 3x\Phi''(Y_k)\left(\frac{\partial^2 Y_k}{\partial u^2}\right)^2 + 4x\Phi''(Y_k)\frac{\partial Y_k}{\partial u}\frac{\partial^3 Y_k}{\partial u^3} + x\Phi'(Y_k)\frac{\partial^4 Y_k}{\partial u^4}.
\end{aligned}$$

Since $Y_k(x,1) = y(x)$ for all $k \ge 0$, this system of recurrence relations has the explicit solutions stated in Lemma 4.27 for $u = 1$.

Proof. (*Proposition 4.26*) First we can use Lemma 4.27 to make (4.44) more explicit. Since

$$\frac{\partial}{\partial u} Y_k(x, uY_h(x, u^{-1}y(x))) = \frac{\partial Y_k}{\partial u}(x, uY_h(x, u^{-1}y(x))) \\ \times \left(Y_h(x, u^{-1}y(x)) - u^{-1}y(x)\frac{\partial Y_h}{\partial u}(x, u^{-1}y(x)) \right),$$

we obtain

$$\left[\frac{\partial}{\partial u} Y_k(x, uY_h(x, u^{-1}y(x))) \right]_{u=1} = y(x)\alpha^k(1-\alpha^h).$$

Similarly

$$\begin{aligned}
&\left[\frac{\partial^2}{\partial u^2} Y_k(x, uY_h(x, u^{-1}y(x))) \right]_{u=1} \\
&\quad = y(x)^2 \frac{\partial^2 Y_k}{\partial u^2}(x,1)(1-\alpha^h)^2 + y(x)^2 \alpha^k \frac{\partial^2 Y_h}{\partial u^2}(x,1), \\
&\left[\frac{\partial^3}{\partial u^3} Y_k(x, uY_h(x, u^{-1}y(x))) \right]_{u=1} \\
&\quad = y(x)^3 \frac{\partial^3 Y_k}{\partial u^3}(x,1)(1-\alpha^h)^3 + 3y(x)^3 \frac{\partial^2 Y_k}{\partial u^2}(x,1)\frac{\partial^2 Y_h}{\partial u^2}(x,1)(1-\alpha^h), \\
&\quad - 3y(x)^2\alpha^k \frac{\partial^2 Y_h}{\partial u^2}(x,1) - y(x)^3\alpha^k \frac{\partial^3 Y_h}{\partial u^3}(x,1),
\end{aligned}$$

and

$$\begin{aligned}
&\left[\frac{\partial^4}{\partial u^4} Y_k(x, uY_h(x, u^{-1}y(x))) \right]_{u=1} \\
&\quad = y(x)^4 \frac{\partial^4 Y_k}{\partial u^4}(x,1)(1-\alpha^h)^4 + 7y(x)^4 \frac{\partial^3 Y_k}{\partial u^3}(x,1)\frac{\partial^2 Y_h}{\partial u^2}(x,1)(1-\alpha^h)^2 \\
&\quad - 12y(x)^4 \frac{\partial^2 Y_k}{\partial u^2}(x,1)\frac{\partial^2 Y_h}{\partial u^2}(x,1)(1-\alpha^h) \\
&\quad + 3y(x)^4 \frac{\partial^2 Y_k}{\partial u^2}(x,1)\left(\frac{\partial^2 Y_h}{\partial u^2}(x,1) \right)^2 \\
&\quad - 4y(x)^4 \frac{\partial^2 Y_k}{\partial u^2}(x,1)\frac{\partial^3 Y_h}{\partial u^3}(x,1)(1-\alpha^h) + 12y(x)^2\alpha^k \frac{\partial^2 Y_h}{\partial u^2}(x,1) \\
&\quad + 8y(x)^3\alpha^k \frac{\partial^3 Y_h}{\partial u^3}(x,1) + y(x)^4\alpha^k \frac{\partial^4 Y_h}{\partial u^4}(x,1),
\end{aligned}$$

yielding an explicit expression of $H_{kh}(x)$ in terms of $y(x)$.

Now notice that

$$\sup_{x\in\Delta} |\alpha| = 1, \tag{4.47}$$

since $\alpha = x\Phi'(y(x))$ has the local expansion (4.34). Hence, a representation of the form (4.46) follows immediately with functions $G_{j,kh}(x)$, $1 \le j \le 4$, which are uniformly bounded for $x \in \Delta$.

The final step of the proof of Theorem 4.24 is to use (4.46) and the Transfer Lemma 2.12.

Proof. (*Theorem 4.24*) Since $y_n \sim (\tau/\sqrt{2\pi\sigma^2})x_0^{-n}n^{-3/2}$, Theorem 4.24 is equivalent to

$$[x^n]H_{kh}(x) = O\left(x_0^{-n}\frac{h^2}{\sqrt{n}}\right) \tag{4.48}$$

uniformly for all $k, h \ge 0$. Note that $H_{k0}(x) \equiv 0$. So we may assume that $h \ge 1$.

First, let us consider the first term of $H_{kh}(x)$ (in the representation (4.46)):

$$\begin{aligned} G_{1,kh}(x)\frac{(1-\alpha^h)^2}{(1-\alpha)^3} &= G_{1,kh}(x)\frac{1}{1-\alpha}\sum_{i=0}^{h-1}\alpha^i\sum_{j=0}^{h-1}\alpha^j \\ &= \sum_{i,j=0}^{h-1} G_{1,kh}(x)\frac{\alpha^{i+j}}{1-\alpha} = O\left(h^2\frac{1}{|1-\alpha|}\right). \end{aligned}$$

Since

$$\frac{1}{1-\alpha} = O\left((1-x/x_0)^{-1/2}\right),$$

we can apply Lemma 2.12 (with $\beta = 1/2$) and obtain

$$G_{1,kh}(x)\frac{(1-\alpha^h)^2}{(1-\alpha)^3} = O\left(x_0^{-n}h^2n^{-1/2}\right).$$

The coefficient of the second term is even smaller:

$$\begin{aligned} [x^n]G_{2,kh}(x)\frac{(1-\alpha^h)}{(1-\alpha)^2} &= [x^n]G_{2,kh}(x)\frac{1}{1-\alpha}\sum_{i=0}^{h-1}\alpha^i \\ &= O\left(x_0^{-n}hn^{-1/2}\right) = O\left(x_0^{-n}h^2n^{-1/2}\right). \end{aligned}$$

Similarly we can handle the remaining terms:

$$[x^n]G_{3,kh}(x)\frac{1}{1-\alpha} = O\left(x_0^{-n}n^{-1/2}\right) = O\left(x_0^{-n}h^2n^{-1/2}\right)$$

and

$$[x^n]G_{4,kh}(x) = O\left(x_0^{-n}n^{-1}\right) = O\left(x_0^{-n}h^2n^{-1/2}\right).$$

Thus we have proved (4.48) which is equivalent to (4.41).

This also completes the proof of Theorem 4.10 in the case that the offspring distribution ξ has finite exponential moments.

4.2.7 The Height of Galton-Watson Trees

We have already mentioned that the limiting distribution of the height H_n of Galton-Watson trees is asymptotically given by

$$\frac{1}{\sqrt{n}} H_n \xrightarrow{d} \frac{2}{\sigma} \max_{0 \le t \le 1} e(t),$$

where $e(t)$ denotes the Brownian excursion of duration 1 (compare with Theorem 4.8).

In what follows we will shortly present a combinatorial proof of this limiting relation that goes back to Flajolet and Odlyko [82] who derived asymptotics for all moments $\mathbb{E}\, H_n^r$ in the binary case.

The generating functions $y_k(x,u)$ that describe the profile of Galton-Watson trees are closely related to the height distribution. Let $y_k(x)$ denote the generating function of trees with height $< k$, that is,

$$y_k(x) = \sum_{n \ge 1} \left(\sum_{|T|=n,\, h(T)<k} \omega(T) \right) x^n.$$

Then we have

$$y_k(x) = y_k(x,0),$$

since $L_T(k) = 0$, if and only if $h(T) < k$. In particular we have

$$y_0(x) = 0,$$
$$y_{k+1}(x) = x\Phi(y_k(x)).$$

In the case of planted plane trees, that is $\Phi(x) = 1/(1-x)$, this recurrence can be explicitly solved and is given by

$$y_k(x) = y(x) - \frac{y(x)\alpha(x)^k}{1 + \frac{y(x)}{\sqrt{1-4x}}(1 - \alpha(x)^k)},$$

where $y(x) = (1 - \sqrt{1-4x})/2$ and $\alpha(x) = x/(1-y(x))^2$. As we did in the case of the profile we will show that such a relation holds asymptotically in the general case. For convenience we set

$$e_k(x) = y(x) - y_k(x),$$

which is the generating function of the (weighted) numbers of trees of size n of height $\ge k$.

We will also make use of a properly chosen Δ-region

$$\Delta = \Delta(x_0, \eta, \delta) = \{x : |x| < x_0 + \eta,\ |\arg(x/x_0 - 1)| > \delta\},$$

where $\eta > 0$ and $0 < \delta < \frac{\pi}{2}$. In particular we will choose η sufficiently small and δ very close to $\frac{\pi}{2}$.

Proposition 4.28. *We have uniformly for $x \in \Delta$ and $k \ge 0$*

$$e_k(x) = \frac{\alpha(x)^k}{\frac{\beta(x)}{\alpha(x)} \frac{1-\alpha(x)^k}{1-\alpha(x)} + O\left(\min\{\log k, \log \frac{1}{1-|\alpha(x)|}\}\right)},$$

where $\alpha(x) = x\Phi'(y(x))$ and $\beta(x) = \frac{1}{2}x\Phi''(y(x))$.

With the help of these representations it is relatively easy to derive the corresponding limiting result for the height.

Theorem 4.29. *For every $\kappa > 0$ we have*

$$\mathbb{P}\{H_n \ge \kappa\sqrt{n}\} = \frac{\sigma}{i\sqrt{2\pi}} \int_{-1-i\infty}^{-1+i\infty} \frac{\sqrt{-2s}\, e^{-s}}{\exp\left(\sigma\kappa\sqrt{-2s}\right) - 1}\, ds + o(1). \tag{4.49}$$

Furthermore, for all integers $r \ge 1$

$$\mathbb{E}\, H_n^r \sim \left(\frac{2}{\sigma}\right)^r 2^{-r/2} r(r-1)\Gamma(r/2)\zeta(r)\, n^{r/2}, \tag{4.50}$$

(where we make the convention $(r-1)\zeta(r) = 1$ for $r = 1$).

The integral representation (4.49) of the distribution of $M = \max_{0 \le t \le 1} e(t)$ is unusual; however, it perfectly fits into our asymptotic frame.

The proof of Proposition 4.28 relies on a precise analysis of the recurrence

$$e_{k+1}(x) = y(x) - x\Phi(y(x) - e_k(x)). \tag{4.51}$$

We start with some preliminary properties.

Lemma 4.30. *Let $h(x, v)$ be defined by*

$$\frac{\alpha(x)v}{y(x) - x\Phi(y(x) - v)} = 1 + \frac{\beta(x)}{\alpha(x)}v + v^2 h(x, v),$$

where $\alpha(x) = x\Phi'(y(x))$ and $\beta(x) = \frac{1}{2}x\Phi''(y(x))$. Then the functions $e_k(x)$ satisfy the recurrence

$$e_k(x) = \frac{e_0(x)\alpha(x)^k}{1 + e_0(x)\frac{\beta(x)}{\alpha(x)}\frac{1-\alpha(x)^k}{1-\alpha(x)} + e_0(x)\sum_{\ell<k} \alpha(x)^\ell e_\ell(x) h(x, e_\ell(x))}. \tag{4.52}$$

Note that $h(x, v)$ is actually an analytic function in v, since the Taylor series expansion of $y(x) - x\Phi(y(x) - v)$ is given by

$$y(x) - x\Phi(y(x) - v) = \alpha(x)v + \beta(x)v^2 + O(v^3) = \frac{\alpha(x)v}{1 - \frac{\beta(x)}{\alpha(x)}v + O(v^2)}.$$

Proof. By definition we have

$$e_{k+1} = y - x\Phi(y - e_k) = \frac{\alpha e_k}{1 + \frac{\beta}{\alpha} e_k + e_k^2 h}$$

or equivalently

$$\frac{\alpha}{e_{k+1}} = \frac{1}{e_k}\left(1 + \frac{\beta}{\alpha} e_k + e_k^2 h\right) = \frac{1}{e_k} + \frac{\beta}{\alpha} + e_k h.$$

By multiplying by α^k we obtain

$$\frac{\alpha^{k+1}}{e_{k+1}} = \frac{\alpha^k}{e_k} + \frac{\beta}{\alpha}\alpha^k + \alpha^k e_k h,$$

which gives

$$\frac{\alpha^k}{e_k} = \frac{1}{e_0} + \frac{\beta}{\alpha}\frac{1-\alpha^k}{1-\alpha} + \sum_{\ell<k} \alpha^\ell e_\ell h(x, e_\ell).$$

This proves the lemma.

Lemma 4.31. *For $|x| \le x_0$ we have*

$$|e_k(x)| \le e_k(x_0) \sim \frac{1}{\beta k}.$$

Proof. The inequality $|e_k(x)| \le e_k(x_0)$ is obvious, since the coefficients of e_k are non-negative. Since $y(x)$ is convergent at $x = x_0$, it follows that $\lim_{k\to\infty} e_k(x_0) = 0$. This implies that

$$\sum_{\ell<k} e_\ell(x_0) h(x_0, e_\ell(x_0)) = o(k)$$

as $k \to \infty$. Hence, since $\alpha(x_0) = 1$, by (4.52) we obtain

$$e_k(x_0) = \frac{1}{\beta k + o(k)} \sim \frac{1}{\beta k}.$$

Lemma 4.32. *There exists a constant C with the property that if $x \in \Delta$ and*

$$|e_k(x)| < \frac{1}{C}\left(\frac{1}{|\alpha(x)|} - 1\right) \tag{4.53}$$

then we have

$$|e_{k+1}(x)| < |e_k(x)|.$$

In particular, if there exists some k with (4.53) then $e_k(x) = O(\alpha(x)^k)$ as $k \to \infty$.

Proof. By Taylor's expansion we have

$$y(x) - x\Phi(y(x) - v) = \alpha(x)v(1 + O(v))$$

and consequently

$$|y(x) - x\Phi(y(x) - v)| < |\alpha(x)|\,|v|\,(1 + C|v|)$$

for a properly chosen constant $C > 0$. Hence, if (4.53) is satisfied then

$$\begin{aligned}|e_{k+1}(x)| &= |y(x) - x\Phi(y(x) - e_k(x))| \\ &< |\alpha(x)|\,|e_k(x)|\,(1 + C|e_k(x)|) \\ &\le |e_k(x)|.\end{aligned}$$

Finally, if (4.53) is satisfied for some $k = k_0$ then we can apply (4.53) recursively and consequently $e_k(x)$ stays bounded. We now use a shifted variant of (4.52), namely

$$e_k(x) = \frac{\alpha(x)^{k-k_0}}{\frac{1}{e_{k_0}} + \frac{\beta}{\alpha}\frac{1-\alpha^{k-k_0}}{1-\alpha} + \sum_{\ell<k-k_0} \alpha^\ell e_{k_0+\ell} h(x, e_{k_0+\ell})}, \tag{4.54}$$

in order to prove that $e_k = O(\alpha^k)$. We only have to check that

$$\left|\frac{\beta}{\alpha}\frac{1-\alpha^k}{1-\alpha}\right| \le \frac{1}{4}\,\frac{C|\alpha|}{1-|\alpha|}, \tag{4.55}$$

and that

$$\left|\sum_{\ell<k-k_0} \alpha^\ell e_{k_0+\ell} h(x, e_{k_0+\ell})\right| \le \|h\|_\infty \frac{1-|\alpha|}{C|\alpha|}\frac{1-|\alpha^k|}{1-|\alpha|} \le \frac{1}{4}\,\frac{C|\alpha|}{1-|\alpha|}, \tag{4.56}$$

so that (4.54) implies $e_k = O(\alpha^k)$. However, (4.55) and (4.56) are certainly satisfied, if C is chosen sufficiently large, since $x \in \Delta$ implies $|1-\alpha| \ge c(1-|\alpha|)$ for some constant $c > 0$.

Remark 4.33 *The value of the constant C can be made more precise, if we assume, for example, that we already know that $e_k(x)$ is sufficiently small. In this case we can assume that C is arbitrarily close to $\beta(x_0)$.*

With the help of these preliminaries we obtain a first approximation for $e_k(x)$, if x is far-off from the singularity.

Lemma 4.34. *For every $\epsilon > 0$ there exists $\delta > 0$ such that*

$$e_k(x) = E(x)\alpha(x)^k + O(\alpha(x)^{2k}) \tag{4.57}$$

uniformly for $|x| \le x_0 + \delta$ and $|x - x_0| \ge \epsilon$, where $E(x) = \alpha(x)(1-\alpha(x))/\beta(x)$.

Proof. If $|x - x_0| \ge \epsilon$ and $|x| \le x_0 + \delta$ then we have uniformly $|\alpha(x)| \le 1 - \eta$ for some $\eta > 0$. Now fix k_0 such that

$$e_{k_0}(x_0) < \frac{1}{2C}\frac{\eta}{1-\eta}$$

where C is the constant given in Lemma 4.32. By continuity there exists $\delta > 0$ such that

$$|e_{k_0}(x)| < \frac{1}{C}\frac{\eta}{1-\eta}$$

uniformly for $|x - x_0| \ge \epsilon$ and $|x| \le x_0 + \delta$. Without loss of generality we can assume that $\delta \le \epsilon$. Hence, we also get

$$|e_{k_0}(x)| < \frac{1}{C}\frac{\eta}{1-\eta} \le \frac{1}{C}\left(\frac{1}{|\alpha|} - 1\right).$$

Thus, the assumptions of Lemma 4.32 are satisfied and we have $e_k(x) = O(\alpha(x)^k)$. Now with help of this a priori bound and (4.52) it follows that $e_k(x)$ can be represented in the form (4.57).

It is also quite easy to discuss the case $|x - x_0| \le \epsilon$ but $|x| \le x_0$.

Lemma 4.35. *Suppose that* $|x-x_0| \le \epsilon$ *but* $|x| \le x_0$*. Then we have uniformly in this range*

$$e_k(x) = \frac{\alpha(x)^k}{\frac{\beta(x)}{\alpha(x)}\frac{1-\alpha(x)^k}{1-\alpha(x)} + O\left(\min\{\log k, \log\frac{1}{1-|\alpha(x)|}\}\right)}.$$

Proof. Since we know a priori, that $|e_k(x)| \le e_k(x_0) \sim \alpha/(\beta k)$,

$$\sum_{\ell<k} \alpha^l e_\ell h(x, e_\ell) = O\left(\min\{\log k,\, \log 1/(1-|\alpha|)\}\right),$$

as $k \to \infty$. Hence, a direct application of (4.52) completes the proof of the lemma.

It remains to discuss the asymptotic behaviour of $e_k(x)$, if $x \in \Delta$ but $|x| \ge x_0$. The problem is that it is not immediately clear that $e_k(x) \to 0$. Note, however, that $e_k(x_0) \sim 1/(\beta k)$. Thus, by continuity for every fixed k_0 there is definitely $\epsilon > 0$ such that $|e_{k_0}(x)| \le (2\alpha)/(\beta k_0)$ for $|x - x_0| \le \epsilon$ and $x \in \Delta$. However, this does not guarantee that Lemma 4.32 can be applied. We have to use a different argument.

Lemma 4.36. *There exists* $\epsilon > 0$ *and a constant* $c_1 > 0$ *such that for all* $x \in \Delta$ *with* $|x - x_0| \le \epsilon$ *we have*

$$|e_k(x)| \le \frac{1}{C}\left(\frac{1}{|\alpha(x)|} - 1\right)$$

for

$$k = K(x) = \left\lfloor \frac{c_1}{|\arg(\alpha(x))|} \right\rfloor,$$

where C is the constant from Lemma 4.32.

Proof. The idea of the proof is to use formula (4.54), where k_0 is chosen in a way that $e_{k_0}(x)$ is sufficiently small and k is approximately $c_1/|\arg(\alpha(x))|$, where c_1 has to be chosen in a proper way. Suppose that we already know that

$$|e_{k_0+\ell}(x)| \le |e_{k_0}(x)| \quad \text{for} \quad \ell \le k - k_0, \tag{4.58}$$

then

$$\left| \sum_{\ell < k-k_0} \alpha^l e_{k_0+\ell} h(x, e_{k_0+\ell}) \right| \le |e_{k_0}(x)| \, \|h\|_\infty \frac{1 - |\alpha|^{k-k_0}}{1 - |\alpha|}.$$

Since $|1-\alpha| \le c(1-|\alpha|)$ for $x \in \Delta$ (with a suitable constant $c > 0$) and $|1-\alpha^{k-k_0}| \ge 1 - |\alpha|^{k-k_0}$, it follows that this term can be made arbitrarily small compared to $(\beta/\alpha)(1-\alpha^{k-k_0})/(1-\alpha)$. Furthermore, if we assume that $\arg(x_0 - x) = \vartheta \in \left[-\frac{\pi}{2} - \epsilon', \frac{\pi}{2} + \epsilon'\right]$ (for some $\epsilon' > 0$ that has to be sufficiently small) and $r = \sigma\sqrt{2}|\sqrt{1 - x/x_0}|$, then

$$\begin{aligned} |\alpha| &= 1 - r\cos\frac{\vartheta}{2} + O(r^2), \\ \log|\alpha| &= -r\cos\frac{\vartheta}{2} + O(r^2), \\ \arg(\alpha) &= -r\sin\frac{\vartheta}{2} + O(r^2). \end{aligned}$$

Hence with $k = K(x) = \lfloor c_1/|\arg(\alpha)| \rfloor$ we have

$$|\alpha^{k-k_0}| \sim e^{-c_1\cot(\vartheta/2)+O(r^2)} \le e^{-c_1\cot(\frac{\pi}{4}+\frac{\epsilon'}{2})+O(r^2)} \le e^{-c_1(1-\epsilon'')}$$

for some arbitrarily small $\epsilon'' > 0$ (depending on ϵ'). Consequently

$$\left| \frac{\beta}{\alpha} \frac{1-\alpha^{k-k_0}}{1-\alpha} \right| \ge \left| \frac{\beta}{\alpha} \right| \frac{1 - e^{-c_1(1-\epsilon'')}}{r}.$$

Thus, if r is sufficiently small (that we can assume without loss of generality) then this term is actually much larger than $1/e_{k_0}$.

In other words, if (4.58) holds then for $k = K(x) = \lfloor c_1/|\arg(\alpha)| \rfloor$ we have

$$\frac{|\alpha|^{k-k_0}}{|e_k|} \ge \left| \frac{\beta}{\alpha} \frac{1-\alpha^{k-k_0}}{1-\alpha} \right| (1 - \epsilon'''),$$

where $\epsilon''' > 0$ can be chosen arbitrarily small, provided that k_0 and ϵ are properly chosen. Hence, if we want to conclude that $|e_k| \le (1/C)(1/|\alpha| - 1)$, or equivalently, that

$$\frac{|\alpha|^{k-k_0}}{|e_k|} \ge \frac{C|\alpha|^{k-k_0+1}}{1-|\alpha|}$$

then we just have to show that

$$\left|\frac{\beta}{\alpha}\frac{1-\alpha^{k-k_0}}{1-\alpha}\right|(1-\epsilon''') \ge \frac{C|\alpha|^{k-k_0+1}}{1-|\alpha|} \tag{4.59}$$

for some $\epsilon''' > 0$. Note that C is a constant that can be chosen arbitrarily close to β. Hence, if $k = K(x) = \lfloor c_1/|\arg(\alpha)|\rfloor$ then (4.59) is implied by the relation

$$e^{-c_1} < \frac{1}{1+1/\sqrt{2}}. \tag{4.60}$$

Namely, if (4.60) is satisfied then there exists $\epsilon' > 0$, $\epsilon''' > 0$ and some $C > \beta/\alpha$ such that

$$\left|\frac{\beta}{\alpha}\right| \frac{1-e^{-c_1\cot(\frac{\pi}{4}+\frac{\epsilon'}{2})}}{r}(1-\epsilon''') \ge C\,\frac{e^{-c_1\cot(\frac{\pi}{4}+\frac{\epsilon'}{2})}}{r\cos(\frac{\pi}{4}+\frac{\epsilon'}{2})}$$

which is asymptotically equivalent to (4.59).

It remains to be shown that (4.58) is satisfied for $k = K(x) = \lfloor c_1/|\arg(\alpha)|\rfloor$ with $c_1 > \log(1+1/\sqrt{2}) = 0.534799...$

For this purpose we show that the argument of $e_k(x)$ can be controlled. In particular, we show that $|\arg(e_\ell)| < \frac{\pi}{4} = 0.78539...$ for $\ell \le k$. If this condition is satisfied, it is clear that for $x \in \Delta$ and $|x-x_0| \le \epsilon$ we have (if e_ℓ is sufficiently small, too)

$$\left|1-\frac{\beta}{\alpha}e_\ell + O(e_\ell^2)\right| \le 1,$$

and consequently

$$|e_{\ell+1}| = |\alpha|\,|e_\ell|\,\left|1-\frac{\beta}{\alpha}e_\ell + O(e_\ell^2)\right| \le |e_\ell|.$$

Moreover, the argument of $e_{\ell+1}$ is given by

$$\begin{aligned}\arg(e_{\ell+1}) &= \arg(\alpha)+\arg(e_\ell)+\arg\left(1-\frac{\beta}{\alpha}e_\ell(1+O(e_\ell))\right)\\ &= \arg(\alpha)+\arg(e_\ell)\\ &\quad -\left|\frac{\beta}{\alpha}e_\ell\right|\sin\left(c''r\sin\frac{\vartheta}{2}+O(r^2)+\arg(e_\ell)+O(e_\ell)\right)\end{aligned}$$

for some real number c'' and $\vartheta = \arg(x_0-x)$. Suppose that $\arg(e_\ell) \ge \epsilon'''' > 0$ is large enough so that

$$c''r\sin\frac{\vartheta}{2}+O(r^2)+\arg(e_\ell)+O(e_\ell) > 0.$$

Then it follows that

$$\arg(e_{\ell+1}) \le \arg(\alpha) + \arg(e_\ell).$$

In particular, it follows by induction

$$\arg(e_\ell) \le \arg(e_{k_0}) + (\ell - k_0)\arg(\alpha) + \epsilon''''.$$

We can adapt ϵ such that $|\arg(e_{k_0})| \le \epsilon''''$. Hence, if we apply this inequality for $k - k_0 = \lfloor c_1/|\arg(\alpha)|\rfloor$ with $c_1 = \frac{\pi}{4} - 2\epsilon''''$ then it follows that

$$|\arg(e_\ell)| \le \frac{\pi}{4}$$

for $k_0 \le \ell \le k$. Consequently we actually have $|e_{\ell+1}| \le |e_\ell|$ and the induction works.

This also completes the proof of the lemma.

It is now easy to complete the proof of Proposition 4.28. Lemma 4.34 and Lemma 4.35 cover the range $|x| \le x_0 + \eta$ and $|x - x_0| \ge \epsilon$ resp. the range $|x| \le x_0$ and $|x - x_0| \le \epsilon$. In the remaining range $x \in \Delta$, $x \ge x_0$, $|x - x_0| \le \epsilon$ we use Lemma 4.36 and Lemma 4.32 to show that $e_k(x) \to 0$. From that it follows that

$$\sum_{\ell<k} \alpha^\ell e_\ell h(x, e_\ell) = o\left(\frac{1 - |\alpha|^k}{1 - |\alpha|}\right)$$

and consequently (by applying Lemma 4.30)

$$e_k = \frac{\alpha^k}{\frac{\beta}{\alpha}\frac{1-\alpha^k}{1-\alpha}(1 + o(1))}.$$

This implies the upper bound

$$e_k = O\left(\min\left\{\frac{1}{k}, (1 - |\alpha|)\alpha^k\right\}\right),$$

and consequently

$$\sum_{\ell<k} \alpha^\ell e_\ell h(x, e_\ell) = O\left(\min\left\{\log k,\, \log\frac{1}{1 - |\alpha|}\right\}\right).$$

Hence a second application of Lemma 4.30 completes the proof of Proposition 4.28.

The proof of Theorem 4.29 is based on the asymptotic formula for $e_k(x)$ that is given in Proposition 4.28. Recall that

$$\mathbb{P}\{H_n \ge k\} = \frac{1}{y_n}[x^n]e_k(x) = \frac{1}{y_n}\frac{1}{2\pi i}\int_\Gamma e_k(x)\,\frac{dx}{x^{n+1}},$$

where we use the same contour $\Gamma = \gamma' \cup \Gamma'$ as in the proof of Proposition 4.21. Again it is sufficient to concentrate on the integral over

$$\gamma' = \left\{ x = x_0 \left(1 + \frac{s}{n}\right) : \Re(s) = -1,\ |\Im(s)| \le C \log^2 n \right\}.$$

For $x \in \Gamma'$ the functions $e_k(x)$ are uniformly bounded and hence the integral over Γ' is bounded above by $O(x_0^{-n} e^{-C\log^2 n})$. On the other hand, if $k = \lfloor \kappa\sqrt{n} \rfloor$ the integral over γ' is asymptotically given by

$$\begin{aligned} \frac{1}{2\pi i} \int_{\gamma'} e_k(x)\, \frac{dx}{x^{n+1}} &\sim x_0^{-n} n^{-3/2} \frac{1}{2\pi i} \int_{-1-i\infty}^{-1+i\infty} \frac{\exp\left(-\sigma\kappa\sqrt{-2s}\right)}{\beta \frac{1-\exp\left(-\sigma\kappa\sqrt{-2s}\right)}{\sigma\sqrt{-2s}}} e^{-s}\, ds \\ &= x_0^{-n} n^{-3/2} \frac{\tau}{2\pi i} \int_{-1-i\infty}^{-1+i\infty} \frac{\sqrt{-2s}e^{-s}}{\exp\left(\sigma\kappa\sqrt{-2s}\right) - 1}\, ds. \end{aligned}$$

Note that $\beta(x_0) = \sigma/\tau$. Since $y_n \sim \tau/(\sigma\sqrt{2\pi}) x_0^{-n} n^{-3/2}$, we obtain the proposed asymptotic representation (4.49).

In order to obtain the asymptotics for the moments (4.50) we observe that the generating function of the r-th moment $\mathbb{E}\, H_n^r$ is given by

$$\sum_{n\ge 1} \mathbb{E}\, H_n^r\, y_n x^n = \sum_{k \ge 1} \left(k^r - (k-1)^r\right) e_k(x).$$

Since $k^r - (k-1)^r \sim r k^{r-1}$ and $e_k \sim (\alpha/\beta)(1-\alpha)\alpha^k/(1-\alpha^k)$, we have (as $x \to x_0$ in Δ)

$$\sum_{n\ge 1} \mathbb{E}\, H_n^r\, y_n x^n \sim r \frac{\alpha}{\beta}(1-\alpha) \sum_{k\ge 1} k^{r-1} \frac{\alpha^k}{1-\alpha^k}.$$

This kind of sum can be analysed with help of the following lemma.

Lemma 4.37. *Let $R \ge 0$ be a fixed real number. Then the function*

$$D_R(z) = \sum_{k\ge 1} k^R \frac{z^k}{1-z^k}$$

has radius of convergence 1 *and we have, as $z \to 1$ with $|\arg(1-z)| \le \phi$ (for some $\phi < \pi/2$),*

$$D_R(z) \sim \begin{cases} \dfrac{1}{1-z} \log \dfrac{1}{1-z} & \text{for } R = 0, \\ \dfrac{R!\zeta(R+1)}{(1-z)^{R+1}} & \text{for } R > 0. \end{cases}$$

Proof. We use the substitution $z = e^{-x}$. Then the Mellin transform of the function

$$F_R(x) = \sum_{k\ge 1} k^R \frac{e^{-kx}}{1-e^{-kx}}$$

is given by

$$F_R^*(s) = \int_0^\infty F_R(x) x^{s-1}\, dx = \Gamma(s)\zeta(s)\zeta(s-R), \qquad (\Re(s) > R+1).$$

Hence, a standard inversion by using a shift of the line of integration in the inversion formula

$$F_R(x) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} F_R^*(s) x^s\, ds$$

to $\sigma < R+1$ and collecting the contributions of the polar singularity at $s = R+1$ leads to the asymptotic leading term for the behaviour $x \to 0+$. For $R = 0$ we have a double pole corresponding to $(1/x)\log(1/x)$, whereas for $R > 0$ there is a single pole with residue $\Gamma(R+1)\zeta(R+1)x^{R+1}$ (for details see [81]).

Note further that this procedure also works, if $x \to 0$ in the complex plane, provided that $|\arg(x)| \le \phi$ with $\phi < \pi/2$.

This implies that for $r = 1$ we obtain

$$\sum_{n\ge 1} \mathbb{E}\, H_n\, y_n x^n \sim \frac{1}{2\beta} \log \frac{1}{1 - x/x_0}$$

and consequently

$$y_n \mathbb{E}\, H_n \sim \frac{x_0^{-n}}{2\beta} \frac{1}{n}.$$

Since $\beta = \sigma^2/(2\tau)$ and $y_n \sim \tau/(\sigma\sqrt{2\pi}) x_0^{-n} n^{-3/2}$, we get

$$\mathbb{E}\, H_n \sim \frac{\sqrt{2\pi}}{\sigma} \sqrt{n}$$

as proposed.

Finally, for $r > 1$ we have

$$\sum_{n\ge 1} \mathbb{E}\, H_n^r\, y_n x^n \sim \frac{r!\zeta(r)}{\beta(\sigma\sqrt{2})^{r-1}} \left(1 - \frac{x}{x_0}\right)^{\frac{r-1}{2}}$$

which leads to

$$\mathbb{E}\, H_n^r \sim \frac{2r!\zeta(r)\sqrt{2\pi}}{\sigma^r 2^{(r-1)/2} \Gamma(\frac{r-1}{2})} n^{r/2}.$$

In view of the duplication formula for the Gamma function this is equivalent to (4.50).

Remark 4.38 *With slightly more case it is also possible to derive a local version of Theorem 4.29, that is, an asymptotic expansion for* $\mathbb{P}\{H_n = k\}$ *if* k *is of order* $\sqrt{n}$ *(see [80]).*

4.2.8 Depth-First Search

Aldous' result on the normalised depth-first search process of Galton-Watson trees (Theorem 4.7), saying that

$$\left(\frac{\sigma}{2\sqrt{n}}X_n(2nt), 0 \le 1 \le t\right) \xrightarrow{d} (e(t), 0 \le t \le 1)$$

is a fundamental observation in the framework of Galton-Watson trees. It implies convergence to the continuum random tree (Theorem 4.6) and also proves that the normalised occupation measure μ_n converges to the occupation measure of the Brownian excursion. Recall that Theorem 4.10 is a local version of the latter observation.

Theorem 4.7 was first proved in the framework of the continuum random tree [4]. One major step in Aldous' method is to consider, for every fixed integer $k \ge 1$, random subtrees of Galton-Watson trees of size n that are spanned by k random vertices and the root and to show that there is a proper limiting object $\mathcal{R}_k$.[5] These limiting objects characterise then the limit of the depth-first search process. Aldous shows that (up to a scaling factor $1/\sigma$) all Galton-Watson trees have the same limiting object $\mathcal{R}_k$. Since the normalised depth-first search process for planted plane trees ($\Phi(x) = 1/(2-x)$) is known to be the Brownian excursion, we obtain the same result for all Galton-Watson trees. However, in order to check the existence of the limiting objects $\mathcal{R}_k$ one has to invest several precise asymptotic properties of Galton-Watson trees, in particular of the height distribution (compare with [4] and [132]). The advantage of that approach is its universality, since it is not restricted to Galton-Watson trees.

In what follows we indicate an alternative proof with the help of combinatorial and analytic methods. As in the case of the profile we first prove finite dimensional convergence and then tightness (compare with [92]).

Lemma 4.39. *Let $x(\cdot)$ denote the depth-first search process of a rooted tree and set*

$$a_{k,m,n} = \sum_{|T|=n,\, x(m)=k} \omega(T) = y_n \, \mathbb{P}\{X_n(m) = k\}.$$

Then the generating function $A_k(x,u) = \sum_{m,n} a_{k,m,n} x^n u^k$ is given by

$$A_k(x,u) = A(x,u)\Phi_1(x,u,1)^{k-1}, \tag{4.61}$$

where

$$A(x,u) = xu\frac{y(xu^2)\Phi(y(xu^2)) - y(x)\Phi(y(x))}{y(xu^2) - y(x)}$$

and

$$\Phi_1(x,u,v) = xuv\frac{\Phi(y(xu^2)) - \Phi(y(xv^2))}{y(xu^2) - y(xv^2)}.$$

[5] We do not give a precise definition of these objects. For detail we refer to [4]).

Proof. Suppose that $k = 1$, that is, the depth-first search process passes the root. Recall that a Galton-Watson tree can be described as a root with $i \geq 0$ subtrees. Thus, there exists j with $0 \leq j \leq i$ such that the first j subtrees have been already traversed but not the $(j+1)$-st one. The corresponding generating function is then given by

$$\begin{aligned} A_1(x,u) &= xu \sum_{i \geq 0} \phi_i \sum_{j=0}^{i} y(xu^2)^j y(x)^{i-j} \\ &= A(x,u). \end{aligned}$$

If $k > 1$ then there exists a subtree of the root, in which the depth-first search process passes a corresponding node at level $k-1$. Hence, we get

$$\begin{aligned} A_k(x,u) &= xu \sum_{i \geq 0} \phi_i \sum_{j=0}^{i-1} y(xu^2)^j A_{k-1}(x,u) y(x)^{i-j-1} \\ &= A_{k-1}(x,u) \Phi_1(x,u,1). \end{aligned}$$

By induction this completes the proof of the lemma.

The explicit representation of $A_k(x,u)$ leads to an asymptotic expansion for $a_{k,m,n}$

Proposition 4.40. *Suppose that $\kappa > 0$ and $0 < t < 1$ are fixed real numbers. Then for $k = \lfloor \kappa \sqrt{n} \rfloor$ and $m = \lfloor 2tn \rfloor$ we have, as $n \to \infty$,*

$$\frac{a_{k,m,n}}{y_n} \sim \frac{1}{4\sqrt{2\pi n}} \frac{\sigma^3 \kappa^2}{(t(1-t))^{3/2}} \exp\left(-\frac{\sigma^2 \kappa^2}{8t(1-t)}\right). \tag{4.62}$$

Proof. By Cauchy's formula

$$a_{k,m,n} = \frac{1}{(2\pi i)^2} \int_{\Gamma} \int_{\tilde{\Gamma}} A_k(x,u) \frac{du}{u^{m+1}} \frac{dx}{x^{n+1}},$$

where Γ and $\tilde{\Gamma}$ a properly chosen contours around the origin. Here we use (as in the proof of Proposition 4.21) $\Gamma = \gamma' \cup \Gamma'$, where

$$\gamma' = \left\{ x = x_0 \left(1 + \frac{s}{n}\right) : \Re(s) = -1, \, |\Im(s)| \leq C \log^2 n \right\}$$

with an arbitrarily chosen fixed constant $C > 0$ and Γ' is a circular arc centred at the origin and making Γ a closed curve. Similarly, $\tilde{\Gamma} = \tilde{\gamma}' \cup \tilde{\Gamma}'$ with

$$\tilde{\gamma}' = \left\{ u = 1 + \frac{s_1}{m} : \Re(s_1) = -1, \, |\Im(s_1)| \leq \tilde{C} \log^2 m \right\}$$

with a properly chosen constant $\tilde{C}$ and a circular arc $\tilde{\Gamma}'$ making $\tilde{\Gamma}$ a closed curve.

We will only concentrate on γ' resp. on $\tilde{\gamma}'$. The remaining parts of the integral are negligible (similarly to the proof of Proposition 4.21).

For $x \in \gamma'$ and $u \in \tilde{\gamma}'$ we use the approximations

$$y(x) = \tau - \frac{\tau\sqrt{2}}{\sigma}\sqrt{-\frac{s}{n}} + O\left(\frac{|s|}{n}\right),$$
$$y(xu^2) = \tau - \frac{\tau\sqrt{2}}{\sigma}\sqrt{-\frac{s}{n} - \frac{2s_1}{m}} + O\left(\frac{|s|}{n} + \frac{|s_1|}{m}\right),$$
$$A(x,u) = 2\tau + O\left(\sqrt{\frac{|s|}{n}} + \sqrt{\frac{|s_1|}{m}}\right),$$
$$\Phi_1(x,u,1) = 1 - \frac{\sigma}{\sqrt{2}}\left(\sqrt{-\frac{s}{n}} + \sqrt{-\frac{s}{n} - \frac{2s_1}{m}}\right) + O\left(\frac{|s|}{n} + \frac{|s_1|}{m}\right).$$

We fix $\kappa > 0$ and $0 < t < 1$ and set $k = \lfloor \kappa\sqrt{n} \rfloor$ and $m = \lfloor 2tn \rfloor$. Thus, we obtain the integral approximation

$$\frac{1}{(2\pi i)^2}\int_{\gamma'}\int_{\tilde{\gamma}'} A_k(x,u)\,\frac{dx}{x^{n+1}}\,\frac{du}{u^{m+1}}$$
$$\sim \frac{x_0^{-n}}{mn}\,\frac{1}{(2\pi i)^2}\int_{-1-iC\log^2 n}^{-1+iC\log^2 n}\int_{-1-i\tilde{C}\log^2 m}^{-1+i\tilde{C}\log^2 m}$$
$$2\tau\exp\left(-\frac{\sigma\kappa}{\sqrt{2}}\left(\sqrt{-s} + \sqrt{-s - \frac{2n}{m}s_1}\right)\right) e^{-s-s_1}\,ds_1\,ds$$
$$\sim \frac{x_0^{-n}}{mn}\,\frac{1}{(2\pi i)^2}\int_{-1-i\infty}^{-1+i\infty}\int_{-1-i\infty}^{-1+i\infty}$$
$$2\tau\exp\left(-\frac{\sigma\kappa}{\sqrt{2}}\left(\sqrt{-s} + \sqrt{-s - \frac{1}{t}s_1}\right)\right) e^{-s-s_1}\,ds_1\,ds.$$

By using the substitution $s_1 = w - ts$, a proper shift of integration and by Lemma 4.22 we can evaluate the double integral:

$$\frac{1}{(2\pi i)^2}\int_{-1-i\infty}^{-1+i\infty}\int_{-1-i\infty}^{-1+i\infty}\exp\left(-\frac{\sigma\kappa}{\sqrt{2}}\left(\sqrt{-s} + \sqrt{-s - \frac{1}{t}s_1}\right)\right) e^{-s-s_1}\,ds_1\,ds$$
$$= \frac{1}{(2\pi i)^2}\int_{-1-i\infty}^{-1+i\infty}\exp\left(-\frac{\sigma\kappa}{\sqrt{2}}\sqrt{-s}\right) e^{-s(1-t)}$$
$$\times\int_{-1+ts-i\infty}^{-1+ts+i\infty}\exp\left(-\frac{\sigma\kappa}{\sqrt{2t}}\sqrt{-w}\right) e^{-w}\,dw\,ds$$
$$= \frac{1}{(2\pi i)^2}\int_{-1-i\infty}^{-1+i\infty}\exp\left(-\frac{\sigma\kappa}{\sqrt{2}}\sqrt{-s}\right) e^{-s(1-t)}$$
$$\times\int_{-1-i\infty}^{-1+i\infty}\exp\left(-\frac{\sigma\kappa}{\sqrt{2t}}\sqrt{-w}\right) e^{-w}\,dw\,ds$$

$$= \frac{1}{4\pi} \frac{\sigma^2\kappa^2}{t^{1/2}(1-t)^{3/2}} \exp\left(-\frac{\sigma^2\kappa^2}{8(1-t)} - \frac{\sigma^2\kappa^2}{8t}\right).$$

Hence, (4.62) follows.

Note that (4.62) is in accordance with the density of the one dimensional density of $\frac{2}{\sigma}e(t)$ (compare with (4.3)). By inspecting the above proof it is clear that all estimates are also uniform, if κ varies in a compact interval (contained in the positive real line). Hence, we also get weak convergence for every $0 < t < 1$:

$$\frac{\sigma}{2\sqrt{n}} X_n(2nt) \xrightarrow{d} e(t).$$

Similarly one can work out the finite dimensional case. For example, set

$$\begin{aligned} b_{k_1,m_1,k_2,m_2,n} &= \sum_{|T|=n,\, x(m_1)=k_1,\, x(m_2)=k_2} \omega(T) \\ &= y_n \mathbb{P}\{X_n(m_1) = k_1,\ X_n(m_2) = k_2\}. \end{aligned}$$

Then we have

$$\begin{aligned} B_{k_1,k_2}(x,u_1,u_2) &= \sum_{m_1,m_2,n\ge 0} b_{k_1,m_1,k_2,m_2,n} x^n u_1^{m_1} u_2^{m_2} \\ &= A(x(u_1u_2)^2, 1)\, A(x,u)\, \Phi_2(x,u_1u_2,u_2) \\ &\quad \times \sum_{r=0}^{\min\{k_1,k_2\}-1} \Phi_1(x,u_1u_2,u_2)^{k_1-1-r} \Phi_1(x,u_2,1)^{k_2-1-r} \Phi_1(x,u_1u_2,1)^r, \end{aligned}$$

where

$$\Phi_2(x,u,v) = x \sum_{i\ge 2} \phi_i \sum_{j_1+j_2+j_3=i-2} y(xu^2)^{j_1} y(xv^2)^{j_2} y(x)^{j_3}.$$

With the help of these explicit representations one obtains two dimensional convergence for $0 < t_1 < t_2 < 1$:

$$\left(\frac{\sigma}{2\sqrt{n}} X_n(2nt_1), \frac{\sigma}{2\sqrt{n}} X_n(2nt_2)\right) \xrightarrow{d} (e(t_1), e(t_2)).$$

The general case is even more involved (compare with [58, 92]).

Finally, with the help of $B_{k_1,k_2}(x,u_1,u_2)$ it is also possible to prove tightness (see [92]).

Proposition 4.41. *Set*

$$e_n(t) := \frac{\sigma}{2\sqrt{n}} X_n(2nt).$$

Then there exist constants $C, D > 0$ such that for all $s,t \in [0,1]$ and $\varepsilon > 0$,

$$\mathbb{P}\{|e_n(s) - e_n(t)| \ge \epsilon\} \le \frac{C}{\sqrt{|s-t|}} \exp\left(-D\frac{\epsilon}{\sqrt{|s-t|}}\right). \tag{4.63}$$

Hence it follows that

$$\mathbb{P}\{|e_n(s) - e_n(t)| \ge \epsilon\} \le C \frac{|s-t|^{\alpha - \frac{1}{2}}}{\epsilon^{2\alpha}}$$

for every $\alpha > 3/2$. Thus, by [18, Theorem 12.3] the sequence of processes $e_n(t)$ is tight.

Proof. The proof of Proposition 4.41 uses the explicit representation for

$$\begin{aligned} C_\ell(x,u_1,u_2) &= \sum_{|k_1-k_2|\ge \ell} B_{k_1k_2}(x,u_1,u_2) \\ &= \frac{A(x(u_1u_2)^2,1)\, A(x,u_2)\, \Phi_2(x,u_1u_2,u_2)}{(1-\Phi_1(x,u_1u_2,u_2)\Phi_1(x,u_2,1))(1-\Phi_1(x,u_1u_2,1))} \\ &\quad \times \left(\frac{\Phi_1(x,u_1u_2,u_2)^\ell}{1-\Phi_1(x,u_1u_2,u_2)} + \frac{\Phi_1(x,u_2,1)^\ell}{1-\Phi_1(x,u_2,1)} \right). \end{aligned}$$

Since $u_1^{m_1}u_2^{m_2} = (u_1u_2)^{m_1}u_2^{m_2-m_1}$ and we are interested in the dependence of the difference $d = m_2 - m_1$, it is appropriate to use the substitution $u = u_1u_2$ and to estimate the coefficient of $x^n u^{m_1} u_2^d$.

For this purpose one uses Cauchy's formula for each variable, where we use the same kind of integration as in the proof of Proposition 4.40, that is $x \in \Gamma$ and $u, u_2 \in \Gamma'$. By using the local expansion of

$$\Phi_1(x,u,v) = 1 - \frac{\sigma}{\sqrt{2}} \left(\sqrt{1-\frac{xu^2}{x_0}} + \sqrt{1-\frac{xv^2}{x_0}} \right) + O\left(\left|1-\frac{xu^2}{x_0}\right| + \left|1-\frac{xv^2}{x_0}\right| \right)$$

one obtains for $x \in \Gamma$ and $u, u_2 \in \Gamma'$ the bounds

$$\begin{aligned} |(1-\Phi_1(x,u,u_2)\Phi_1(x,u_2,1))(1-\Phi_1(x,u,1))(1-\Phi_1(x,u,u_2))| &\ge \frac{C_1}{m_1\sqrt{d}}, \\ |(1-\Phi_1(x,u,u_2)\Phi_1(x,u_2,1))(1-\Phi_1(x,u,1))(1-\Phi_1(x,u_2,1))| &\ge \frac{C_1}{m_1\sqrt{d}}, \end{aligned}$$

and

$$\max\{|\Phi_1(x,u,u_2)|, |\Phi_1(x,u_2,1)|\} \le e^{-C_2\sqrt{|u_2|/d}}$$

for some constants $C_1, C_2 > 0$. Hence it follows that

$$[x^n u^{m_1} u^d]\, C_\ell(x,u/u_2,u_2) = O\left(x_0^{-n} \frac{m_1\sqrt{d}}{nm_1 d} e^{-C_3\ell/\sqrt{d}} \right)$$

for some constant $C_3 > 0$. Consequently

$$\begin{aligned} \mathbb{P}\{|X_n(m_1) - X_n(m_1+d)| \ge \ell\} &= \frac{[x^n u^{m_1} u_2^d]\, C_\ell(x,u/u_2,u_2)}{y_n} \\ &= O\left(\sqrt{n/d}\, e^{-C_3\ell/\sqrt{d}} \right). \end{aligned}$$

This proves (4.63), if $m_1 = 2t_1 n$ and $m_2 = 2t_2 n$ are integers. However, it is an easy exercise to extend this estimate for the interpolated process for all $0 < t_1 < t_2 < 1$ (compare with the comments following Theorem 4.24). This completes the proof of Proposition 4.41.

4.3 The Profile of Pólya Trees

We analyse next the profile of unlabelled rooted random trees, also called Pólya trees. In Section 3.1.5 we have already shown that the analysis of this kind of trees is quite similar to that of Galton-Watson trees. Moreover, the shape of Pólya trees has the same asymptotic behaviour as that of Galton-Watson trees. The height is of order $\sqrt{n}$ and – this is the main focus of this section – the profile $L_n(k)$ can be approximated by the local time of Brownian excursion, too, although the uniform probability model cannot be realised on Pólya trees as Galton-Watson trees. (We follow Drmota and Gittenberger [65].)

Theorem 4.42. *Let $L_n(k)$ denote the number of nodes at distance k to the root in random Pólya trees of size n and $l(t)$ the local time of a Brownian excursion of duration* 1*. Then*

$$\left(\frac{1}{\sqrt{n}} L_n\left(t\sqrt{n}\right), t \geq 0\right) \xrightarrow{d} \left(\frac{b\sqrt{\rho}}{2\sqrt{2}} \cdot l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}} \cdot t\right), t \geq 0\right)$$

in $\mathrm{C}[0,\infty)$*, as $n \to \infty$, where $b \approx 2.6811266$ is the constant and $\rho \approx 0.3383219$ the singularity appearing in equation* (3.7)*.*

The structure of the proof of Theorem 4.42 is very similar to that of Theorem 4.10. Following Theorem 4.14 we prove convergence of finite dimensional distributions and tightness. We start with the combinatorial setup that is based on an involved recurrence; then we asymptotically solve this recurrence which leads to the proof of weak convergence of the finite dimensional distributions. Finally, we check tightness according to the moment conditions of Theorem 4.14.

We will also consider the height of Pólya trees in Section 4.3.5.

4.3.1 Combinatorial Setup

First we recall some facts on unlabelled rooted trees $\mathcal{T}$. Let t_n denote the number of these trees of size n. Then the generating function

$$t(x) = \sum_{n \geq 1} t_n x^n$$

satisfies the functional equation

$$t(x) = x \exp\left(t(x) + \frac{1}{2} t(x^2) + \frac{1}{3} t(x^3) + \cdots \right). \tag{4.64}$$

It has radius of convergence $\rho \approx 0.3383219$ and a local singular expansion of the form $t(x) = 1 - b(\rho - x)^{1/2} + c(\rho - x) + \cdots$, where $b \approx 2.6811266$, and $c = b^2/3 \approx 2.3961466$. Recall, too, that this singular expansion can be used to obtain an asymptotic formula for t_n (compare with (3.9)).

In order to compute the distribution of the number of nodes in some given levels $k_1 < k_2 < \cdots < k_d$ in a tree of size n we have to calculate the number $t_{k_1,m_1,k_2,m_2,\cdots k_d,m_d;n}$ of trees of size n with m_i nodes in level k_i, $i = 1, \ldots, d$ and normalise by t_n. As in the case of Galton-Watson trees we introduce the generating functions

$$t_{k_1,k_2,\ldots,k_d}(x, u_1, \ldots, u_d) = \sum_{m_1,\ldots,m_d,n} t_{k_1,m_1,k_2,m_2,\cdots k_d,m_d;n}\, u_1^{m_1} u_2^{m_2} \cdots u_d^{m_d}\, x^n.$$

They can be recursively calculated. However, the procedure is more involved than that of Galton-Watson trees.

Lemma 4.43. *For $d = 1$ we have*

$$t_0(x, u) = u t(x) \tag{4.65}$$

$$t_{k+1}(x, u) = x \exp\left(\sum_{i \ge 1} \frac{t_k(x^i, u^i)}{i} \right), \qquad (k \ge 0), \tag{4.66}$$

and for $d > 1$ with integers $0 \le k_1 < k_2 < k_3 < \cdots < k_d$

$$t_{0,k_2,\ldots,k_d}(x, u_1, \ldots, u_d) = u_1 t_{k_2,\ldots,k_d}(x, u_2, \ldots, u_d) \tag{4.67}$$

$$t_{k_1+1,k_2+1,\ldots,k_d+1}(x, u_1, \ldots, u_d) = x \exp\left(\sum_{i \ge 1} \frac{t_{k_1,k_2,\ldots,k_d}(x^i, u_1^i, u_2^i, \ldots, u_d^i)}{i} \right). \tag{4.68}$$

Proof. The proof is principally the same as that of Lemma 4.18. The only difference is that one has to apply the multiset construction for ordinary generating functions (compare with Section 2.1.1), now applied for several variables, if we proceed from one level to the next.

There is, however, a slight difference to the Galton-Watson case. There we could define a function $Y_k(x, u)$ (see (4.31)) so that

$$\begin{aligned} &y_{k_1,\ldots,k_d}(x, u_1, \ldots, u_d) \\ &= Y_{k_1}\Big(x, u_1 Y_{k_2-k-1}\big(x, u_2 Y_{k_3-k_2}(x, \ldots, u_{d-1} Y_{k_d-k_{d-1}}(x, u y(x)) \cdots)\big)\Big). \end{aligned}$$

Such a representation is not present for Pólya trees. Therefore we have to be more precise in the asymptotic analysis. Even the case $d = 1$ requires much more care than in the Galton-Watson case.

4.3.2 Asymptotic Analysis of the Main Recurrence

The asymptotic structure of $t_k(x,u)$ is very similar to that of $y_k(x,u)$ (in the context of Galton-Watson trees).

Proposition 4.44. *Let $x = \rho\left(1+\frac{s}{n}\right)$, where $\Im(s) = -1$, $u = e^{i\tau/\sqrt{n}}$, and $k = \lfloor \kappa\sqrt{n} \rfloor$. Moreover, assume that, as $n \to \infty$, we have $s = O\left(\log^2 n\right)$ whereas τ and κ are fixed. Then we have, as $n \to \infty$,*

$$\begin{aligned} t_k(x,u) - t(x) &\sim \frac{b^2\rho}{2\sqrt{n}} \cdot \frac{i\tau\sqrt{-s}\exp\left(-\frac{1}{2}\kappa b\sqrt{-\rho s}\right)}{\sqrt{-s}\exp\left(\frac{1}{2}\kappa b\sqrt{-\rho s}\right) - \frac{i\tau b\sqrt{\rho}}{2}\sinh\left(\frac{1}{2}\kappa b\sqrt{-\rho s}\right)} \\ &= \frac{b\sqrt{2\rho}}{\sqrt{n}}\Psi_{\kappa b\sqrt{\rho}/(2\sqrt{2})}\left(s, \frac{i\tau b\sqrt{\rho}}{2\sqrt{2}}\right). \end{aligned} \tag{4.69}$$

As in the case of Galton-Watson trees we introduce the quantity

$$w_k(x,u) = t_k(x,u) - t(x). \tag{4.70}$$

Obviously, $w_k(x,1) = 0$. Hence, we expect that $w_k(x,u) \to 0$ as $k \to \infty$, if u is sufficiently close to 1 and x is in the analyticity range of $t(x)$.

We start with a useful property of $t(x)$.

Lemma 4.45. *The generating function $t(x)$ has the following properties:*

1. *For $|x| \le \rho$ we have $|t(x)| \le 1$. Equality holds only for $x = \rho$.*
2. *Let $x = \rho\left(1+\frac{s}{n}\right)$ with $\Im(s) = -1$ and $|s| \le C\log^2 n$ for some fixed $C > 0$. Then there is a $c > 0$ such that*

$$|t(x)| \le 1 - c\sqrt{\frac{|s|}{n}}.$$

Proof. The first statement follows from the fact that $t(x)$ has only positive coefficients (except $t_0 = 0$), $t(\rho) = 1$ and there are no periodicities. The second statement is an immediate consequence of the singular expansion (3.7) of $t(x)$.

Next let us consider the derivative

$$\gamma_k(x,u) = \frac{\partial}{\partial u} t_k(x,u).$$

Lemma 4.46. *For $|x| < \rho + \eta$ and $\arg(x-\rho) \ne 0$ (where $\eta > 0$ is sufficiently small) the functions $\gamma_k(x,1)$ can be represented as*

$$\gamma_k(x,1) = C_k(x) t(x)^{k+1}, \tag{4.71}$$

where $C_k(x)$ are analytic and converge uniformly to an analytic limit function $C(x)$ (for $|x| < \rho + \eta$ and $\arg(x-\rho) \ne 0$) with convergence rate

$$C_k(x) = C(x) + O(L^k),$$

for some L with $0 < L < 1$. Furthermore we have $C(\rho) = \frac{1}{2} b^2 \rho$.

There also exist constants $c_1, c_2, c_3 > 0$ with $c_3 < \frac{\pi}{2}$ such that

$$|\gamma_k(x,u)| = O(|t(x)|^k) \tag{4.72}$$

uniformly for $|u| \le 1$, $k|u-1| \le c_1$, $|x-\rho| < c_2$ and $|\arg(x-\rho)| \ge c_3$.

Proof. A tree that has nodes at level k must have a size larger than k. Thus $[x^r]t_k(x,u)$ does not depend on u for $r \le k$. Consequently, the lowest order non-vanishing term in the power series expansion of $\gamma_k(x) = \gamma_k(x,1)$ is of order $k+1$. The power series expansion of $t(x)$ starts with x. Hence $C_k(x) = \gamma_k(x) t(x)^{-k-1}$ is analytic for $|x| < \rho + \eta$ and $\arg(x-\rho) \ne 0$. We will show that the sequence $(C_k(x))_{k \ge 0}$ has a uniform limit $C(x)$ which has the required properties.

Using the recurrence relation of $t_k(x,u)$ we get

$$\begin{aligned}
\gamma_{k+1}(x,u) &= \frac{\partial}{\partial u} x \exp\left(\sum_{i \ge 1} \frac{t_k(x^i,u^i)}{i} \right) \\
&= x \exp\left(\sum_{i \ge 1} \frac{t_k(x^i,u^i)}{i} \right) \sum_{i \ge 1} \frac{\partial}{\partial u} t_k(x^i,u^i)\, u^{i-1} \\
&= t_{k+1}(x,u) \sum_{i \ge 1} \gamma_k(x^i,u^i)\, u^{i-1}.
\end{aligned} \tag{4.73}$$

Setting $u = 1$ this rewrites to

$$C_{k+1}(x)t(x)^{k+2} = C_k(x)t(x)^{k+2} + t(x)\left(C_k(x^2)t(x^2)^{k+1} + C_k(x^3)t(x^3)^{k+1} + \dots\right), \tag{4.74}$$

resp. to

$$C_{k+1}(x) = \sum_{i \ge 1} C_k(x^i) \frac{t(x^i)^{k+1}}{t(x)^{k+1}}. \tag{4.75}$$

Set

$$L_k := \sup_{|x|<\rho+\eta,\ \arg(x-\rho)\ne 0} \sum_{i \ge 2} \frac{|t(x^i)|^{k+1}}{|t(x)|^{k+1}}.$$

If $\eta > 0$ is sufficiently small then

$$\sup_{|x|<\rho+\eta,\ \arg(x-\rho)\ne 0} \frac{|t(x^i)|}{|t(x)|} < 1$$

for all $i \ge 2$ and

$$\sup_{|x|<\rho+\eta,\ \arg(x-\rho)\ne 0} \frac{|t(x^i)|}{|t(x)|} = O(\overline{L}^i)$$

for some $\overline{L}$ with $0 < \overline{L} < 1$. Consequently we also get

$$L_k = O(L^k)$$

for some L with $0 < L < 1$. Thus, if we use the notation

$$\|f\| = \sup_{|x|<\rho+\eta,\ \arg(x-\rho)\neq 0} |f(x)|$$

then (4.75) implies

$$\|C_{k+1}\| \le \|C_k\|(1 + L_k) \tag{4.76}$$

and also

$$\|C_{k+1} - C_k\| \le \|C_k\| L_k. \tag{4.77}$$

Now (4.76) implies that the functions $C_k(x)$ are uniformly bounded in Δ by

$$\|C_k\| \le c_0 := \prod_{\ell \ge 1} (1 + L_\ell).$$

Furthermore, (4.77) implies that there exists a limit $\lim_{k\to\infty} C_k(x) = C(x)$ that is analytic for $|x| < \rho+\eta$, $\arg(x-\rho) \neq 0$; and we have uniform exponential convergence rate

$$\|C_k - C\| \le c_0 \sum_{\ell \ge k} L_\ell = O(L^k).$$

Hence, we get (4.71).

Finally, note that (for $|x| \le \rho$)

$$\sum_{k\ge 0} \gamma_k(x,1) = \sum_{n\ge 1} n t_n x^n.$$

However, if we use (4.71) we get uniformly for $|x| < \rho + \eta$ and $\arg(x - \rho) \neq 0$

$$\begin{aligned}\sum_{k\ge 0} \gamma_k(x,1) &= \sum_{k\ge 0} (C(x) + L^k) t(x)^{k+1} \\ &= C(x) \frac{t(x)}{1 - t(x)} + O(1)\end{aligned}$$

which leads to an asymptotic expansion of the n-th coefficient of the form

$$n t_n \sim \frac{C(\rho)}{b\sqrt{\pi\rho}} n^{-\frac{1}{2}} \rho^{-n}.$$

Note that here we have used the fact that $C(x)$ is a continuous function as the uniform limit of continuous functions $C_k(x)$, where $x = \rho$ is also a point of continuity. By comparing this with the (known) expansion for $t_n \sim \frac{1}{2} b\sqrt{\rho/\pi} n^{-3/2} \rho^{-n}$ we obtain $C(\rho) = \frac{1}{2} b^2 \rho$.

In order to obtain the upper bound (4.72) we set for $\ell \le k$

$$\overline{C}_\ell = \sup |\gamma_\ell(x,u) t(x)^{-\ell-1}|,$$

where the supremum is over (x,u) with $|u|\le 1$, $k|u-1|\le c_1$, $|x-\rho|<c_2$ and $|\arg(x-\rho)|\ge c_3$, where the constants $c_1,c_2,c_3>0$ (with $c_3<\frac{\pi}{2}$) will be chosen in the sequel. From (4.73) we get (as above) the inequality

$$\overline{C}_{\ell+1}\le \overline{C}_\ell\, e^{c_1\overline{C}_\ell/k}\,(1+O(L^\ell)), \tag{4.78}$$

where we have implicitly used the inequality

$$\begin{aligned}|t_{\ell+1}| &\le |t(x)|\exp\left(\sum_{i\ge 1}\frac{|w_\ell(x^i,u^i)|}{i}\right)\\ &\le |t(x)|e^{\overline{C}_\ell|u-1|+O(L^\ell)}.\end{aligned}$$

Set

$$c_0=\prod_{j\ge 0}(1+O(L^j))$$

and choose $c_1>0$ such that $e^{2c_0c_1}\le 2$. We also choose $c_2\le\eta$ and $0<c_3<\frac{\pi}{2}$ such that $|t(x)|\le 1$ for $|x-\rho|<c_2$ and $|\arg(x-\rho)|\ge c_3$. Now if $k>0$ is fixed, it follows by induction that if $|u-1|\le c_1/k$

$$\overline{C}_\ell\le\prod_{j<\ell}(1+O(L^j))\cdot e^{2c_0c_1\ell/k}\le 2c_0 \qquad (\ell\le k).$$

This completes the proof of the lemma.

The representation (4.71) from Lemma 4.46 gives us a first indication of the behaviour of $w_k(x,u)$ for u close to 1. We expect that

$$w_k(x,u)\approx (u-1)\gamma_k(x,1)\sim (u-1)C(x)t(x)^k. \tag{4.79}$$

This actually holds (up to constants) in a proper range for u and x, although it is only partially true in the range of interest (see Proposition 4.44).

In order to make this more precise we derive estimates for the second derivative

$$\gamma_k^{[2]}(x,u)=\frac{\partial^2}{\partial u^2}t_k(x,u).$$

Lemma 4.47. *Suppose that $|x|\le\rho-\eta$ for some $\eta>0$ and $|u|\le 1$. Then uniformly*

$$\gamma_k^{[2]}(x,u)=O(t(|x|)^{k+1}). \tag{4.80}$$

There also exist constants $c_1,c_2,c_3>0$ with $c_3<\frac{\pi}{2}$ such that

$$\gamma_k^{[2]}(x,u)=O(k\,|t(x)|^{k+1}) \tag{4.81}$$

uniformly for $|u|\le 1$, $k|u-1|\le c_1$, $|x-\rho|<c_2$ and $|\arg(x-\rho)|\ge c_3$.

Proof. By definition we have the recurrence

$$\begin{aligned}\gamma_{k+1}^{[2]}(x,u) &= t_{k+1}(x,u)\sum_{i\ge 1} i\gamma_k^{[2]}(x^i,u^i)u^{2i-2}\\ &\quad + t_{k+1}(x,u)\left(\sum_{i\ge 1}\gamma_k(x^i,u^i)u^{i-1}\right)^2 \\ &\quad + t_{k+1}(x,u)\sum_{i\ge 2}(i-1)\gamma_k(x^i,u^i)u^{i-2}\end{aligned} \tag{4.82}$$

with initial condition $\gamma_0^{[2]}(x) = 0$.

First suppose that $|x| \le \rho - \eta$ for some $\eta > 0$ and $|u| \le 1$. Then we have $|\gamma_k^{[2]}(x,u)| \le \gamma_k^{[2]}(|x|,1)$. Thus, in this case it is sufficient to consider non-negative real $x \le \rho - \eta$. We proceed by induction. Suppose that we already know that $\gamma_k^{[2]}(x) = \gamma_k^{[2]}(x,1) \le D_k t(x)^{k+1}$ (where $D_0 = 0$). Then we get from (4.82) and the already known bound $\gamma_k(x,1) \le Ct(x)^k$ the upper bound

$$\begin{aligned}\gamma_{k+1}^{[2]}(x) &\le D_k t(x)^{k+2} + D_k t(x)^{k+2}\sum_{i\ge 2} i\frac{t(x^i)^{k+1}}{t(x)^{k+1}}\\ &\quad + C^2 t(x)^{2k-1} + Ct(x)^{k+2}\sum_{i\ge 2}(i-1)\frac{t(x^i)^{k+1}}{t(x)^{k+1}}\\ &\le t(x)^{k+2}\left(D_k(1+O(L^k)) + C^2 t(\rho-\eta)^k + O(L^k)\right).\end{aligned}$$

Consequently we can set

$$D_{k+1} = D_k(1+O(L^k)) + C^2 t(\rho-\eta)^k + O(L^k)$$

and obtain that $D_k = O(1)$ as $k \to \infty$ which proves (4.80).

Next set (for $\ell \le k$)

$$\overline{D}_\ell = \sup\left|\gamma_\ell^{[2]}(x,u)t(x)^{-\ell-1}\right|,$$

where the supremum is taken over $|u| \le 1$, $k|u-1| \le c_1$, $|x-\rho| < c_2$ and $|\arg(x-\rho)| \ge c_3$ and where $c_1, c_2, c_3 > 0$ are chosen as in the proof of Lemma 4.46. By the same reasoning as in the proof of Lemma 4.46, where we use the already proved bound $|\gamma_\ell(x,u) \le \overline{C}|t(x)|^{\ell+1}$ we obtain

$$\overline{D}_{\ell+1} \le \overline{D}_\ell\, e^{c_1 C/k}(1+O(L^\ell)) + C^2\, e^{c_1 C/k} + O(L^\ell),$$

that is, we have

$$\overline{D}_{\ell+1} \le \alpha_\ell \overline{D}_\ell + \beta_\ell$$

with $\alpha_\ell = e^{c_1 C/k}(1+O(L^\ell))$ and $\beta_\ell = C^2\, e^{c_1 C/k} + O(L^\ell)$. Hence we get

$$\begin{aligned}\overline{D}_k &\le \sum_{j=0}^{k-1} \beta_j \prod_{i=j+1}^{k-1} \alpha_i \\ &\le k \max_j \beta_j\, e^{c_1 C} \prod_{\ell \ge 0} (1 + O(L^\ell)) \\ &= O(k).\end{aligned}$$

This completes the proof of (4.81).

Using the estimates for $\gamma_k(x,u)$ and $\gamma_k^{[2]}(x,u)$ we derive the following representations (4.83) for w_k and $\Sigma_k(x,u)$, where

$$\Sigma_k(x,u) = \sum_{i\ge 2} \frac{w_k(x^i, u^i)}{i}.$$

Lemma 4.48. *There exist $c_1, c_2, c_3 > 0$ with $c_3 < \frac{\pi}{2}$ such that*

$$w_k(x,u) = C_k(x)(u-1)t(x)^{k+1}\left(1 + O(k|u-1|)\right) \tag{4.83}$$

uniformly for $|u| \le 1$, $k|u-1| \le c_1$, $|x-\rho| < c_2$ and $|\arg(x-\rho)| \ge c_2$, where $C_k(x) = \gamma_k(x,1)/t(x)^{k+1}$ is given in Lemma 4.46.

Furthermore we have for $|x| \le \rho + \eta$ (for some $\eta > 0$) and $|u| \le 1$

$$\Sigma_k(x,u) = \tilde{C}_k(x)(u-1)t(x^2)^{k+1} + O\left(|u-1|^2 t(|x|^2)^k\right), \tag{4.84}$$

where the analytic functions $\tilde{C}_k(x)$ are given by

$$\tilde{C}_k(x) = \sum_{i\ge 2} C_k(x^i) \left(\frac{t(x^i)}{t(x^2)}\right)^{k+1}. \tag{4.85}$$

They have a uniform limit $\tilde{C}(x)$ with convergence rate

$$\tilde{C}_k(x) = \tilde{C}(x) + O(L^k)$$

for some constant L with $0 < L < 1$.

Proof. The first relation (4.83) follows from Lemma 4.46, Lemma 4.47 and Taylor's theorem.

In order to prove (4.84) we first note that $|x^i| \le \rho - \eta$ for $i \ge 2$ and $|x| \le \rho + \eta$ (if $\eta > 0$ is sufficiently small). Hence, by a second use of Taylor's theorem we get uniformly

$$w_k(x^i, u^i) = C_k(x^i)(u^i - 1)t(x^i)^{k+1} + O\left(|u^i - 1|^2 t(|x^i|)^{k+1}\right)$$

and consequently

$$\begin{aligned}\Sigma_k(x,u) &= \sum_{i\ge 2} \frac{1}{i} C_k(x^i)(u^i-1)t(x^i)^{k+1} + O\left(|u-1|^2 t(|x|^2)^k\right) \\ &= (u-1)\tilde{C}_k(x)t(x^2)^{k+1} + O\left(|u-1|^2 t(|x|^2)^k\right).\end{aligned}$$

Here we have used the property that the sum

$$\sum_{i\ge 2} C_k(x^i) \frac{u^i-1}{i(u-1)} \frac{t(x^i)^{k+1}}{t(x^2)^{k+1}}$$

represents an analytic function in x and u. Finally, since $C_k(x) = C(x) + O(L^k)$, it also follows that $\tilde{C}_k(x)$ has a limit $\tilde{C}(x)$ and the same order of convergence.

With these auxiliary results we are able to get a precise result for $w_k(x,u)$.

Lemma 4.49. *There exists positive constants c_1, c_2, c_3 with $c_3 < \frac{\pi}{2}$ such that*

$$w_k(x,u) = \frac{C_k(x) w_0(x,u)\, t(x)^k}{1 - \frac{1}{2} C_k(x) w_0(x,u) \frac{1-t(x)^k}{1-t(x)} + O\left(|u-1|\right)}$$

as long as $|u| \le 1$, $k|u-1| \le c_1$, $|x-\rho| \le c_2$, and $|\arg(x-\rho)| \ge c_3$.

Proof. Since $\Sigma_k(x,u) = O(w_k(x,u)L^k) = O(w_k(x,u))$ (see Lemma 4.48), we observe that $w_k(x,u)$ satisfies the recurrence relation (we omit the arguments now)

$$\begin{aligned}w_{k+1} &= t\left(e^{w_k+\Sigma_k} - 1\right) \\ &= t\left(w_k + \frac{w_k^2}{2} + \Sigma_k + O(w_k^3) + O(\Sigma_k^2)\right) \\ &= t w_k \left(1 + \frac{w_k}{2} + O\left(w_k^2\right) + O\left(\Sigma_k\right)\right)\left(1 + \frac{\Sigma_k}{w_k}\right).\end{aligned}$$

Equivalently we have

$$\begin{aligned}\frac{t}{w_{k+1}} + \frac{t\,\Sigma_k}{w_k w_{k+1}} &= \frac{1}{w_k}\left(1 - \frac{w_k}{2} + O\left(w_k^2\right) + O\left(\Sigma_k\right)\right) \\ &= \frac{1}{w_k} - \frac{1}{2} + O\left(w_k\right) + O\left(\frac{\Sigma_k}{w_k}\right),\end{aligned}$$

and consequently

$$\frac{t^{k+1}}{w_{k+1}} = \frac{t^k}{w_k} - \frac{\Sigma_k t^{k+1}}{w_k w_{k+1}} - \frac{1}{2} t^k + O\left(w_k t^k\right) + O\left(\frac{\Sigma_k t^k}{w_k}\right).$$

Thus we get

$$\frac{t^k}{w_k} = \frac{1}{w_0} - \sum_{\ell=0}^{k-1} \frac{\Sigma_\ell}{w_\ell w_{\ell+1}} t^{\ell+1} - \frac{1}{2}\frac{1-t^k}{1-t} + O\left(\frac{1-L^k}{1-L}\right) + O\left(w_0 \frac{1-t^{2k}}{1-t^2}\right).$$

Now we use again Lemma 4.48 to obtain

$$\begin{aligned} w_0 \sum_{\ell=0}^{k-1} \frac{\Sigma_\ell}{w_\ell w_{\ell+1}} t^{\ell+1} &= \sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell(x) t(x^2)^{\ell+1} + O\left(|u-1| t(|x|^2)^\ell\right)}{C_\ell(x) C_{\ell+1}(x) t(x)^{\ell+1} (1 + O(\ell|u-1|))} \\ &= \sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell(x)}{C_\ell(x) C_{\ell+1}(x)} \frac{t(x^2)^{\ell+1}}{t(x)^{\ell+1}} + O(|u-1|) \\ &= c_k(x) + O(u-1) \end{aligned}$$

with a proper function $c_k(x)$. Observe, too, that $w_0 \frac{1-t^{2k}}{1-t^2} = O(1)$, if $k|u-1| \le c_1$. Hence we obtain the representation

$$w_k = \frac{w_0 t^k}{1 - c_k(x) - \frac{w_0}{2}\frac{1-t^k}{1-t} + O\left(|u-1|\right)}. \tag{4.86}$$

Thus, it remains to verify that $1 - c_k(x) = 1/C_k(x)$. By using (4.75) and (4.85) it follows that

$$\tilde{C}_k(x) = \sum_{i\ge 2} C_k(x^i) \left(\frac{t(x^i)}{t(x^2)}\right)^{k+1} = \left(C_{k+1}(x) - C_k(x)\right) \left(\frac{t(x)}{t(x^2)}\right)^{k+1} \tag{4.87}$$

and consequently by telescoping

$$c_k(x) = \sum_{\ell=0}^{k-1} \frac{C_{\ell+1}(x) - C_\ell(x)}{C_\ell(x) C_{\ell+1}(x)} = \frac{1}{C_0(x)} - \frac{1}{C_k(x)} = 1 - \frac{1}{C_k(x)}.$$

Alternatively we can compare (4.86) with (4.83) for $u \to 1$ which also shows $1 - c_k(x) = 1/C_k(x)$. This completes the proof of the lemma.

The proof of Proposition 4.44 is now immediate. We substitute $x = \rho\left(1 + \frac{s}{n}\right)$ (where $\Im(s) = -1$), $u = e^{i\tau/\sqrt{n}}$ and set $k = \kappa\sqrt{n}$. We also use the local expansion $t(x) = 1 - b\sqrt{\rho}\sqrt{1 - x/\rho} + O(|x-\rho|)$. That leads to

$$t(x)^k = \exp\left(-\kappa b\sqrt{-\rho s}\right)\left(1 + O\left(\frac{\kappa}{\sqrt{n}}\right)\right).$$

Finally, since the functions $C_k(x)$ are continuous and uniformly convergent to $C(x)$, they are also uniformly continuous and, thus, $C_k(x) \sim C(\rho) = \frac{1}{2}b^2\rho$. Altogether this leads to

$$\frac{w_0(x,u)\,t(x)^k}{\frac{1}{C_k(x)} - \frac{w_0(x,u)}{2}\frac{1-t(x)^k}{1-t(x)} + O\left(|u-1|\right)}$$
$$= \frac{C_k(x)(u-1)(1-t(x))\,t(x)^{k+1}}{1-t(x)-C_k(x)\frac{w_0(x,u)}{2}(1-t(x)^k)+O\left(|u-1|\cdot|1-t(x)|\right)}$$
$$\sim \frac{b^2\rho}{2\sqrt{n}}\cdot\frac{i\tau\sqrt{-s}\exp\left(-\frac12\kappa b\sqrt{-\rho s}\,\right)}{\sqrt{-s}\exp\left(\frac12\kappa b\sqrt{-\rho s}\,\right)-\frac{i\tau b\sqrt{\rho}}{2}\sinh\left(\frac12\kappa b\sqrt{-\rho s}\,\right)}$$

as proposed.

4.3.3 Finite Dimensional Limiting Distributions

We will use the results of the previous section to prove finite dimensional convergence. We only discuss the cases $d=1$ and $d=2$ in detail. As above, $(l(t), t\ge 0)$ denotes the local time of the Brownian excursion.

Proposition 4.50. *Let $\kappa>0$ and τ be given with $|\kappa\tau|\le c$. Then we have*

$$\lim_{n\to\infty}\mathbb{E}e^{i\tau\frac{1}{\sqrt{n}}L_n(\kappa\sqrt{n})} = \mathbb{E}\,e^{i\tau\frac{b\sqrt{\rho}}{2\sqrt{2}}\,l(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa)}, \tag{4.88}$$

and consequently

$$\frac{1}{\sqrt{n}}L_n(\kappa\sqrt{n}) \xrightarrow{d} \frac{b\sqrt{\rho}}{2\sqrt{2}}\cdot l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa\right).$$

Proof. By definition the generating function $t_k(x,u)$ can be interpreted as

$$t_k(x,u)=\sum_{n\ge1}t_n\,\mathbb{E}\,u^{L_n(k)}\,x^n$$

and consequently

$$\mathbb{E}e^{i\tau L_n(k)/\sqrt{n}} = \frac{1}{t_n}[x^n]t_k\left(x, e^{i\tau/\sqrt{n}}\right).$$

In order to get asymptotics for this characteristic function we will use the local representation for $t_k(x,u)$ of Proposition 4.44 and the same contour integration as in the proof of Proposition 4.21.

Let $\Gamma=\gamma'\cup\Gamma'$ consist of a line

$$\gamma'=\left\{x=\rho\left(1+\frac{s}{n}\right): \Re(s)=-1,\ |\Im(s)|\le C\log^2 n\right\}$$

with an arbitrarily chosen fixed constant $C>0$ and of Γ', a circular arc centred at the origin and making Γ a closed curve. Then

$$\mathbb{E}e^{i\tau L_n(k)/\sqrt{n}} = \frac{1}{t_n}\frac{1}{2\pi i}\int_{\Gamma} t_k\left(x, e^{i\tau/\sqrt{n}}\right)\frac{dx}{x^{n+1}} \tag{4.89}$$

$$= 1 + \frac{1}{t_n}\frac{1}{2\pi i}\int_{\Gamma} w_k\left(x, e^{i\tau/\sqrt{n}}\right)\frac{dx}{x^{n+1}}. \tag{4.90}$$

The contribution of Γ' is exponentially small, since for $x \in \Gamma'$ we have $|x^{-n-1}| \sim \rho^{-n}e^{-C\log^2 n}$, whereas $\left|y_k\left(x, e^{i\tau/\sqrt{n}}\right)\right|$ is bounded.

If $x \in \gamma'$ then the local expansion (4.69) is valid for a proper range for k. In particular, we replace k by $\lfloor \kappa\sqrt{n}\rfloor$, u by $e^{i\tau/\sqrt{n}}$ and x by $x_0\left(1+\frac{s}{n}\right)$. If we assume that $\kappa > 0$ is given then the assumptions of Proposition 4.44 are satisfied, if $\kappa\tau \le c$. Finally, by applying the approximation $x^{-n} = x_0^{-n}e^{-s}(1+O(|s|^2/n))$ we observe that the integral

$$\frac{1}{2\pi i}\int_{\gamma'} w_k\left(x, e^{i\tau/\sqrt{n}}\right)\frac{dx}{x^{n+1}}$$

is approximated by

$$\frac{b\sqrt{\rho}\,\rho^{-n}}{2\sqrt{\pi}\,n^{3/2}}\frac{\sqrt{2}}{\sqrt{\pi}}\int_{-1-i\infty}^{1+i\infty}\frac{\frac{i\tau b}{2}\sqrt{-\rho s}\exp(-\frac12\kappa b\sqrt{-\rho s}\,)}{\sqrt{-2s}\exp(\frac12\kappa b\sqrt{-\rho s}\,) - \frac{i\tau b\sqrt{\rho}}{\sqrt{2}}\sinh\left(\frac12\tau b\sqrt{-\rho s}\right)}\,e^{-s}\,ds.$$

Since $t_n \sim \frac12 b\sqrt{\rho/\pi}\rho^{-n}n^{-3/2}$ we thus obtain (4.88) for $\tau \le c/\kappa$. But this is sufficient to prove weak convergence (compare with the proof of Proposition 4.21).

The corresponding property in the two dimensional case is the following one.

Proposition 4.51. *Let $\kappa_2 > \kappa_1 > 0$ and τ_1, τ_2 be given with $|\kappa_2\tau_1| \le c$ and $|\kappa_2\tau_2| \le c$. Then we have*

$$\lim_{n\to\infty}\mathbb{E}\,e^{i\tau_1\frac{1}{\sqrt{n}}L_n(\kappa_1\sqrt{n})+i\tau_2\frac{1}{\sqrt{n}}L_n(\kappa_2\sqrt{n})} = \mathbb{E}\,e^{i\tau_1\frac{b\sqrt{\rho}}{2\sqrt{2}}\,l(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa_1)+i\tau_2\frac{b\sqrt{\rho}}{2\sqrt{2}}\,l(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa_2)} \tag{4.91}$$

and consequently

$$\left(\frac{1}{\sqrt{n}}L_n(\kappa_1\sqrt{n}),\ \frac{1}{\sqrt{n}}L_n(\kappa_2\sqrt{n})\right) \xrightarrow{d} \left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\cdot l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa_1\right), \frac{b\sqrt{\rho}}{2\sqrt{2}}\cdot l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa_1\right)\right).$$

Proof. In order to prove two dimensional convergence we need precise asymptotic properties of $t_{k_1,k_2}(x,u_1,u_2)$ for $k_1 = \kappa_1\sqrt{n}$, $k_2 = \kappa_2\sqrt{n}$, $u_1 = e^{i\tau_1/\sqrt{n}}$ and $u_2 = e^{i\tau_2/\sqrt{n}}$. Similarly to the above we define

$$w_{k_1,k_2}(x,u_1,u_2) = t_{k_1,k_2}(x,u_1,u_2) - t(x)$$

so that $w_{k_1,k_2}(x,1,1) = 0$. We further have $w_{k_1,k_2}(x,u_1,1) = w_{k_1}(x,u_1)$ and $w_{k_1,k_2}(x,1,u_2) = w_{k_2}(x,u_2)$. Thus, the first derivatives are given by

$$\left[\frac{\partial}{\partial u_1} w_{k_1,k_2}(x,u_1,u_2)\right]_{u_1=u_2=1} = \gamma_{k_1}(x,1),$$
$$\left[\frac{\partial}{\partial u_2} w_{k_1,k_2}(x,u_1,u_2)\right]_{u_1=u_2=1} = \gamma_{k_2}(x,1).$$

It is also possible to get bounds for the second derivatives of the form $O(k_2 t(x)^{k_1})$, if $k_1|u_1-1| \le c$, $k_2|u_2-1| \le c$, $|x-\rho| < c$, and $|\arg(x-\rho)| \ge c$. Hence, we can approximate $w_{k_1,k_2}(x,u_1,u_2)$ by

$$\begin{aligned} w_{k_1,k_2}(x,u_1,u_2) &= C_{k_1}(x)(u_1-1)t(x)^{k_1+1} + C_{k_2}(x)(u_2-1)t(x)^{k_2+1} \\ &\quad + O\left(k_2 t(x)^{k_1}(|u_1-1|^2+|u_2-1|^2)\right). \end{aligned} \tag{4.92}$$

Similarly (and even more easily) we obtain a representation for

$$\begin{aligned} \Sigma_{k_1,k_2}(x,u_1,u_2) &= \sum_{i\ge 2} \frac{w_{k_1,k_2}(x^i,u_1^i,u_2^i)}{i} \\ &= \tilde{C}_{k_1}(x)(u_1-1)t(x^2)^{k_1+1} + \tilde{C}_{k_2}(x)(u_2-1)t(x^2)^{k_2+1} \\ &\quad + O\left(t(|x|^2)^{k_1}|u_1-1|^2 + t(|x|^2)^{k_2}|u_2-1|^2\right). \end{aligned} \tag{4.93}$$

Since $k_2-k_1 \sim (\kappa_2-\kappa_1)\sqrt{n}$ and $t(x^2)<1$ for $|x| \le \rho+\eta$, the term $t(x^2)^{k_2+1}$ is significantly smaller than $t(x^2)^{k_1+1}$. Thus, in what follows it will be sufficient to consider just the first term $\tilde{C}_{k_1}(x)(u_1-1)t(x^2)^{k_1+1}$.

We also assume that $u_1 = e^{i\tau_1/\sqrt{n}}$ and $u_2 = e^{i\tau_2/\sqrt{n}}$, that is, u_1-1 and u_2-1 are (asymptotically) proportional. Since we have for $x=\rho(1+s/n)$

$$t(x)^{k_2-k_1} \sim \exp\left(-(\kappa_2-\kappa_1)b\sqrt{-\rho s}\right)$$

and since we can shift the line $\Re(s)=-1$ of integration to any parallel line $\Re(s)=\sigma_0$ with $\sigma_0<0$, we can always assure that the leading terms of w_{k_1,k_2}, that is, $C_{k_1}(x)(u_1-1)t(x)^{k_1+1}$ and $C_{k_2}(x)(u_2-1)t(x)^{k_2+1}$, do not cancel to lower order terms.

By using the same reasoning as in the proof of Lemma 4.49 we get the representation

$$w_{k_1,k_2} = \frac{w_{0,k_2-k_1}t^{k_1}}{1-f_{k_1}-\frac{w_{0,k_2-k_1}}{2}\frac{1-t^{k_1}}{1-t}+O(|u_1-1|+|u_2-1|)},$$

where

$$\begin{aligned} f_{k_1} &= f_{k_1}(x,u_1,u_2) \\ &= w_{0,k_2-k_1}(x,u_1,u_2)\sum_{\ell=0}^{k_1-1}\frac{\Sigma_{\ell,k_2-k_1+\ell}(x,u_1,u_2)\,t(x)^{\ell+1}}{w_{\ell,k_2-k_1+\ell}(x,u_1,u_2)w_{\ell+1,k_2-k_1+\ell+1}(x,u_1,u_2)}. \end{aligned}$$

Note that

$$\begin{aligned}w_{0,k_2-k_1} &= u_1 t_{k_2-k_1}(x,u_2) - t(x)\\ &= (u_1-1)t(x) + u_1 w_{k_2-k_1}(x,u_2)\\ &= U+W,\end{aligned}$$

where U and W abbreviate $U = (u_1-1)t(x)$ and $W = u_1 w_{k_2-k_1}(x,u_2)$. By the above assumption we can assume that $|W| < \frac{1}{2}|U|$ so that there is no cancellation.

Next by (4.92) and Lemma 4.49 it follows that $w_{\ell,k_2-k_1+\ell}$ can be represented by

$$w_{\ell,k_2-k_1+\ell} = (C_\ell(x)U + W)\, t(x)^\ell \,(1 + O(\ell|u_1-1| + (\ell+k_2-k_1)|u_2-1|))\,.$$

Hence, f_k can be approximated by (for simplicity we omit the error terms)

$$\begin{aligned}f_{k_1} &\sim U(U+W)\sum_{\ell=0}^{k_1-1} \frac{\tilde{C}_\ell(x)\,(t(x^2)/t(x))^{\ell+1}}{(C_\ell(x)U+W)(C_{\ell+1}(x)U+W)}\\ &= (U+W)\sum_{\ell=0}^{k_1-1} \frac{(C_{\ell+1}(x)U+W)-(C_\ell(x)U+W)}{(C_\ell(x)U+W)(C_{\ell+1}(x)U+W)}\\ &= (U+W)\left(\frac{1}{U+W} - \frac{1}{C_kU+W}\right)\\ &= 1 - \frac{U+W}{C_kU+W},\end{aligned}$$

where we have used the formula (4.87) and telescoping. Consequently, it follows that

$$\begin{aligned}w_{k_1,k_2} &\sim \frac{(U+W)t^{k_1}}{\frac{U+W}{C_kU+W} - \frac{U+W}{2}\frac{1-t^{k_1}}{1-t}}\\ &= \frac{(C_kU+W)t^{k_1}}{1-\frac{C_kU+W}{2}\frac{1-t^{k_1}}{1-t}}.\end{aligned}$$

Now if we use the approximations

$$\begin{aligned}C_k(x) &\sim \frac{b^2\rho}{2},\\ U = t(u_1-1) &\sim \frac{i\tau_1}{\sqrt{n}},\\ W = u_1 w_{k_2-k_1}(x,u_2) &\sim \frac{b\sqrt{2\rho}}{\sqrt{n}}\,\Psi_{\frac{(\kappa_2-\kappa_1)b\sqrt{\rho}}{2\sqrt{2}}}\left(s, \frac{i\tau_2 b\sqrt{\rho}}{2\sqrt{2}}\right),\\ t^{k_1} &\sim \exp\left(-\frac{1}{2}\kappa_1 b\sqrt{-\rho s}\right),\\ 1-t &\sim b\sqrt{\rho}\sqrt{\frac{s}{n}},\end{aligned}$$

we finally get

$$w_{k_1,k_2} \sim \frac{b\sqrt{2\rho}}{\sqrt{n}} \Psi_{\frac{\kappa_1 b\sqrt{\rho}}{2\sqrt{2}}}\left(s, \frac{i\tau_1 b\sqrt{\rho}}{2\sqrt{2}} + \Psi_{\frac{(\kappa_2-\kappa_1)b\sqrt{\rho}}{2\sqrt{2}}}\left(s, \frac{i\tau_2 b\sqrt{\rho}}{2\sqrt{2}}\right)\right).$$

(Recall, that $\Psi_\kappa(s,t)$ is defined in (4.14).)

Using this approximation it follows that the characteristic function of the random vector $\left(\frac{1}{\sqrt{n}}L_n(k_1), \frac{1}{\sqrt{n}}L_n(k_2)\right)$ where $k_1 = \lfloor \kappa_1\sqrt{n}\rfloor$ and $k_2 = \lfloor \kappa_2\sqrt{n}\rfloor$ is asymptotically given by

$$\begin{aligned}
&\mathbb{E}\, e^{it_1 L_n(k_1)/\sqrt{n} + it_2 L_n(k_2)/\sqrt{n}} \\
&\quad = \frac{1}{t_n}[x^n] t_{k_1,k_2}\left(x, e^{it_1/\sqrt{n}}, e^{it_2/\sqrt{n}}\right) \\
&\quad = 1 + \frac{1}{t_n}\frac{1}{2\pi i}\int_\Gamma w_{k_1,k_2}\left(x, e^{it_1/\sqrt{n}}, e^{it_2/\sqrt{n}}\right)\frac{dx}{x^{n+1}} \\
&\quad = 1 + \frac{\sqrt{2}}{i\sqrt{\pi}}\int_{\sigma_0 - i\infty}^{\sigma_0+i\infty} \Psi_{\frac{\kappa_1 b\sqrt{\rho}}{2\sqrt{2}}}\left(s, \frac{i\tau_1 b\sqrt{\rho}}{2\sqrt{2}} + \Psi_{\frac{(\kappa_2-\kappa_1)b\sqrt{\rho}}{2\sqrt{2}}}\left(s, \frac{i\tau_2 b\sqrt{\rho}}{2\sqrt{2}}\right)\right) e^{-s}\, ds \\
&\quad + o(1).
\end{aligned}$$

This completes the proof of the two dimensional case.

The general finite dimensional case now follows from an iterative use of the above techniques.

4.3.4 Tightness

In this section we will show that the sequence of random variables $l_n(t) = n^{-1/2}L_n(t\sqrt{n})$, $t \ge 0$ is tight in C$[0,\infty)$. As in the case of Galton-Watson trees it is sufficient to prove the following property (compare with Theorem 4.24).

Theorem 4.52. *There exists a constant $C > 0$ such that*

$$\mathbb{E}\,(L_n(k) - L_n(k+h))^4 \le C\,h^2 n \tag{4.94}$$

holds for all non-negative integers n, k, h.

The fourth moment in the above equation can be expressed as the coefficient of a suitable generating function. We have

$$\begin{aligned}
&\mathbb{E}\,(L_n(k) - L_n(k+h))^4 \\
&\quad = \frac{1}{t_n}[x^n]\left[\left(\frac{\partial}{\partial u} + 7\frac{\partial^2}{\partial u^2} + 6\frac{\partial^3}{\partial u^3} + \frac{\partial^4}{\partial u^4}\right) t_{k,k+h}\left(x, u, \frac{1}{u}\right)\right]_{u=1}
\end{aligned}$$

where $t_{k,k+h}(x, u_1, u_2)$ is recursively given in Lemma 4.43. Thus, (4.94) is equivalent to

$$[x^n]\left[\left(\frac{\partial}{\partial u}+7\frac{\partial^2}{\partial u^2}+6\frac{\partial^3}{\partial u^3}+\frac{\partial^4}{\partial u^4}\right)t_{k,k+h}\left(x,u,\frac{1}{u}\right)\right]_{u=1}\le C\,\frac{h^2}{\sqrt{n}}\rho^{-n}. \tag{4.95}$$

Hence, it is sufficient to show that

$$\begin{aligned}\left[\left(\frac{\partial}{\partial u}+7\frac{\partial^2}{\partial u^2}+6\frac{\partial^3}{\partial u^3}+\frac{\partial^4}{\partial u^4}\right)t_{k,k+h}\left(x,u,\frac{1}{u}\right)\right]_{u=1}&\\ =O\left(\frac{h^2}{1-|t(x)|}\right)&\\ =O\left(\frac{h^2}{\sqrt{|1-x/\rho|}}\right)&\end{aligned} \tag{4.96}$$

for $x\in\Delta$ and $h\ge 1$. Recall that we have $1-|t(x)|\ge c\sqrt{|1-x/\rho|}$ for $x\in\Delta$ (for some constant $c>0$).

For convenience we use the notation[6]

$$\gamma_k^{[j]}(x)=\left[\frac{\partial^j t_k(x,u)}{\partial u^j}\right]_{u=1}\qquad\text{and}\qquad\gamma_{k,h}^{[j]}(x)=\left[\frac{\partial^j t_{k,k+h}\left(x,u,\frac{1}{u}\right)}{\partial u^j}\right]_{u=1}.$$

The main part of the proof of Theorem 4.52 is to provide upper bounds for these derivatives. We have already derived upper bounds for $\gamma^{[1]}(x)$ and $\gamma^{[2]}(x)$ (compare with Lemma 4.46 and Lemma 4.47). However, we have to be more precise. On the other hand we use the convexity bound $|t(x)|\le t(|x|)\le|x/\rho|$ for $|x|\le\rho$ in order to simplify the calculations.

Lemma 4.53. *We have*

$$\gamma_k^{[1]}(x)=\begin{cases}O(1) & \textit{uniformly for } x\in\Delta,\\ O\left(|x/\rho|^k\right) & \textit{uniformly for } |x|\le\rho,\end{cases} \tag{4.97}$$

and

$$\gamma_{k,h}^{[1]}(x)=\begin{cases}O\left(\frac{h}{k+h}\right) & \textit{uniformly for } x\in\Delta,\\ O\left(|x/\rho|^k\right) & \textit{uniformly for } |x|\le\rho,\end{cases} \tag{4.98}$$

where L is constant with $0<L<1$.

Proof. We already know that $\gamma_k^{[1]}(x)=C_k(x)t(x)^k$, where $C_k(x)=O(1)$ and $|t(x)|\le 1$ for $x\in\Delta$ (see Lemma 4.46). By convexity we also have $|t(x)|\le|x/\rho|$ for $|x|\le\rho$. Hence, we obtain $\gamma_k^{[1]}(x)=O\left(|x/\rho|^k\right)$ for $|x|\le\rho$.

The functions $\gamma_{k,h}^{[1]}(x)$ are given by the recurrence

$$\gamma_{k+1,h}^{[1]}(x)=t(x)\sum_{i\ge 1}\gamma_{k,h}^{[1]}(x^i)$$

[6] In Section 4.3.3 we already used the notation $\gamma(x,u)$ for the first derivative and $\gamma^{[2]}(x,u)$ for the second derivative. There will be no confusion, the only difference is that we now always have $u=1$.

with initial value $\gamma_{0,h}^{[1]}(x) = t(x) - \gamma_h(x)$. Hence, the representation $\gamma_{k,h}^{[1]}(x) = \gamma_k^{[1]}(x) - \gamma_{h+k}^{[1]}(x)$ follows by induction. Since $\gamma_k^{[1]}(x) = (C(x) + O(L^k))t(x)^k$, we thus get that

$$\gamma_{k,h}^{[1]}(x) = O\left(\sup_{x\in\Delta} |t(x)^k(1-t(x)^h)| + L^k\right).$$

However, it is an easy exercise to show that

$$\sup_{x\in\Delta} |t(x)^k(1-t(x)^h)| = O\left(\frac{h}{k+h}\right). \tag{4.99}$$

For this purpose observe that if $x \in \Delta$ then we either have $|t(x)-1| \le 1$ and $|t(x)| \le 1$, or $|t(x)| \le 1-\eta$ for some $\eta > 0$. In the second case we surely have

$$|t(x)^k(1-t(x)^h)| \le 2(1-\eta)^k = O\left(L^k\right).$$

For the first case we set $t = 1 - \rho e^{i\phi}$ and observe that

$$\left|1 - (1-\rho e^{i\phi})^h\right| \le (1+\rho)^h - 1.$$

Hence, if $k \ge 3h$ we thus obtain that

$$|t(x)^k(1-t(x)^h)| \le \max_{0\le\rho\le1} (1-\rho)^k\left((1+\rho)^h - 1\right) \le \frac{h}{k-h} \le \frac{2h}{k+h}.$$

If $k < 3h$,we obviously have

$$|t(x)^k(1-t(x)^h)| \le 2 \le \frac{4h}{k+h}$$

which completes the proof of (4.99). Of course, we also have $L^k = O\left(\frac{h}{h+k}\right)$. This completes the proof of the upper bound of $\gamma_{k,h}^{[1]}(x)$ for $x \in \Delta$.

Finally, the upper bound $\gamma_{k,h}^{[1]}(x) = O\left(|x/\rho|^k\right)$ follows from (4.97).

Lemma 4.54. *We have*

$$\gamma_k^{[2]}(x) = \begin{cases} O\left(\min\left\{k, \frac{1}{1-|t(x)|}\right\}\right) & \text{uniformly for } x \in \Delta, \\ O\left(|x/\rho|^k\right) & \text{uniformly for } |x| \le \rho - \eta, \end{cases} \tag{4.100}$$

and

$$\gamma_{k,h}^{[2]}(x) = \begin{cases} O\left(\min\left\{h, \frac{1}{1-|t(x)|}\right\}\right) & \text{uniformly for } x \in \Delta, \\ O\left(|x/\rho|^k\right) & \text{uniformly for } |x| \le \rho - \eta, \end{cases} \tag{4.101}$$

for every $\eta > 0$.

Proof. The bound $\gamma_k^{[2]}(x) = O(|x/\rho|^k)$ (for $|x| \le \rho - \eta$) and the bound $\gamma_k^{[2]}(x) = O(k)$ follow from Lemma 4.47. In order to complete the analysis for $\gamma_k^{[2]}(x)$ we recall the recurrence

$$\gamma_{k+1}^{[2]}(x) = t(x)\sum_{i\ge 1} i\gamma_k^{[2]}(x^i) + t(x)\left(\sum_{i\ge 1}\gamma_k^{[1]}(x^i)\right)^2 + t(x)\sum_{i\ge 2}(i-1)\gamma_k^{[1]}(x^i) \tag{4.102}$$

that we rewrite to

$$\gamma_{k+1}^{[2]}(x) = t(x)\gamma_k^{[2]}(x) + b_k(x),$$

where

$$b_k(x) = t(x)\sum_{i\ge 2} i\gamma_k^{[2]}(x^i) + t(x)\left(\sum_{i\ge 1}\gamma_k^{[1]}(x^i)\right)^2 + t(x)\sum_{i\ge 2}(i-1)\gamma_k^{[1]}(x^i).$$

Since $\gamma_0^{[2]}(x) = 0$, the solution of this recurrence can be written as

$$\gamma_k^{[2]}(x) = b_{k-1}(x) + t(x)b_{k-2}(x) + \cdots + t(x)^{k-1}b_0(x).$$

Hence $b_k(x) = O(1)$ uniformly for $x \in \Delta$ and consequently

$$\gamma_k^{[2]}(x) = O\left(\sum_{j=0}^{k-1}|t(x)|^j\right) = O\left(\frac{1}{1-|t(x)|}\right).$$

This completes the proof of (4.100).

The recurrence for $\gamma_{h,k}^{[2]}(x)$ is similar to that of $\gamma_k^{[2]}(x)$:

$$\gamma_{k+1,h}^{[2]}(x) = t(x)\sum_{i\ge 1} i\gamma_{k,h}^{[2]}(x^i) + t(x)\left(\sum_{i\ge 1}\gamma_{k,h}^{[1]}(x^i)\right)^2 + t(x)\sum_{i\ge 2}(i-1)\gamma_{k,h}^{[1]}(x^i) \tag{4.103}$$

with initial value $\gamma_{0,h}^{[2]}(x) = \gamma_h^{[2]}(x)$. We again use induction. Assume that we already know that $|\gamma_{k,h}^{[2]}(x)| \le D_{k,h}|x/\rho|^k$ for $|x| \le \rho - \eta$ and for some constant $D_{k,h}$. By (4.101) we can set $D_{0,h} = D_h$ which is bounded as $h \to \infty$. We also assume that $|\gamma_{k,h}^{[1]}(x)| \le C|x/\rho|^k$ for $|x| \le \rho - \eta$. Then by (4.103) we get

$$\begin{aligned}|\gamma_{k+1,h}^{[2]}(x)| &\le D_{k,h}|x/\rho|^{k+1} + D_{k,h}|x/\rho|\frac{2|x|^{2k}/\rho^k}{(1-|x|^k)^2}\\ &\quad + C^2|x/\rho|\left(\frac{|x/\rho|^k}{1-|x|^k}\right)^2 + C|x/\rho|\frac{2|x|^{2k}/\rho^k}{(1-|x|^k)^2}.\end{aligned}$$

Thus, we can set

$$D_{k+1,h} = D_{k,h}\left(1+\frac{2(\rho-\eta)^k}{(1-\rho^k)^2}\right) + C^2\frac{(\rho-\eta)^k}{(1-\rho^k)^2} + C\frac{2(\rho-\eta)^k}{(1-\rho^k)^2}$$

which shows that the constants $D_{k,h}$ are uniformly bounded. Consequently $\gamma^{[1]}_{k,h}(x) = O\left(|x/\rho|^k\right)$ for $|x| \le \rho-\eta$.

Next we assume that $|\gamma^{[2]}_{k,h}(x)| \le \bar{D}_{k,h}$ for $x \in \Delta$. We already know that $|\gamma^{[1]}_{k,h}(x)| \le C\frac{h}{h+k}$ for $x \in \Delta$. Hence,

$$\begin{aligned}|\gamma^{[2]}_{k+1,h}(x)| &\le \bar{D}_{k,h} + D_{k,h}\sum_{i\ge 2} i|x^i/\rho|^k \\ &\quad + C^2\left(\frac{h}{k+h} + \sum_{i\ge 2}|x^i/\rho|^k\right)^2 + C\sum_{i\ge 2}(i-1)|x^i/\rho|^k \\ &\le \bar{D}_{k,h} + 8D_{k,h}(\rho+\eta)^{2k}/\rho^k \\ &\quad + C^2\left(\frac{h}{k+h} + 2(\rho+\eta)^{2k}/\rho^k\right)^2 + 4C(\rho+\eta)^{2k}/\rho^k.\end{aligned}$$

Thus, we can set

$$\begin{aligned}\bar{D}_{k+1,h} &= \bar{D}_{k,h} + 8D_{k,h}(\rho+\eta)^{2k}/\rho^k + C^2\left(\frac{h}{k+h} + 2(\rho+\eta)^{2k}/\rho^k\right)^2 \\ &\quad + 4C(\rho+\eta)^{2k}/\rho^k\end{aligned}$$

with initial value $\bar{D}_{0,h} = \bar{D}_h = O(h)$ and obtain a uniform upper bound of the form

$$\bar{D}_{k,h} = O(h).$$

Consequently $\gamma^{[2]}_{k,h}(x) = O(h)$ for $x \in \Delta$.

Thus, in order to complete the proof of (4.101) it remains to prove $\gamma^{[2]}_{k,h}(x) = O(1/(1-|t(x)|))$ for $x \in \Delta$. Similarly to the above we represent $\gamma^{[2]}_{k,h}(x)$ as

$$\gamma^{[2]}_{k,h}(x) = \gamma^{[2]}_{0,h}(x) + c_{k-1,h}(x) + t(x)c_{k-2,h}(x) + \cdots + t(x)^{k-1}c_{0,h}(x), \quad (4.104)$$

where

$$c_{j,h}(x) = t(x)\sum_{i\ge 2} i\gamma^{[2]}_{j,h}(x^i) + t(x)\left(\sum_{i\ge 1}\gamma^{[1]}_{j,h}(x^i)\right)^2 + t(x)\sum_{i\ge 2}(i-1)\gamma^{[1]}_{j,h}(x^i).$$

Observe that there exists $\eta > 0$ such that $|x^i| \le \rho-\eta$ for $i \ge 2$ and $x \in \Delta$. Hence it follows that $c_{j,h}(x) = O(1)$ for $x \in \Delta$. Since $\gamma^{[2]}_{0,h}(x) = \gamma^{[2]}_h(x) = O(1/(1-|t(x)|))$, we consequently get

$$\gamma^{[2]}_{k,h}(x) = \gamma^{[2]}_h(x) + O\left(\frac{1}{1-|t(x)|}\right) = O\left(\frac{1}{1-|t(x)|}\right).$$

Remark 4.55 *Note that the estimates of Lemma 4.54 already prove that*

$$\mathbb{E}\left(L_n(k) - L_n(k+h)\right)^2 = O\left(h\sqrt{n}\right).$$

Unfortunately, this estimate is not sufficient to prove tightness. We have to deal with the 4-th moments.

Before we start with bounds for $\gamma_k^{[3]}(x)$ and $\gamma_k^{[4]}(x)$ we need an auxiliary bound.

Lemma 4.56. *We have uniformly for $x \in \Delta$*

$$\sum_{k\ge 0} |\gamma_{k,h}^{[1]}(x)\gamma_{k,h}^{[2]}(x)| = O\left(h^2\right). \tag{4.105}$$

Proof. We use the representation (4.104), where we can approximate $c_{j,h}(x)$ by

$$c_{j,h}(x) = t(x)\gamma_{j,h}^{[1]}(x)^2 + O\left(L^j\right) = O\left(\frac{h^2}{(k+h)^2}\right)$$

uniformly for $x \in \Delta$ with some constant L that satisfies $0 < L < 1$. Furthermore, we use the approximation

$$\gamma_{k,h}^{[1]}(x) = C(x)t(x)^k(1 - t(x)^h) + O\left(L^k\right)$$

that is uniform for $x \in \Delta$. This shows

$$\sum_{k\ge 0} |\gamma_{k,h}^{[1]}(x)| = |C(x)|\frac{|1 - t(x)^h|}{1 - |t(x)|} + O\left(1\right).$$

Now observe that for $x \in \Delta$ there exists a constant $c > 0$ with $|1 - t(x)| \ge c(1 - |t(x)|)$. Hence it follows that

$$\frac{|1 - t(x)^h|}{1 - |t(x)|} = O\left(\frac{1 - t(x)^h}{1 - t(x)}\right) = O\left(h\right),$$

and consequently

$$\sum_{k\ge 0} |\gamma_{k,h}^{[1]}(x)| = O\left(h\right).$$

Similarly we get

$$\begin{aligned}
\sum_{k\ge 1}|\gamma_{k,h}^{[1]}(x)| \left|\sum_{j<k} t(x)^{k-j-1}c_{j,h}(x)\right| &\le \sum_{j\ge 0}|c_{j,h}(x)|\,|t(x)|^{-j-1}\sum_{k>j}|t(x)|^k|\gamma_{k,h}^{[1]}(x)| \\
&= \sum_{j\ge 0}|c_{j,h}(x)|\,|t(x)|^{-j-1}\left(|C(x)|\,|t(x)|^{2j+2}\frac{|1 - t(x)^h|}{1 - |t(x)|^2} + O\left(|t(x)|^{j+1}L^j\right)\right) \\
&= O\left(\sum_{j\ge 0}\frac{h^3}{(j+h)^2}\right) \\
&= O\left(h^2\right).
\end{aligned}$$

Hence, we finally obtain

$$\begin{aligned}\sum_{k\ge 0} |\gamma_{k,h}^{[1]}(x)\gamma_{k,h}^{[2]}(x)| &\le \sum_{k\ge 0} |\gamma_{k,h}^{[1]}(x)|\,|\gamma_{k,h}^{[2]}(x)| + \sum_{k\ge 1} |\gamma_{k,h}^{[1]}(x)| \left|\sum_{j<k} t(x)^{k-j-1} c_{j,h}(x)\right| \\ &= O\left(h^2\right).\end{aligned}$$

Lemma 4.57. *We have*

$$\gamma_k^{[3]}(x) = \begin{cases} O\left(\min\{k^2, \frac{k}{1-|t(x)|}\}\right) & \textit{uniformly for } x \in \Delta, \\ O\left(|x/\rho|^k\right) & \textit{uniformly for } |x| \le \rho - \eta, \end{cases} \tag{4.106}$$

and

$$\gamma_{k,h}^{[3]}(x) = \begin{cases} O\left(\min\{h^2, \frac{h}{1-|t(x)|}\}\right) & \textit{uniformly for } x \in \Delta, \\ O\left(|x/\rho|^k\right) & \textit{uniformly for } |x| \le \rho - \eta, \end{cases} \tag{4.107}$$

for every $\eta > 0$.

Proof. The recurrence for $\gamma_k^{[3]}(x)$ is given by

$$\begin{aligned}\gamma_{k+1}^{[3]}(x) &= t(x)\sum_{i\ge 1} i^3\gamma_k^{[3]}(x^i) + t(x)\left(\sum_{i\ge i}\gamma_k^{[1]}(x^i)\right)^3 \\ &\quad + 3t(x)\left(\sum_{i\ge 1}\gamma_k^{[1]}(x^i)\right)\left(\sum_{i\ge 1} i\gamma_k^{[2]}(x^i)\right) \\ &\quad + 3t(x)\left(\sum_{i\ge 1}\gamma_k^{[1]}(x^i)\right)\left(\sum_{i\ge 1}(i-1)\gamma_k^{[i]}(x^i)\right) \\ &\quad + 3t(x)\sum_{i\ge 1} i(i-1)\gamma_k^{[2]}(x^i) + t(x)\sum_{i\ge 1}(i-1)(i-2)\gamma_k^{[1]}(x^i).\end{aligned} \tag{4.108}$$

By inspecting the proof of Lemmas 4.53 and 4.54 one expects that the only *important* part of this recurrence is given by

$$\gamma_{k+1}^{[3]}(x) = t(x)\gamma_k^{[3]}(x) + t(x)\gamma_k^{[1]}(x)^3 + 3t(x)\gamma_k^{[1]}(x)\gamma_k^{[2]}(x) + R_k \tag{4.109}$$

and R_k collects the *less important* remainder terms that only contribute exponentially small terms. Thus, in order to shorten our presentation we will not focus on R_k. In particular, it is easy to show the bound $\gamma_k^{[3]}(x) = O\left(|x/\rho|^k\right)$ for $|x| \le \rho - \eta$. (We omit the details.)

Next, since $t(x)\gamma_k^{[1]}(x)^3 + 3t(x)\gamma_k^{[1]}(x)\gamma_k^{[2]}(x) + R_k = O(k)$, it follows that $\gamma_k^{[3]}(x) = O\left(k^2\right)$.

Now we proceed by induction and observe that a bound of the form $|\gamma_k^{[3]}(x)| \le E_k/(1-|t(x)|)$ leads to

$$|\gamma_{k+1}^{[3]}(x)| \le \frac{E_k}{1-|t(x)|} + O\left(\frac{1}{1-|t(x)|}\right) + |R_k|$$

and consequently to $E_{k+1} \le E_k + O(1)$. Hence, $E_k = O(k)$ and $\gamma_k^{[3]}(x) = O(k/(1-|t(x)|))$.

Similarly, the leading part of the recurrence for $\gamma_{k,h}^{[3]}(x)$ is given by

$$\begin{aligned}\gamma_{k+1,h}^{[3]}(x) &= t(x)\gamma_{k,h}^{[3]}(x) + t(x)\gamma_{k,h}^{[1]}(x)^3 + 3t(x)\gamma_{k,h}^{[1]}(x)\gamma_{k,h}^{[2]}(x) + \bar{R}_{k,h} \qquad (4.110)\\ &= t(x)\gamma_{k,h}^{[3]}(x) + d_{k,h}(x),\end{aligned}$$

where

$$d_{k,h}(x) = t(x)\gamma_{k,h}^{[1]}(x)^3 + 3t(x)\gamma_{k,h}^{[1]}(x)\gamma_{k,h}^{[2]}(x) + \bar{R}_{k,h} = O(h),$$

and the initial value is given by

$$\gamma_{0,h}^{[3]}(x) = -\gamma_h^{[3]}(x) - 3\gamma_h^{[2]}(x) = O\left(\min\left\{h^2, \frac{h}{1-|t(x)|}\right\}\right).$$

Note that we also assume that $\gamma_{k,h}^{[3]}(x) = O\left(|x/\rho|^k\right)$ for $|x| \le \rho - \eta$ (which can be proved easily). Consequently it follows that

$$\begin{aligned}\gamma_{k,h}^{[3]}(x) &= \gamma_{0,h}^{[3]}(x) + d_{k-1,h}(x) + t(x)d_{k-1,h}(x) + \cdots + t(x)^{k-1}d_{0,h}(x)\\ &= O\left(\frac{h}{1-|t(x)|}\right).\end{aligned}$$

Next observe that Lemmas 4.53–4.56 ensure that

$$\sum_{j\ge 0} |d_{j,h}(x)| = O\left(h^2\right)$$

uniformly for $x \in \Delta$. Hence, we finally get

$$\gamma_{k,h}^{[3]}(x) = O\left(h^2\right),$$

which completes the proof of Lemma 4.57.

Lemma 4.58. *We have*

$$\gamma_k^{[4]}(x) = \begin{cases} O\left(\frac{k^2}{1-|t(x)|}\right) & \text{uniformly for } x \in \Delta,\\ O\left(|x/\rho|^k\right) & \text{uniformly for } |x| \le \rho - \eta,\end{cases} \qquad (4.111)$$

and

$$\gamma_{k,h}^{[4]}(x) = \begin{cases} O\left(\frac{h^2}{1-|t(x)|}\right) & \text{uniformly for } x \in \Delta,\\ O\left(|x/\rho|^k\right) & \text{uniformly for } |x| \le \rho - \eta,\end{cases} \qquad (4.112)$$

for every $\eta > 0$.

Proof. The proof is very similar to that of Lemma 4.57. First, the recurrence for $\gamma_k^{[4]}(x)$ is essentially of the form

$$\gamma_{k+1}^{[4]}(x) = t(x)\gamma_k^{[4]}(x) + t(x)\gamma_k^{[1]}(x)^4 + 4t(x)\gamma_k^{[1]}(x)\gamma_k^{[3]}(x) \\ + 6t(x)\gamma_k^{[1]}(x)^2\gamma_k^{[2]}(x) + 3t(x)\gamma_k^{[2]}(x)^2 + R_k, \tag{4.113}$$

where R_k collects all exponentially small terms. We assume that we have already proved the upper bound $\gamma_k^{[4]}(x) = O\left(|x/\rho|^k\right)$ for $|x| \le \rho - \eta$. Now, by assuming that $|\gamma_k^{[4]}(x)| \le F_k/(1-|t(x)|)$ and by using the known estimates $\gamma_k^{[1]}(x) = O(1)$, $\gamma_k^{[2]}(x) = O(\min\{k, 1/(1-|t(x)|)\})$, and $\gamma_k^{[3]}(x) = O(k/(1-|t(x)|))$, we get

$$|\gamma_{k+1}^{[4]}(x)| \le \frac{F_k}{1-|t(x)|} + O\left(|t(x)|^k\right) + O\left(\frac{k}{1-|t(x)|}\right) + O\left(\frac{1}{1-|t(x)|}\right) + |R_k| \\ \le \frac{F_{k+1}}{1-|t(x)|}$$

with $F_{k+1} = F_k + O(k)$. Consequently $F_k = O\left(k^2\right)$.

Finally, the essential part of the recurrence for $\gamma_{k,h}^{[4]}(x)$ is given by

$$\gamma_{k+1,h}^{[4]}(x) = t(x)\gamma_{k,h}^{[4]}(x) + t(x)\gamma_{k,h}^{[1]}(x)^4 + 4t(x)\gamma_{k,h}^{[1]}(x)\gamma_{k,h}^{[3]}(x) \\ + 6t(x)\gamma_{k,h}^{[1]}(x)^2\gamma_{k,h}^{[2]}(x) + 3t(x)\gamma_{k,h}^{[2]}(x)^2 + \bar{R}_{k,h} \\ = t(x)\gamma_{k,h}^{[4]}(x) + e_{k,h}(x), \tag{4.114}$$

where

$$e_{k,h}(x) = t(x)\gamma_{k,h}^{[1]}(x)^4 + 4t(x)\gamma_{k,h}^{[1]}(x)\gamma_{k,h}^{[3]}(x) \\ + 6t(x)\gamma_{k,h}^{[1]}(x)^2\gamma_{k,h}^{[2]}(x) + 3t(x)\gamma_{k,h}^{[2]}(x)^2 + \bar{R}_{k,h}.$$

As above, $\bar{R}_{k,h}$ collects all exponentially small terms. Thus,

$$\gamma_{k,h}^{[4]}(x) = \gamma_{0,h}^{[4]}(x) + e_{k-1,h}(x) + t(x)e_{k-1,h}(x) + \cdots + t(x)^{k-1}e_{0,h}(x).$$

By applying the known estimates $\gamma_{k,h}^{[1]}(x) = O(1)$, $\gamma_{k,h}^{[2]}(x) = O(h)$ and $\gamma_{k,h}^{[3]}(x) = O\left(h^2\right)$ we obtain $e_{k,h} = O\left(h^2\right)$. By combining that with the initial condition

$$\gamma_{0,h}^{[4]}(x) = 12\gamma_h^{[2]}(x) + 8\gamma_h^{[3]}(x) + \gamma_h^{[4]}(x) = O\left(\frac{h^2}{1-|t(x)|}\right),$$

we finally get

$$\gamma_{k,h}^{[4]}(x) = O\left(\frac{h^2}{1-|t(x)|}\right)$$

which completes the proof of Lemma 4.58.

The proof of (4.96) is now immediate. As already noted this implies (4.95) and proves Theorem 4.52.

4.3.5 The Height of Pólya Trees

Since the profile of Pólya trees behaves similarly to the profile of conditioned Galton-Watson trees, there is no doubt that there is a similar correspondence for the height (compare with [65]).

Theorem 4.59. *Let H_n denote the height of an unlabelled rooted random tree with n vertices. Then we have*

$$\frac{1}{\sqrt{n}} H_n \xrightarrow{d} \frac{2\sqrt{2}}{b\sqrt{\rho}} \max_{0\le t\le 1} e(t)$$

and

$$\mathbb{E}\, H_n^r \sim \left(\frac{b\sqrt{\rho}}{4}\right)^r r(r-1)\Gamma(r/2)\zeta(r)\, n^{r/2}$$

for every integer $r \ge 1$.

A similar theorem was proved by Broutin and Flajolet [28] for binary unlabelled trees who also provide a local version, that is, asymptotic expansion for $\mathbb{P}\{H_n = k\}$ (compare with [80]). For the sake of brevity we will not make this explicit but a local version of Theorem 4.59 still holds.

Let $t_{n,k}$ denote the number of trees with n nodes and height at most k. Then the generating function $t_k(x) = \sum_{n\ge 1} t_{n,k} x^n$ satisfies the recurrence relation

$$\begin{aligned} t_0(x) &= 0 \\ t_{k+1}(x) &= x \exp\left(\sum_{i\ge 1} \frac{t_k(x^i)}{i} \right), \qquad (k \ge 0). \end{aligned}$$

Obviously $t_k(x) = t_k(x,0)$ where the function on the right-hand-side is the generating function of (4.28) which we used to analyse the profile in the previous sections.

Set

$$e_k(x) = t(x) - t_k(x),$$

that is $e_k(x) = -w_k(x,0)$. Then e_k satisfies the recurrence

$$e_{k+1}(x) = t(x)\left(1 - e^{-e_k(x) - E_k(x)}\right), \tag{4.115}$$

where

$$E_k(x) = \sum_{i\ge 1} \frac{e_k(x^i)}{i} = -\Sigma_k(x,0).$$

The proof follows the same principles as the proof of the corresponding properties of the height of Galton-Watson trees (see Section 4.2.7). However, the term $E_k(x)$ needs some additional considerations. In particular we will always have to verify that in the range of interest, that is $x \in \Delta$ and $|x-\rho| < \epsilon$, we have an estimate of the kind

$$\frac{|E_k(x)|}{|e_k(x)|^2} = O(L^k) \tag{4.116}$$

for some $L < 1$. Suppose for a moment that (4.116) is satisfied and that we also know that $e_k = O(1)$. Then (4.115) rewrites to

$$e_{k+1} = te_k \left(1 - \frac{e_k}{2} + O\left(e_k^2 + \frac{E_k}{e_k}\right)\right),$$

resp. to

$$\frac{t}{e_{k+1}} = \frac{1}{e_k} + \frac{1}{2} + O\left(e_k + \frac{E_k}{e_k^2}\right).$$

This leads to the representation

$$\frac{t^k}{e_k} = \frac{1}{e_0} + \frac{1}{2}\frac{1-t^k}{1-t} + O\left(\sum_{\ell<k} |e_\ell t^\ell|\right) + O\left(\sum_{\ell<k} \frac{|E_\ell|}{|e_\ell^2|}|t^\ell|\right). \tag{4.117}$$

Now, if (4.116) is satisfied then it follows that

$$e_k = \frac{t^k}{\frac{1}{2}\frac{1-t^k}{1-t} + +O\left(\sum_{\ell<k} |e_\ell t^\ell|\right) + O(1)}$$

which can be handled as in Section 4.2.7 to obtain

$$e_k(x) = \frac{t(x)^k}{\frac{1}{2}\frac{1-t(x)^k}{1-t(x)} + O\left(\min\left\{\log k, \log\frac{1}{1-|t(x)|}\right\}\right)}. \tag{4.118}$$

If $e_k(x)$ has this kind of asymptotic representation Theorem 4.59 follows. We just have to repeat the corresponding steps from Section 4.2.7.

Note that (4.117) and (4.118) can be made more precise. Set

$$S_k = \frac{e_k^2\left(e^{-E_k} - 1\right)}{\left(e^{e_k} - 1\right)\left(1 - e^{-e_k - E_k}\right)},$$

and define a function $h(v)$ by

$$\frac{v}{1-e^{-v}} = 1 + \frac{v}{2} + v^2 h(v).$$

Then the recurrence $e_{k+1} = t(1 - e^{-e_k - E_k})$ rewrites to

$$\frac{t}{e_{k+1}} = \frac{1}{e_k} + \frac{1}{2} + e_k h(e_k) + \frac{S_k}{e_k^2}$$

and leads to the explicit representations

$$\frac{t^k}{e_k} = \frac{1}{e_0} + \frac{1}{2}\frac{1-t^k}{1-t} + \sum_{\ell<k} e_\ell h(e_\ell) t^\ell + \sum_{\ell<k} \frac{S_\ell}{e_\ell^2} t^\ell \tag{4.119}$$

and

$$e_k = \frac{t^k}{\frac{1}{e_0} + \frac{1}{2}\frac{1-t^k}{1-t} + \sum_{\ell<k} e_\ell h(e_\ell) t^\ell + \sum_{\ell<k} \frac{S_\ell}{e_\ell^2} t^\ell}. \tag{4.120}$$

Note also that if we just assume $e_k \to 0$ and $E_k = o(e_k)$ as $k \to \infty$ then

$$S_k \sim -E_k.$$

We start our precise analysis with an a priori bound for $e_k(x)$.

Lemma 4.60. *Let $|x| \le \rho$. Then there is a $C > 0$ such that*

$$|e_k(x)| \le \frac{C}{\sqrt{k}} \left|\frac{x}{\rho}\right|^k.$$

Proof. Obviously, we have

$$|e_k(x)| = \sum_{n>k} (t_n - t_{kn})|x|^n \le \sum_{n>k} t_n |x|^n.$$

The assertion follows now from $t_n \sim c\rho^{-n} n^{-3/2}$ for some constant $c > 0$.

Lemma 4.60 applies to $E_k(x)$.

Corollary 4.61 *Suppose that $|x| < \sqrt{\rho}$. Then there exists a constant $C_0 > 0$ with*

$$|E_k(x)| \le \frac{C_0}{\sqrt{k}} \left|\frac{x^2}{\rho}\right|^k.$$

The next lemma shows that $e_k(x)$ behaves as expected if x is on the positive real axis.

Lemma 4.62. *Suppose that $0 \le x \le \rho$ is real. Then (4.118) holds.*

Proof. Let $\tilde{e}_k(x)$ be defined by $\tilde{e}_0(x) = t(x)$ and by $\tilde{e}_{k+1}(x) = t(x)(1 - e^{-\tilde{e}_k(x)})$ (for $k \ge 0$). Then it follows by the methods of Section 4.2.7 that $\tilde{e}_k(x)$ behaves like (4.118), even in a proper Δ-domain.

However, if $0 \le x \le \rho$ then we obtain by induction that $e_k(x) \ge \tilde{e}_k(x)$. Hence, by combining (4.118) with the upper bound from Lemma 4.60 we have

$$\frac{E_k(x)}{e_k(x)^2} \le \frac{E_k(x)}{\tilde{e}_k(x)^2} = O(L^k)$$

for some L with $0 < L < 1$. Thus, (4.116) is satisfied and we are done.

The analysis of $e_k(x)$ for complex x with $|x| \le \rho$ is also not too difficult. The next two lemmas consider the case $|x| \le \rho$ and $|x-\rho| \le \epsilon$ and the case $|x| \le \rho - \epsilon$.

Lemma 4.63. *There exists $\epsilon > 0$ such that (4.118) holds for all x with $|x| \le \rho$ and $|x-\rho| \le \epsilon$.*

Proof. Recall that $|e_k(x)| \le C/\sqrt{k}$. Suppose that we can show that $|E_k/e_k^2| \le 1$. Then it follows that

$$\begin{aligned} |e_{k+1}| &\ge |t|\,|e_k| \exp\left(-C_1|e_k|\left(1+\frac{|E_k|}{|e_k|^2}\right)\right) \\ &\ge |t|\,|e_k| e^{-C_2 k^{-1/2}}. \end{aligned} \tag{4.121}$$

with $C_2 = 2C_1C$.

We now choose k_0 sufficiently large that

$$e^{-2C_2\sqrt{k}} \le \frac{1}{k} \quad \text{and} \quad C_2\rho^{k/2}e^{4C_0\sqrt{k}} \le 1$$

hold for all $k \ge k_0$. We already know that $e_k(\rho) \sim 2/k$. Hence, by continuity there exists $\epsilon > 0$ with $|e_{k_0}(x)| \ge \frac{1}{k_0}$ and $|t(x)| \ge \rho^{1/4}$ for $|x| \le \rho$ and $|x-\rho| \le \epsilon$. These assumptions imply

$$|e_{k_0}(x)| \ge \frac{1}{k_0} \ge e^{-2C_2\sqrt{k_0}} \ge |t(x)|^{k_0} e^{-2C_2\sqrt{k_0}}$$

and (by Corollary 4.61)

$$\frac{|E_{k_0}|}{|e_{k_0}^2|} \le C_0\rho^k|t|^{-2k}e^{4C_2\sqrt{k}} \le C_0\rho^{k/2}e^{4C_2\sqrt{k}} \le 1$$

The goal is to show by induction that for $k \ge k_0$ and for $|x| \le \rho$ and $|x-\rho| \le \epsilon$

$$|e_k| \ge |t|^k e^{-2C_2\sqrt{k}} \quad \text{and} \quad \left|\frac{E_k}{e_k^2}\right| \le 1. \tag{4.122}$$

By assumption (4.122) is satisfied for $k = k_0$. Now suppose that (4.122) holds for some $k \ge k_0$. Then (4.121) implies

$$\begin{aligned} |e_{k+1}| &\ge |t|\,|e_k| e^{-C_2 k^{-1/2}} \\ &\ge |t|^{k+1} e^{-2C_2\sqrt{k}} e^{-C_2 k^{-1/2}} \\ &\ge |t|^{k+1} e^{-2C_2\sqrt{k+1}}. \end{aligned}$$

Furthermore

$$\frac{|E_{k+1}|}{|e_{k+1}^2|} \le C_0\rho^k|t|^{-2k-2}e^{4C_2\sqrt{k+1}} \le C_0\rho^{(k+1)/2}e^{4C_2\sqrt{k+1}} \le 1.$$

Hence, we have proved (4.122) for all $k \ge k_0$.

In the last step of the induction proof we also obtained the upper bound

$$\frac{|E_k|}{|e_k^2|} \le C_0 \rho^{k/2} e^{4C_2\sqrt{k}}$$

which is sufficient to obtain the asymptotic representation (4.118).

Lemma 4.64. *Suppose that $|x| \le \rho - \epsilon$ for some $\epsilon > 0$. Then we have uniformly*

$$e_k(x) = C_k(x) t(x)^k = (C(x) + o(1)) t(x)^k$$

for some analytic function $C(x)$. Consequently we have uniformly for $|x| \le \sqrt{\rho - \epsilon}$

$$E_k(x) = \tilde{C}_k(x) t(x^2)^k = (\tilde{C}(x) + o(1)) t(x^2)^k$$

with an analytic function $\tilde{C}(x)$.

Proof. If $|x| \le \rho - \epsilon$ then we have $|e_k(x)| \le e_k(\rho - \epsilon) = O(t(\rho - \epsilon)^k)$. Thus, we can replace the upper bound $|e_k(x)| \le C/\sqrt{k}$ in the proof of Lemma 4.63 by an exponential bound which leads to a lower bound for $e_k(x)$ of the form

$$|e_k(x)| \ge c_0 |t(x)|^k.$$

Hence, by using (4.120) the result follows with straightforward calculations.

The disadvantage of the previous two lemmas is that they only work for $|x| \le \rho$. In order to obtain some progress for $|x| > \rho$ we have to find a proper analogue to Lemma 4.32. We fix a constant $C > 0$ such that

$$\left| e^{-E_k(x)} - 1 \right| \le \frac{C}{\sqrt{k}} \left(\frac{|x|^2}{\rho} \right)^k$$

for all $k \ge 1$ and for all $|x| \le \sqrt{\rho}$. Furthermore we choose a proper Δ-domain $\Delta(\rho, \eta, \delta)$ with $\rho + \eta < \sqrt{\rho}$ such that $|t(x)| < 1$ for all $x \in \Delta$.

Lemma 4.65. *Suppose that $x \in \Delta$ such that $|t(x)| > |x^2/\rho|$ and that there exist real numbers D_1 and D_2 with $0 < D_1, D_2 < 1$ and some integer $K \ge 1$ with*

$$|e_K(x)| < D_1, \quad |t(x)| \frac{e^{D_1} - 1}{D_1} < D_2, \quad D_1 D_2 + e^{D_1} \frac{C}{\sqrt{K}} \left(\frac{|x|^2}{\rho} \right)^K < D_1. \tag{4.123}$$

Then we have $|e_k(x)| < D_1$ for all $k \ge K$ and

$$e_k(x) = O(t(x)^k)$$

as $k \to \infty$, where the implicit constant might depend on x.

Proof. By definition we have $e_{K+1} = t(1 - e^{-e_K - E_K})$. Hence, if we write $e^{-E_k} = 1 + R_k$ we obtain

$$|e_{K+1}| \le |e_K|\,|t|\frac{e^{|e_K|} - 1}{|e_K|} + e^{|e_K|} R_K.$$

If (4.123) is satisfied then it follows that

$$|e_{K+1}| \le D_1 D_2 + e^{D_1} \frac{C}{\sqrt{K}} \left(\frac{|x|^2}{\rho}\right)^K < D_1.$$

Now we can proceed by induction and obtain $|e_k| < D_1$ for all $k \ge K$. Note that $D_2 < 1$ and $\sum_k R_k = O(1)$ also imply that

$$e_k = O(D_2^k).$$

If we set $a_k = t(1 - e^{-e_k})/e_k$ and $b_k = -e^{-e_k} R_k$ we obtain the recurrence

$$e_{k+1} = e_k a_k + b_k$$

with an explicit solution of the form

$$e_k = e_K \prod_{K \le i < k} a_i + \sum_{K \le j < k} b_j \prod_{j < i < k} a_i.$$

Since $e_k = O(D_2^k)$, we have

$$\prod_{j<i<k} a_i = t(x)^{k-j} e^{O(1)}$$

and consequently

$$b_j \prod_{j<i<k} a_i = O\left(t(x)^{k-j} |x^2/\rho|^j\right) = O\left(t(x)^k L^j\right)$$

for some L with $0 < L < 1$. Hence,

$$e_k = e_K t(x)^{k-K} e^{O(1)} + O\left(t(x)^k L^K\right) = O\left(t(x)^k\right). \tag{4.124}$$

Recall that $e_k(x) \to 0$ for $|x| \le \rho$. Using Lemma 4.65 we deduce that $e_k \to 0$ in a certain region that extends the circle $|x| \le \rho$.

Lemma 4.66. *For every $\epsilon > 0$ there exists $\delta > 0$ such that $e_k(x) \to 0$, if $|x| \le \rho + \delta$ and $|x - \rho| \ge \epsilon$.*

Proof. The same arguments as in the proof of Lemma 4.34 show that we can apply Lemma 4.65.

For the most interesting range, namely for $x \in \Delta$ and $|x - \rho| \le \epsilon$, we need a proper variant of Lemma 4.36.

Lemma 4.67. *There exists $\epsilon > 0$ and a constant $c_1 > 0$ such that for all $x \in \Delta$ with $x \ge \rho$ and $|x - \rho| \le \epsilon$ and* $\arg(x) \ne 0$ *the conditions (4.123) are satisfied for*

$$k = K(x) = \left\lfloor \frac{c_1}{|\arg(t(x))|} \right\rfloor$$

and properly chosen real numbers D_1, D_2. Consequently, $e_k(x) \to 0$.

Proof. The arguments are about the same as in Lemma 4.36, but there are some essential differences.

Suppose that $\arg(x)$ and $\arg(t(x))$ are positive and that $K(x) = \lfloor c_1/|\arg(t(x))|\rfloor$, where $c_1 = \arccos(1/4) - \epsilon_1$ and $\epsilon_1 > 0$ is arbitrarily small.

The first step is to prove that $|e_k(x)| \le \epsilon_2$ and $\arg(e_k(x)) \le k \arg(t(x)) + \epsilon_3$ for $k_0 \le k \le K(x)$, where $\epsilon_2 > 0$ and $\epsilon_3 > 0$ can be chosen arbitrarily small and k_0 is sufficiently large. Here we use the recurrence (4.115) and proceed inductively as in the proof of Lemma 4.36. Note that $0 < \arg(e_k(x)) < \arccos(1/4)$ ensures that $|e_{k+1}| \le |e_k| + |E_k|$ and consequently, by Lemma 4.64 and the property $e_{k_0}(\rho) \sim c/k_0$ it follows that

$$|e_k(x)| \le |e_{k_0}| + \sum_{k_0 \le \ell < k} |E_\ell| < \epsilon_2$$

provided that k_0 is chosen sufficiently large. (We do not repeat the details.)

The second step is to prove a lower bound for $e_k(x)$ for $k_0 \le k \le K(x)$ of the form $|e_k(x)| \ge c|t(x)|^k/k$ for some $c > 0$. By Lemma 4.64 $E_k(x) = (\tilde{C}(x) + o(1))t(x^2)^k$ behaves nicely, if $|x - \rho| \le \epsilon$. Suppose that $|x| \ge \rho$ and $x \in \Delta$. Since $\arg(t(x^2))$ is of order $\arg(t(x))^2$ we deduce that $\arg(E_k(x)) = O(\arg(t(x)))$ for $k_0 \le k \le K(x)$. In particular, it follows that (for $k_0 \le k \le K(x)$)

$$|e_{k+1}(x)| = \left| t(x)(1 - e^{-e_k(x) - E_k(x)}) \right| \ge |t(x)|(1 - e^{-|e_k(x)|}).$$

Hence, by applying the methods of Section 4.2.7 it follows that $|e_k(x)| \ge c|t(x)|^k/k$ for some $c > 0$.

Finally we have to check that the conditions (4.123) are satisfied for $k = K(x)$. First we use the formula

$$e_k = \frac{t^{k-k_0}}{\frac{1}{e_{k_0}} + \frac{1}{2}\frac{1-t^{k-k_0}}{1-t} + \sum_{k_0 \le \ell < k} e_\ell h(e_\ell) t^\ell + \sum_{k_0 \le \ell < k} \frac{S_\ell}{e_\ell^2} t^\ell}. \tag{4.125}$$

to show (as in the proof of Lemma 4.36) that the term

$$\frac{1}{2}\frac{1 - t^{k-k_0}}{1 - t}$$

dominates the denominator on the right-hand-side of the equation. The first and the third term can be handled as in Lemma 4.36. Finally, due to the already obtained bounds $|e_k(x)| \ge c|t(x)|^k/k$, $E_k(x) = O(t(x^2)^k)$ and the property $S_\ell \sim -E_\ell$ the last term

$$\sum_{k_0 \le \ell < k} \frac{S_\ell}{e_\ell^2} t^\ell = O(1)$$

and does not contribute to the main term, either. Summing up we have

$$|e_k| \le \frac{|t|^{k-k_0}}{\left|\frac{1}{2}\frac{1-t^{k-k_0}}{1-t}\right|}(1+\epsilon_4)$$

for an arbitrarily small $\epsilon_4 > 0$.

We set $\arg(\rho - x) = \vartheta$, where we assume that $\vartheta \in \left[-\frac{\pi}{2} - \epsilon_5, \frac{\pi}{2} + \epsilon_5\right]$ (for some $\epsilon_5 > 0$ that has to be sufficiently small), and $r = b|\rho - x|^{1/2}$, where $b \approx 2.6811266$ is the constant appearing in Theorem 3.8. Then we have (as in the proof of Lemma 4.36)

$$\begin{aligned} |t| &= 1 - r\cos\frac{\vartheta}{2} + O(r^2), \\ \log|t| &= -r\cos\frac{\vartheta}{2} + O(r^2), \\ \arg(t) &= -r\sin\frac{\vartheta}{2} + O(r^2). \end{aligned}$$

Hence with $k = K(x) = \lfloor c_1/|\arg(t)|\rfloor$ we have

$$|t^{k-k_0}| \sim e^{-c_1\cot(\vartheta/2)+O(r^2)} \le e^{-c_1\cot(\frac{\pi}{4}+\frac{\epsilon_5}{2})+O(r^2)} \le e^{-c_1}(1-\epsilon_6)$$

for some arbitrarily small $\epsilon_6 > 0$ (depending on ϵ_5). Consequently

$$|e_k| < D_1 := 2\frac{e^{-\arccos(1/4)}}{1-e^{-\arccos(1/4)}} r(1+\epsilon_7) = c'r,$$

where $\epsilon_7 > 0$ can be chosen arbitrarily small. Moreover

$$|t| = 1 - r\cos\frac{\vartheta}{2} + O(r^2) \le 1 - \frac{r}{\sqrt{2}}(1-\epsilon_8)$$

for some (small) $\epsilon_8 > 0$ and consequently

$$|t|\frac{e^{D_1}-1}{D_1} = 1 - \left(\frac{1}{\sqrt{2}} - \frac{e^{-\arccos(1/4)}}{1-e^{-\arccos(1/4)}}\right) r(1-\epsilon_9) + O(r^2).$$

Thus, we are led to set

$$D_2 := 1 - \frac{1}{2}\left(\frac{1}{\sqrt{2}} - \frac{e^{-\arccos(1/4)}}{1-e^{-\arccos(1/4)}}\right) r = 1 - c''r$$

and the first two conditions of (4.123) are satisfied if r is sufficiently small. Since $D_1 - D_1D_2 = c'c''r^2$ we just have to check whether

$$e^{D_1}\frac{C}{\sqrt{k}}\left|\frac{x^2}{\rho}\right|^k < c'c''r^2.$$

However, since $k = K(x) \ge c_1\sqrt{2}\,r^{-1}$ the left hand side of this inequality is definitely smaller than $c'c''r^2$ if r is sufficiently small. Hence all conditions of (4.123) are satisfied for $k = K(x)$.

Lemma 4.68. *There exists $\epsilon > 0$ such that (4.118) holds for all x with $x \in \Delta$ with $|x| \ge \rho$ and $|x-\rho| \le \epsilon$.*

Proof. We recall that the properties $e_k = O(1)$ and (4.116) imply (4.118). By Lemma 4.67 we already know that $e_k \to 0$. Furthermore, we have upper bounds for E_k (see Lemma 4.64). Hence, it remains to provide proper lower bounds for e_k.

Since we already know that $e_k \to 0$ and $|E_k| = O(L^k)$ (for some $L < 1$), the recurrence (4.115) implies

$$|e_{k+1}| \ge (1-\delta)\left(|e_k| - |E_k|\right)$$

for some $\delta > 0$ provided that $x \in \Delta$ and $|x-\rho| < \epsilon$. Without loss of generality we can assume that $L < (1-\delta)^2$. Hence

$$|e_k| \ge (1-\delta)^k - \sum_{\ell<k}|E_\ell|(1-\delta)^{k-\ell} \ge c_0(1-\delta)^k$$

for some constant $c_0 > 0$. Consequently

$$\left|\frac{E_k}{e_k^2}\right| = O\left(\left(\frac{L}{(1-\delta)^2}\right)^k\right).$$

As noted above, this upper bound is sufficient to deduce (4.118).

As already mentioned the proof of Theorem 4.59 is now a direct copy of the corresponding parts of the proof of Theorem 4.29 (compare with Section 4.2.7).

5

The Vertical Profile of Trees

Chapter 4 was devoted to the horizontal profile of certain classes of trees and we have observed that this kind of profile can be approximated by the local time of the Brownian excursion.

However, it is also possible to introduce a vertical profile of trees. For a binary tree, for example, we can define the vertical position of a vertex by the difference between the number of steps to the right and to the left on the path from the root. Then the vertical profile counts the number of nodes of a given vertical position. Interestingly, this new kind of profile can be approximated in terms of the ISE, the Integrated Super-Brownian Excursion.

The study of the vertical profile of binary trees is not only interesting from a probabilistic point of view. The generating functions related to these problem are interesting by themselves, since they have an explicit form that is not understood from a combinatorial point of view.

Similar phenomena appear for embedded trees and so-called well balanced trees. The latter kind of trees is of specific interest, since there is a bijection to quadrangulations – due to Schaeffer [188] – that transfers the (usual) profile of a quadrangulation to the profile of the labels of well balanced trees.

After a description of the Schaeffer bijection we survey the relations between general (vertical) profiles in random trees and the ISE that turns out to be a universal limiting object. However, the main part of this chapter is devoted to some special cases of this limit relation which can be worked out explicitly and lead to integral representations for distribution of certain functionals of the ISE. These results are due to Bousquet-Mélou [23] who extended previous work of Bouttier, Di Francesco and Guitter [25] considerably. The general convergence theorem to the ISE is based on the work of Aldous [5], Janson and Marckert [117], Bousquet-Mélou and Janson [24], and Devroye and Janson [52].

5.1 Quadrangulations and Embedded Trees

We (again) consider quadrangulations Q that are embedded in the plane. We recall that a quadrangulation is a planar map, where each face has valency 4 and where one of the edges on the external face is rooted and directed (so that the external face is on the left of this edge). We will also consider quadrangulations, where we additionally mark one vertex v_0 (see Figure 5.1[1]). If we assume that Q has n faces then there are exactly $2n$ edges and $n+2$ vertices (by Euler's formula).

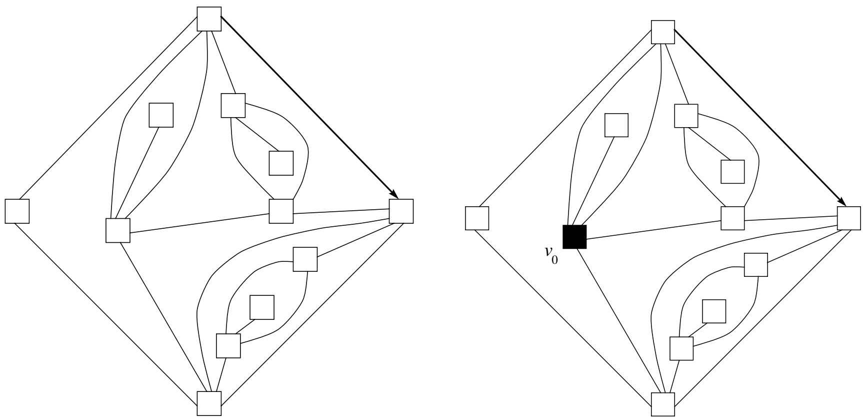


Fig. 5.1. Quadrangulations

We now define the profile of a quadrangulation to be the sequence $(H_k)_{k\ge1}$, where H_k denotes the number of vertices with distance k to the first vertex of the root edge. Similarly, if we additionally mark a vertex v_0 then the profile $(H_k^{v_0})_{k\ge1}$ is the sequence of numbers $H_k^{v_0}$ of vertices with distance k from v_0 (compare with Figure 5.2)

Our main objective in this introductory section is to present a bijection between quadrangulations and certain classes of trees that transfers the profile. For this purpose we introduce well-labelled and so-called embedded trees with increments 0 and ±1.

A *well-labelled tree* is planted plane tree, where the vertices are labelled by positive integers such that the root has label 1 and labels of adjacent vertices differ at most by 1. Similarly, an *embedded tree* with increments 0 and ±1 is a planted plane tree, where the vertices are labelled by integers such that the root has label 0 and, again, labels of adjacent vertices differ by 0 or by ±1 (see Figure 5.3).

[1] The example used here is adopted from [26]

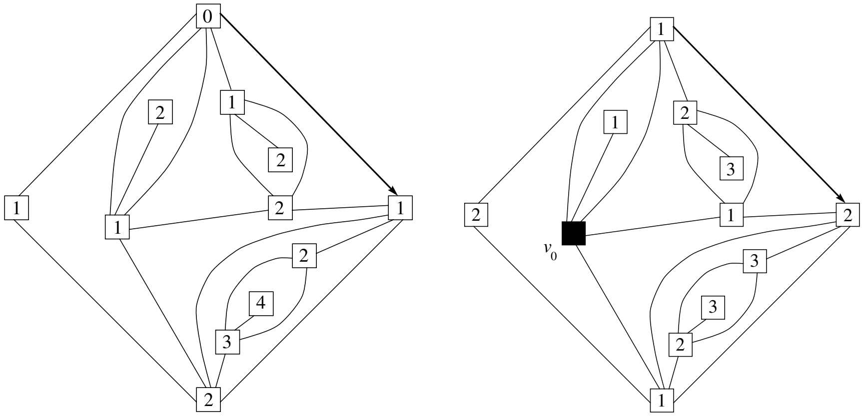


Fig. 5.2. Distances in quadrangulations

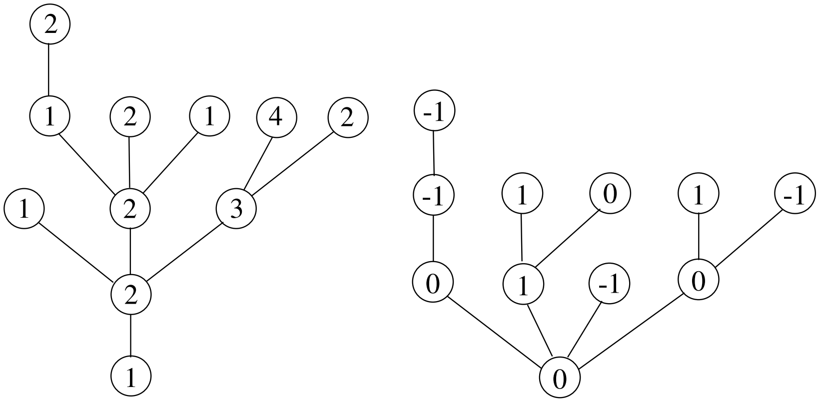


Fig. 5.3. Well-labelled and embedded trees

The label distribution $(\lambda_k)_{k\ge 1}$ of a well-labelled tree T consists of the numbers λ_k of nodes in T with label k. Similarly, the label distribution $(\Lambda_k)_{k\in\mathbb{Z}}$ of embedded trees is defined by the numbers Λ_k of nodes in T with label k.

We start with a bijection between these quadrangulations and well-labelled trees that is due to Schaeffer [188].

Theorem 5.1. *There exists a bijection between edge-rooted quadrangulations with n faces and well-labelled trees with n edges, such that the profile $(H_k)_{k\ge 1}$ of a quadrangulation is mapped onto the label distribution $(\lambda_k)_{k\ge 1}$ of the corresponding well-labelled tree.*

Instead of giving a formal proof of this statement we will work out an explicit example. The first step is to label the vertices of the quadrangulation by the distance from the root vertex (the first vertex of the root edge, see Figure 5.2). By this definition it is clear that there are only two possible configurations of labels on the boundary of a face, since the labels of adjacent vertices can only differ by 1 (see Figure 5.4). We will call a face of type 1,if there are three different labels $k, k+1, k+2$ and of type 2, if there are only two different labels $k, k+1$.

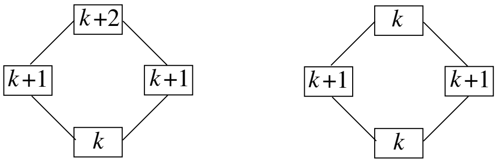


Fig. 5.4. Two types of faces

We now apply to each face (also to the external one where we have to interchange the orientation) the following rule. If a face is of type 1 we include a fat edge between $k+1$ and $k+2$ in counterclockwise direction and if the face is of type 2 then we connect the two vertices labelled by $k+1$ by a fat edge (see Figure 5.5).

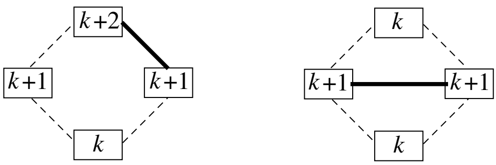


Fig. 5.5. Modifications for the two types of faces

If we apply this rule to all faces of the quadrangulation we obtain the figure that is depicted on the left-hand-side of Figure 5.6. Finally, we delete all non-fat edges; however, we keep track of the second vertex of the rooted edge (that is always labelled by 1).

The observation now is that the resulting figure that contains all vertices different from the first vertex of the root edge (of the original quadrangulation) and all remaining *fat edges* is in fact a well-labelled tree with the second vertex of the root edge as the root. In this particular example we actually get the well-labelled tree depicted in Figure 5.3. Note that we have two kinds of fat edges. The first kind are edges that are also edges of the original quadrangulation. They connect vertices with different labels. The other kind of fat edges are new edges and can be recognised easily, since they connect vertices with the same label.

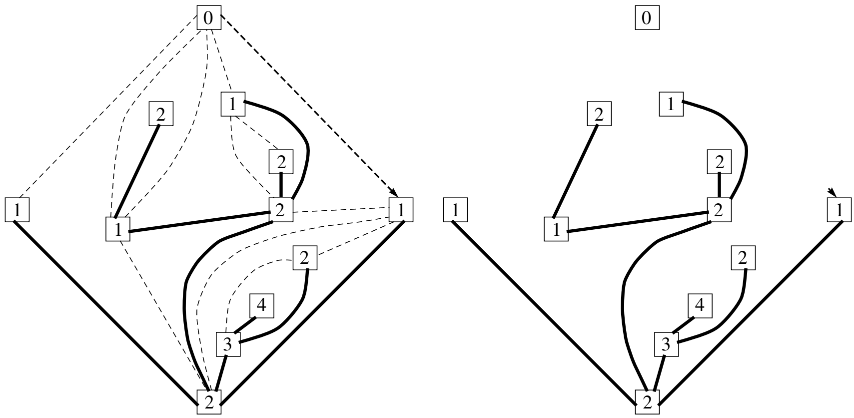


Fig. 5.6. Quadrangulations and well-labelled trees

Next we modify the above procedure to obtain a bijection to embedded trees with increments 0 and ± 1 (that is due to Chassaing and Schaeffer [35]).

Theorem 5.2. *There exists a bijection between edge-rooted quadrangulations with n faces, where one of the $n+2$ vertices is marked, and two copies*[2] *of the set of embedded trees (with increments* 0 *and* ± 1*) with n edges, such that the profile $(H_k^{v_0})_{k\ge 1}$ of a quadrangulation is mapped onto the shifted label distribution of the corresponding embedded tree with increments* 0 *and* ± 1*. (By shifted label distribution we mean the label distribution obtained by translating all labels so that the minimum label is* 0*.)*

Again we present just an example. We start with the quadrangulation depicted in Figure 5.1 and label the vertices by the distance from v_0. Then we apply the same procedure as above in order to get *fat edges* that form a tree (compare with Figure 5.7). There is, however, a slight modification that has to be made at the fat edge e that corresponds to the external face. We have to keep track that the root edge of the quadrangulation can be recovered in the inverse procedure. Since there are exactly 4 possible cases, we can encode them by choosing a direction to e and adding a $+$ or $-$ to the resulting tree (by any rule). The first vertex of the directed edge e will be now the root of an embedded tree and the second vertex of e the left-most successor of root in this tree. Finally, we also shift all labels so that the new root gets label 0. In our example we get precisely the embedded tree of Figure 5.3 with a $+$.

We can use these bijections to obtain explicit formulas for the number of quadrangulations and for the number of well-labelled trees. Let u_n denote the number of embedded trees (with increments 0 and ± 1) with n edges and q_n the number of well-labelled trees (with increments 0 and ± 1). Then the above

[2] Formally, the notion *two copies of a set A* means the union $(A \times \{1\}) \cup (A \times \{2\})$.

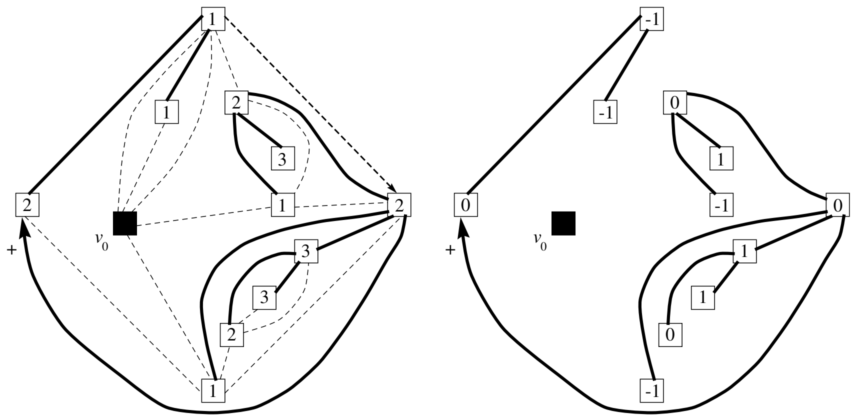


Fig. 5.7. Quadrangulations and embedded trees

two bijections show that

$$2u_n = (n+2)q_n,$$

since there are precisely $n+2$ ways to root a quadrangulation with $n+2$ vertices. On the other hand it is immediately clear that

$$u_n = 3^n p_{n+1} = \frac{3^n}{n+1}\binom{2n}{n},$$

where $p_{n+1} = \frac{1}{n+1}\binom{2n}{n}$ denotes the number of planted plane trees with n edges (or with $n+1$ vertices). Namely, there are precisely 3^n different ways to label a given planted plane tree with n edges to make it an embedded tree. Starting from the root that gets label 0 we have 3 ways to label an adjacent vertex and so on. Hence, we get the following remarkable formula.

Theorem 5.3. *The number q_n of different edge-rooted quadrangulations with n faces, or the number of well-labelled trees (with increments 0 and ± 1) with n edges, is given by the formula*

$$q_n = \frac{2\cdot 3^n}{(n+1)(n+2)}\binom{2n}{n}. \tag{5.1}$$

We also present an alternative proof for this formula by a generating function approach that counts well-labelled trees.

Proof. We consider variants of well-labelled trees, where the root has a label not necessarily equal to 1 but again all vertices are labelled by positive integers and labels of adjacent vertices differ at most by 1. Let $T_j(t)$, $j \ge 1$, denote the (ordinary) generating function of those generalised well-balanced trees where the root has label j and where the exponent of t counts the number of edges. Then, by using the convention $T_0(t) = 0$, we immediately get the relation

$$T_j(t) = \frac{1}{1 - t(T_{j-1}(t) + T_j(t) + T_{j+1}(t))}, \quad (j \ge 1). \tag{5.2}$$

This relation comes from the usual decomposition of a planted plane tree into the root and its subtrees. If the root has label j then the successors of the root must have a label in the set $\{j-1, j, j+1\}$ and the subtrees are again well-balanced trees. This justifies (5.2). Note also that the infinite system (5.2) uniquely determines the functions $T_j(t)$. In particular, the equations (5.2) provide a recurrence for the coefficients $[t^n]\,T_j(t)$ that can be solved by using the initial condition $T_0(t) = 0$. For example we have indeed $[t^0]\,T_0(t) = 0$ and $[t^0]\,T_j(t) = 1$ for $j \ge 1$. Hence, it follows that $[t^1]\,T_1(t) = 2$ and $[t^1]\,T_j(t) = 3$ for $j \ge 2$. In the same way we can proceed further. If we already know $[t^k]\,T_j(t)$ for all $k \le n$ and all $j \ge 0$ then $[t^{n+1}]\,T_j(t)$ can be computed for all $j \ge 1$. Since $[t^{n+1}]\,T_0(t) = 0$ the induction works.

Let $T(t) = (1 - \sqrt{1-12t})/(6t)$ be the solution of the equation

$$T(t) = \frac{1}{1 - 3tT(t)}, \tag{5.3}$$

and let $Z(t)$ be defined by

$$Z(t) + \frac{1}{Z(t)} + 1 = \frac{1}{tT(t)^2}. \tag{5.4}$$

Note that one of the two possible choices of $Z(t)$ represents a power series in t with constant term 0 (and non-negative coefficients) which we will use in the sequel.

By using the ansatz

$$T_j(t) = T(t)\frac{u_j u_{j+3}}{u_{j+1}u_{j+2}}$$

with unknown functions u_j, the recurrence (5.2) is equivalent to

$$\begin{aligned} u_j u_{j+1} u_{j+2} u_{j+3} &= \frac{1}{T(t)} u_{j+1}^2 u_{j+2}^2 \\ &\quad + tT(t)\left(u_{j-1}u_{j+2}^2 u_{j+3} + u_j^2 u_{j+3}^2 + u_j u_{j+1}^2 u_{j+4}\right). \end{aligned} \tag{5.5}$$

By using (5.3) and (5.4) it is easy to check that

$$u_j = 1 - Z(t)^j$$

satisfies (5.5) which is kind of a mystery. Hence, the functions

$$T_j(t) = T(t)\frac{(1 - Z(t)^j)(1 - Z(t)^{j+3})}{(1 - Z(t)^{j+1})(1 - Z(t)^{j+2})} \tag{5.6}$$

satisfy the system of equations (5.2) and we also have $T_0(t) = 0$. Since all these functions $T_j(t)$ are power series in t, we have found the solution of interest of the system (5.2).

In particular, we have

$$\begin{aligned} T_1(t) &= T(t)\frac{(1-Z(t))(1-Z(t)^4)}{(1-Z(t)^2)(1-Z(t)^3)} \\ &= T(t)\frac{1+Z(t)^2}{1+Z(t)+Z(t)^2} \\ &= T(t)(1-tT(t)^2), \end{aligned}$$

and it is an easy exercise (by using Lagrange's inversion formula) to show that

$$q_n = [t^n]T_1(t) = \frac{2\cdot 3^n}{(n+1)(n+2)}\binom{2n}{n}.$$

This completes the proof of the theorem.

Interestingly enough, there is a similar correspondence between Eulerian triangulations and well-labelled trees with increments ± 1. An Eulerian triangulation is a planar map (with a root edge), where all faces are triangles (that is, they have valency 3) and with an even number of triangles around each vertex. These triangulations are characterised by the property that their faces are bicolourable or that their vertices are tricolourable (compare with [26]). They can also be seen as the dual maps of the bicubic (that is, bipartite and trivalent) maps, which were first enumerated by Tutte [204].

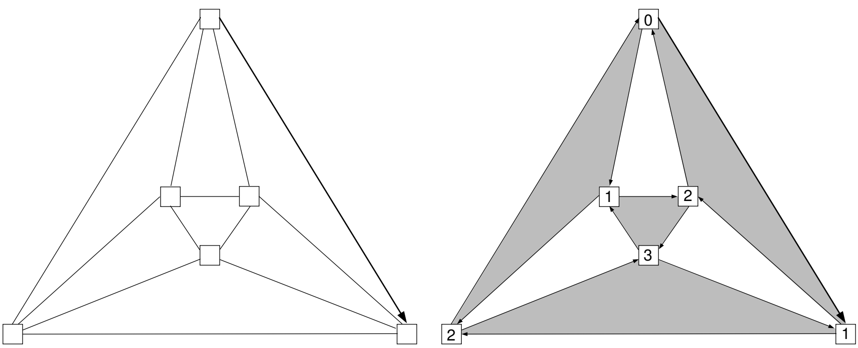


Fig. 5.8. Eulerian triangulation

We will describe the bijection by using the example depicted in Figure 5.8. First, the orientation of the root edge (that is situated at the infinite face and oriented clockwise) induces uniquely an orientation of all other edges by requiring that the orientations alternate around each vertex. We (uniquely) distinguish between black and white faces, with the convention that the infinite face is white. In particular, black faces are then oriented clockwise and white ones counterclockwise. Then we define a distance between the first root vertex and any other vertex by the minimal number of edges in the shortest oriented path between them. This defines a labelling of the vertices such that the sequence of labels around all black faces is of the form k, $k+1$, $k+2$ (see Figure 5.9). This follows from the fact that neighbouring vertices have labels which differ by at most 2 while the tricolorability of the vertices fixes the residue modulo 3. We now build a well-labelled tree by a rule which is similar to that for quadrangulations. We replace the edge in such a black triangle from $k+1$ to $k+2$ by a fat edge (see Figure 5.9). Now the vertices (different from the first root vertex of the triangulation) together with the fat edges form a well-labelled tree (rooted at the second root vertex of the triangulation) with increments ± 1 and root label 1 (see Figure 5.10).

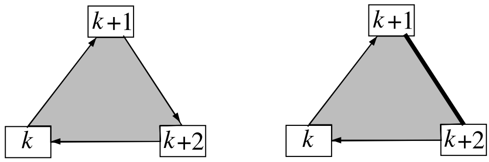


Fig. 5.9. Modification of black faces

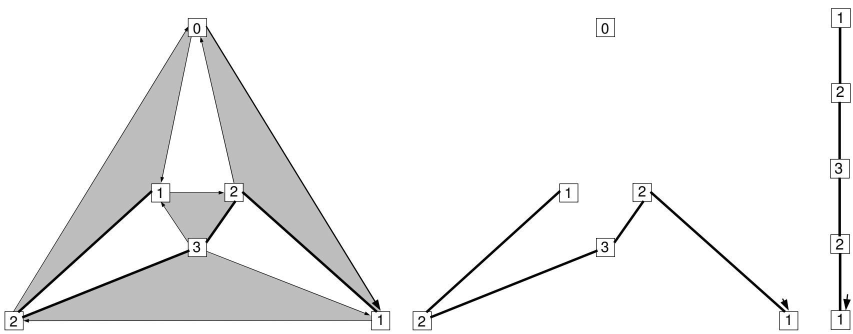


Fig. 5.10. Eulerian triangulation and well-labelled tree

It is an nice exercise to show that this procedure is actually a bijection between Eulerian triangulations and well-labelled trees with increments ± 1. This bijection and a procedure that is similar to proof of Theorem 5.3 yields an explicit formula for the number t_n of Eulerian triangulations with n black faces:

$$t_n = \frac{3 \cdot 2^{n-1}}{(n+1)(n+2)} \binom{2n}{n}.$$

We will come back to this enumeration problem in Section 5.3.1.

5.2 Profiles of Trees and Random Measures

5.2.1 General Profiles

The profile of a rooted tree T was defined as the sequence $(L_T(k))_{k\geq 0}$, where $L_T(k)$ denotes the number of nodes of T at distance k from the root. Similarly we have defined the profile $(H_k)_{k\geq 0}$ of a quadrangulation and the label distribution $(\lambda_k)_{k\geq 0}$ of a well-labelled tree.

All these examples can be seen from a more general point of view. Let T be a tree with n nodes, where the nodes v are labelled by integers $\ell(v)$. Let $X_T(k)$ denote the number of nodes v in T with $\ell(v) = k$, that is,

$$X_T(k) = \#\{v \in V(T) : \ell(v) = k\}.$$

Then the sequence $(X_T(k))_{k\in\mathbb{Z}}$ that encodes the label distribution will be called the *profile* of T corresponding to the labelling ℓ.

This comprises all previous notions. For example, if ℓ denotes the distance from the root then we get the usual profile. The vertical position of vertices in binary trees induces the *vertical profile*.

The distribution of the labels gives rise to a probability distribution on the integers[3]

$$\mu_T = \frac{1}{|V(T)|}\sum_{v\in V(T)} \delta_{\ell(v)} = \frac{1}{|V(T)|}\sum_{k\in\mathbb{Z}} X_T(k)\,\delta_k.$$

If either the tree is random or the labelling is random then the corresponding profiles induce *random measures* and it is a natural problem to describe or to approximate these random measures in a proper way. For example, (4.6) says that the occupation time of the Brownian excursion $e(t)$ is (after scaling) the limiting random measure. The corresponding local version is Theorem 4.10, the local time $l(t)$ is the corresponding random density.

5.2.2 Space Embedded Trees and ISE

In many instances labels encode topological properties like positions or – as above – the distance from the root. This idea was further developed by Aldous [5] by using *several-dimensional labels*. Fix a dimension $d \geq 1$ and embed a rooted tree into the d-dimensional space $\mathbb{R}^d$ in the following way. Put the root at the origin and regard each edge as a step of the form $(0,\ldots,0,\pm 1,0,\ldots,0)$ so that each vertex is sent to a vertex of the lattice $\mathbb{Z}^d$. Figure 5.11 shows an easy example of this procedure. Note that different points of the tree may be sent to the same point in $\mathbb{R}^d$.

By putting mass $1/|V(T)|$ on each vertex of the embedded tree we get an induced measure in $\mathbb{R}^d$. Formally we can do this by using a d-dimensional

[3] δ_x denotes the δ-distribution concentrated at x.

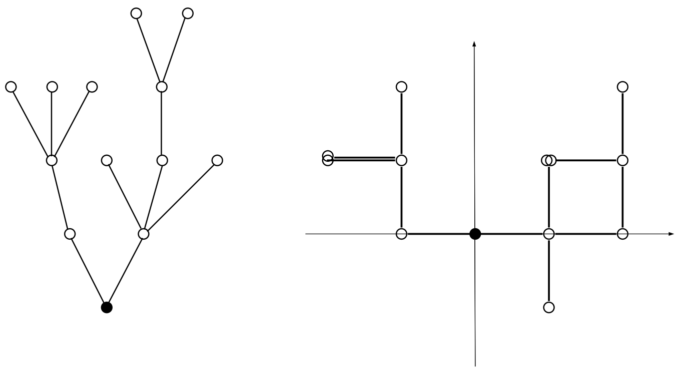


Fig. 5.11. Space embedded tree in dimension $d = 2$

label $\ell(v)$ that encodes the position of v in the lattice $\mathbb{Z}^d$. Then this measure is given by

$$\mu_T = \frac{1}{|V(T)|} \sum_{v \in V(T)} \delta_{\ell(v)}. \tag{5.7}$$

Note that our previous examples fit into this scheme for $d = 1$. For example, if we just use unit steps to the positive direction then this measure corresponds to the usual profile. Other examples are the tree classes discussed above, well-labelled and embedded trees (with increments 0 and ± 1), where both directions appear.

This procedure gets more interesting in a probabilistic framework. Suppose that we have random trees T_n of size n and a d-dimensional random (integer) vector η with zero mean and finite covariance matrix. Then, if we embed T_n by using an independent copy η_e of η for each edge e, we obtain random (mass) distributions μ_n in $\mathbb{R}^d$. Of course, they have to be scaled properly. If the average distance from the root in T_n is of order $\sqrt{n}$ then the average distance of an embedded point from the origin will be of order $\sqrt{\sqrt{n}} = n^{1/4}$, since $\sqrt{n}$ random steps of η will be of that order. After such a scaling one can expect to observe a limiting distribution.

Actually, such a limit exists in many instances and is in fact an universal law, the so-called *Integrated Super-Brownian Excursion* (ISE) (see Aldous [5]).

Formally, the ISE can be defined in the following way with help of the continuum random tree. For simplicity we just consider the one dimensional case here. Let us start (again) with a continuous function $g : [0, \infty) \to [0, \infty)$ of compact support with $g(0) = 0$ and consider the corresponding real tree $\mathcal{T}_g$ that is given by $\mathcal{T}_g = [0, \infty)/\sim$, where the equivalence relation $\sim$ is defined by

$$s \sim t \quad \Longleftrightarrow \quad d_g(s, t) = g(s) + g(t) - 2 \inf_{\min\{s,t\} \le u \le \max\{s,t\}} g(u) = 0.$$

We next introduce a Gaussian process[4] $(W_g(t),\, t \ge 0)$ with $W_g(0) = 0$, mean value $\mathbb{E}(W_g(t)) = 0$, and covariance

$$\mathbb{C}\mathrm{ov}(W_g(s), W_g(t)) = \mathbb{E}(W_g(s)W_g(t)) = \inf_{\min\{s,t\} \le u \le \max\{s,t\}} g(u).$$

In particular this implies

$$\mathbb{V}\mathrm{ar}(W_g(s) - W_g(t)) = d_g(s,t).$$

Consequently $d_g(s,t) = 0$ provides $W_g(s) = W_g(t)$ a.s. and, thus, $W_g(s)$ can also be interpreted as a stochastic process on $\mathcal{T}_g$. Let us have a closer look at this construction, if g represents the depth-first search on a planted plane tree (compare with Figure 4.1). The process starts at the planted vertex and is just a Brownian motion where the path on the edge joining the planted vertex and the original root acts as the time axis. Then at the original root the process splits into several processes (depending on the number of successors of the root and so on); of course, if we consider a path away from the root then this is just Brownian motion with this path as time axis. Suppose that we approximate the Brownian motion on an edge by a discrete step of size 0 or ± 1 (say with equal probability). Then any realisation of this discrete process is just an embedded tree with increments $0, \pm 1$. Hence, the process $W_g(t)$ can be considered as a continuous analogue of an embedded tree.

Suppose that g is supported on $[0,1]$. Then the process $W_g(t)$ defines an occupation measure

$$\mu_g(A) = \int_0^1 \mathbf{1}_A(W_g(s))\, ds,$$

where A is any Borel set. The measure μ_g can be considered as the value distribution of W_g and is, thus, a continuous analogue of the value distribution of the labels of an embedded tree.

Finally, the continuum random tree enters the scene. Let $\overline{W}(s)$ be defined by $W_{2e}(s)$, where e denotes Brownian excursion of duration 1 (this means two random elements are combined). This process is also called head of Brownian snake and by construction it can be seen as a process on the continuum random tree, too.

The (one dimensional) ISE is now defined by the (random) occupation measure

$$\mu_{\mathrm{ISE}}(A) = \int_0^1 \mathbf{1}_A(\overline{W}(s))\, ds$$

of the head of Brownian snake. If we replace the one dimensional Gaussian process by a d-dimensional one then we can define the d-dimensional ISE in the same way but we will not stay on this point.

[4] A Gaussian process $(X(t), t \in I)$ (with zero mean) is completely determined by a positive definite covariance function $B(s,t)$. All finite dimensional random vectors $(X(t_1), \ldots, X(t_k))$ are normally distributed with covariance matrix $(B(t_i,t_j))_{1 \le i,j \le k}$.

By Aldous' result (Theorem 4.6) we know that properly scaled Galton-Watson trees converge to the continuum random tree (where we assume that the offspring distribution ξ satisfies $\mathbb{E}\,\xi = 1$ and $0 < \mathbb{V}\mathrm{ar}\,\xi = \sigma_\xi^2 < \infty$). Let η be a random variable with $\mathbb{E}\,\eta = 0$ and variance $0 < \mathbb{V}\mathrm{ar}\,\eta = \sigma_\eta^2 < \infty$ that embeds the tree into the line (as described above). The displacements are now interpreted as labels: the root r has label $\ell(r) = 0$ and if v' is a successor of v that is linked by an edge e, then set $\ell(v') = \ell(v) + \eta_e$, where η_e is an independent copy of η. For example, if we consider planted plane trees ($\mathbb{E}\,x^\xi = 1/(2-x)$) and η is the uniform distribution on $\{-1, 0, 1\}$ then this construction leads to the embedded trees with increments 0 and ± 1 (also discussed above).

Then we have the following limit theorem (see Aldous [5] and Janson and Marckert [117]).

Theorem 5.4. *Let the labels $\ell(v)$ on Galton-Watson trees T_n be defined as above. Then with $\gamma = \sigma_\eta^{-1}\sigma_\xi^{1/2}$ we have*

$$\frac{1}{n}\sum_{v\in V(T_n)} \delta_{\gamma n^{-1/4}\ell(v)} \xrightarrow{d} \mu_{\mathrm{ISE}}, \tag{5.8}$$

with convergence in the space of probability measures on $\mathbb{R}$.[5]

The left hand side of (5.8) is just a scaled version of the (random) displacement distribution (5.7) of the (random) labels $\ell(v)$. Thus, one dimensional embedded Galton-Watson trees are actually approximated by the ISE (as it should be).

This result is also an analogue to (4.6). We just have to replace the labels $\ell(v)$ by the distances from the root and μ_{ISE} by the occupation time of Brownian excursion; of course, the scaling is different.

It is natural to ask whether there is a local version of Theorem 5.4, that is, an analogue to Theorem 4.10. Here one has to verify first whether the (random) measures μ_{ISE} has a proper (random) density. Actually, there is a corresponding *local time* of $\overline{W}$,

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_0^1 \mathbf{1}_{[t,t+\epsilon]}(\overline{W}(s))\,ds,$$

and thus there exists a continuous stochastic process

$$(f_{\mathrm{ISE}}(t),\ -\infty < t < \infty)$$

with $d\mu_{\mathrm{ISE}} = f_{\mathrm{ISE}}\,dt$, the density of the ISE (see [24]). In fact, a local result on these (vertical) profiles has been recently proved by Bousquet-Mélou and Janson [24] and by Devroye and Janson [52].

[5] The space of probability measures on $\mathbb{R}$ with the weak topology is a Polish space.

Theorem 5.5. *Additionally to the assumptions of Theorem 5.4 we assume that η is integer valued and aperiodic, that is, it is not concentrated on $d\mathbb{Z}$ for some $d > 1$. Let $(X_n(j))_{j\in\mathbb{Z}}$ the corresponding profile, that is, $X_n(j)$ is the number of vertices in T_n with $\ell(v) = j$, and $(X_n(t), -\infty < t < \infty)$ the linearly interpolated process. Then we have*

$$\left(n^{-3/4}X_n(n^{1/4}t),\ -\infty < t < \infty\right) \xrightarrow{d} (\gamma f_{\mathrm{ISE}}(\gamma t),\ -\infty < t < \infty)$$

in the space $\mathrm{C}_0(\mathbb{R})$, *where* $\gamma = \sigma_\eta^{-1}\sigma_\xi^{1/2}$.

Interestingly, Theorems 5.4 and 5.5 appear in other settings, too. We just mention an example, namely the vertical profile of (incomplete) binary trees, where we have precisely the same result although the labels are not random.

We recall that a (complete) binary tree is a plane rooted tree where each node has either two or no successors. In particular we can distinguish between internal nodes that have two successors and external ones. If we disregard the external nodes then the resulting object is a rooted tree, a so-called incomplete binary tree, where each node has either no successor, a right successor, a left successor or a left and a right successor.[6] There is a natural labelling of the nodes, namely the number of right steps minus the number of left steps on the path from the root to this node (see Figure 5.12). This labelling induces a natural embedding of these binary trees and also explains why we call the sequence $(X_n(j))_{j\in\mathbb{Z}}$, where $X_n(j)$ denotes the number of nodes of label j, *vertical profile.*

Theorem 5.6. *Let $X_n(j)$ denote the number of nodes of label j in naturally embedded reduced binary trees with n nodes and $(X_n(t), -\infty < t < \infty)$ the corresponding linearly interpolated process. Then we have*

$$\frac{1}{n}\sum_{v\in V(T_n)} \delta_{(2n)^{-1/4}\ell(v)} \xrightarrow{d} \mu_{\mathrm{ISE}},$$

and

$$\left(n^{-3/4}X_n(n^{1/4}t),\ -\infty < t < \infty\right) \xrightarrow{d} \left(2^{-1/4}f_{\mathrm{ISE}}(2^{-1/4}t),\ -\infty < t < \infty\right).$$

In view of the bijections between quadrangulations (together with their profile) and well-labelled and embedded trees (with increments 0 and ±1) the following theorem on quadrangulations is not unexpected (see [35] for details).

For convenience let $[L, N] = [L_{\mathrm{ISE}}, N_{\mathrm{ISE}}]$ denote the (random) support of μ_{ISE} and $\hat\mu_{\mathrm{ISE}}$ the shifted ISE that is supported by $[0, N - L]$, that is, $\hat\mu_{\mathrm{ISE}}((-\infty,\lambda]) = \mu_{\mathrm{ISE}}((-\infty,\lambda + L])$

[6] Reduced binary trees of this kind are easier to handle; for example, the counting problem for binary trees is based on the number of internal nodes. Therefore we have decided to formulate the following results in terms of this notion.

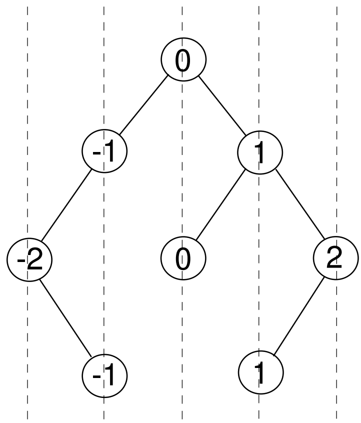


Fig. 5.12. Naturally embedded binary tree

Theorem 5.7. *Let* $(\lambda_{n,k})$ *denote the height profile and* r_n *the maximum distance from the root vertex in random quadrangulations with* n *vertices. Then*

$$\frac{1}{n}\sum_{k\ge 0}\lambda_{n,k}\,\delta_{\gamma n^{-1/4}k}\xrightarrow{d}\hat{\mu}_{\mathrm{ISE}}$$

and

$$\gamma n^{-1/4}r_n\xrightarrow{d}N_{\mathrm{ISE}}-L_{\mathrm{ISE}},$$

where $\gamma=2^{-1/4}$.

In what follows we will draw a brief glance on these general theorems. Due to the complexity of the topic we will not present a complete proof of these results but give some comments. Nevertheless we will discuss some special cases (embedded trees, binary trees) in detail in Sections 5.3 and 5.4. There are very interesting combinatorial aspects in relation to generating function. By using these generating function we prove, for example, $n^{-3/4}X_n(n^{1/4}\lambda)\xrightarrow{d}\gamma f_{\mathrm{ISE}}(\gamma\lambda)$ for every fixed real number λ (see Theorem 5.24).

The proofs of Theorems 5.4 and 5.5 that rely on the work by Aldous [4, 5], Chassaing and Schaeffer [35], Janson and Marckert [117], Bousquet-Mélou and Janson [24], and Devroye and Janson [52] are based on the following principles. First one observes that a Galton-Watson tree together with the labels that are induced by random increments η_e can be interpreted as a discrete version of the Brownian snake. One applies the depth-first search but instead of interpolating the distance from the root $x(i)$, one considers the labels $\ell(i)$, that is the sum of the increments η_e on the path from the root. By applying the (known) convergence of the scaled depth-first search to the

Brownian excursion, the central limit theorem and proper tightness estimates it follows that the *discrete snake process* converges to the Brownian snake (see, for example [117]). Theorem 5.4 is a direct consequence of this convergence result (recall that the ISE is the occupation measure of the Brownian snake).

Theorem 5.7 is deduced by similar arguments, however, one has to use some additional properties of the relation between quadrangulations and embedded trees (see [35]).

Now suppose that the distribution of the increments η is discrete (and aperiodic). Let $X_n(j)$ denote the number of nodes with label j in a random Galton-Watson tree of size n and let $X_n(t)$ be the corresponding linear interpolated process. Then the scaled process $n^{-3/4}X_n(n^{1/4}t)$ can be seen as the (random) density of a measure

$$\frac{1}{n}\sum_{j\ge 0} X_n(j)\,\nu_{jn^{-1/4},n^{-1/4}},$$

that is close to the (scaled) occupation measure

$$\frac{1}{n}\sum_{j\ge 0} X_n(j)\,\delta_{jn^{-1/4}}$$

of the label distribution (recall that $\nu_{h,y}$ is the measure with triangular density $f(x) = h^{-1}\max\{1-|x-y|/h, 0\}$, compare with Theorem 4.17 and its discussion). Thus, in order to prove Theorem 5.5 it is sufficient to prove tightness of $n^{-3/4}X_n(n^{1/4}t)$.

Bousquet-Mélou and Janson [24] have formulated an interesting sufficient condition for tightness (in this context).

Lemma 5.8. *Suppose that there exists a constant C such that for all $n \ge 1$ and $u \in [-\pi,\pi]$,*

$$\mathbb{E}\left(|n^{-1}\hat{X}_n(u)|^2\right) \le \frac{C}{1+nu^4}, \tag{5.9}$$

where

$$\hat{X}_n(u) = \sum_j X_n(j)e^{iju}$$

denotes the Fourier transform of the occupations measure $\sum_j X_n(j)\delta_j$ and that there exists a random variable W with

$$n^{-1/4}\sup\{|j| : X_n(j) \ne 0\} \xrightarrow{d} W. \tag{5.10}$$

Then $n^{-3/4}X_n(n^{1/4}t)$ is tight in $\mathrm{C}_0(\mathbb{R})$.

We sketch the proof of Lemma 5.8. First, by basic Fourier analytic tools it follows that (for $0 \le a < 3$)

$$\mathbb{E}\left(\int_{-\infty}^{\infty} |t|^a \left|n^{-3/4}X_n(n^{1/4}t)\right|^2 dy\right)$$

is finite so that with high probability

$$\int_{-\infty}^{\infty} |t|^a \left| n^{-3/4} X_n(n^{1/4}t) \right|^2 < A$$

for some $A > 0$. The condition (5.10) ensures that again with high probability

$$\left| n^{-3/4} X_n(n^{1/4}t) \right| = 0$$

for $|t| > M$, if M is chosen sufficiently large. Let $K_{M,A}$ denote the set of all functions in $\mathrm{C}_0(\mathbb{R})$ with $f(t) = 0$ for $|t| > M$ and

$$\int_{-\infty}^{\infty} |t|^a \, |f(t)|^2 < A.$$

Then $K_{M,A}$ is a relative compact subset of $\mathrm{C}([-M, M])$, and thus of $\mathrm{C}_0(\mathbb{R})$, too (compare with [24]). Since $n^{-3/4} X_n(n^{1/4}t)$ is contained in $K_{M,A}$ with high probability, this proves tightness.

Note that (5.10) is related to Theorems 5.7 and 5.22 and can be deduced from the convergence of the (scaled) discrete snake process to the Brownian snake. Actually, $W = \gamma^{-1} \max\{N_{\mathrm{ISE}}, -L_{\mathrm{ISE}}\}$.

It remains to verify (5.9). For this purpose we introduce the following notions. For each pair of vertices v, w in a random Galton-Watson tree T_n of size n, the path from v to w consists of two (possibly empty) parts, one going from v towards the root, ending at the last common ancestor $v \wedge w$ of v and w, and another part going from $v \wedge w$ to w in the direction away from the root. We will also prove a more general result, where we consider separately the lengths of these two parts. Define

$$h_n(x_1, x_2) = \mathbb{E}\left(\sum_{v,w \in T_n} x_1^{d(v, v\wedge w)} x_2^{d(w, v \wedge w)} \right), \tag{5.11}$$

where $d(v, w)$ denotes the distance between v and w, and

$$H(z, x_1, x_2) = \sum_{n \ge 1} y_n h_n(x_1, x_2) z^n,$$

where $y_n = \sum_{|T|=n} \omega(T)$. It can be shown (see [52]) that $H(z, x_1, x_2)$ can be explicitly represented as

$$H(z, x_1, x_2) = \frac{z x_1 x_2 \Phi''(y(z)) G(z, x_1) G(z, x_2) + z\Phi'(y(z)) \left(x_1 G(z, x_1) + x_2 G(z, x_2) \right) + y(z)}{1 - z\Phi'(y(z))},$$

where

$$G(z,x) = \frac{y(z)}{1 - zx\Phi'(y(z))}.$$

On the other hand, there is a direct relation to $\hat{X}_n(u)$. Let $\ell(v)$ denote the random label of the vertex v, that is, the sum of the increments on the path from the root to v and let $\varphi(t) = \mathbb{E}\, e^{it\eta}$ the characteristic function of the distribution of the increment η. Then from

$$\begin{aligned}\mathbb{E}\left(e^{iu(\ell(v)-\ell(w))}|T_n\right) &= \mathbb{E}\left(e^{iu(\ell(v)-\ell(v\wedge w))}|T_n\right)\mathbb{E}\left(e^{-iu(\ell(w)-\ell(v\wedge w))}|T_n\right)\\ &= \varphi(u)^{d(v,v\wedge w)}\overline{\varphi(u)}^{d(w,v\wedge w)}\end{aligned}$$

we obtain

$$\begin{aligned}\mathbb{E}\left(|n^{-1}\hat{X}_n(u)|^2\right) &= \frac{1}{n^2}\mathbb{E}\left(\sum_{v,w\in T_n}\left(e^{iu(\ell(v)-\ell(w))}\right)\right)\\ &= \frac{1}{n^2}\mathbb{E}\left(\sum_{v,w\in T_n}\varphi(u)^{d(v,v\wedge w)}\overline{\varphi(u)}^{d(w,v\wedge w)}\right)\\ &= \frac{1}{n^2}\, h_n(\varphi(u),\overline{\varphi(u)}).\end{aligned}$$

This means that we have almost direct access to $\mathbb{E}|n^{-1}\hat{X}_n(u)|^2$ with help of analytic tools. In particular, by singularity analysis it follows that

$$|h_n(x_1,x_2)| \le \frac{C}{n\,|1-x_1|\,|1-x_2|}$$

uniformly for x_1, x_2 in a Δ-domain (compare with [52]). We also have $|\varphi(u) - 1| \ge cu^2$ for $u \in [-\pi,\pi]$ and consequently

$$\begin{aligned}(1+nu^4)\,\mathbb{E}\left(|n^{-1}\hat{X}_n(u)|^2\right) &\le 1 + nu^4\frac{C}{n|1-\varphi(u)|^2}\\ &\le 1 + \frac{C}{c^2}.\end{aligned}$$

Thus, Lemma 5.8 can be applied and Theorem 5.5 follows.

The proof of Theorem 5.6 can be worked out in a similar way.

5.2.3 The Distribution of the ISE

The definition of the ISE does not provide directly analytic expressions for the distribution of the ISE. Nevertheless there are integral representations for several statistics of the ISE. In this section we list some of them, however, they are derived an indirect way. More precisely, they will appear in the proofs of the Theorems 5.22–5.26 (in Section 5.4) where we analyse limiting distributional properties of the profile of embedded trees. However, by Theorems 5.4 and 5.5 it follows that these limiting distributions have to be corresponding functionals of the ISE (see also [23] and [24]).

We start with the distribution of μ_{ISE}. Let $G^+(\lambda) = \mu_{\mathrm{ISE}}((\lambda,\infty)) = 1 - \mu_{\mathrm{ISE}}((-\infty,\lambda])$ denote the (random) tail distribution function. Then the Laplace transform (for $|a| < 1$) is given by

$$\mathbb{E}\, e^{aG^+(\lambda)} = 1 + \frac{48}{i\sqrt{\pi}} \int_{\Gamma} \frac{B(a/v^4)e^{-2\lambda v}}{(1+B(a/v^4)e^{-2\lambda v})^2} v^5 e^{v^4}\, dv,$$

where

$$B(x) = -\frac{(1-D(x))(1-2D(x))}{(1+D(x))(1+2D(x))}, \qquad D(x) = \sqrt{\frac{1+\sqrt{1-x}}{2}}, \tag{5.12}$$

and the integral is taken over

$$\Gamma = \{1 - te^{-i\pi/4}, t \in (-\infty, 0]\} \cup \{1 + te^{i\pi/4}, t \in [0,\infty)\} \tag{5.13}$$

(see Figure 5.13). The expected value has also a series expansion

$$\mathbb{E}\, G^+(\lambda) = \frac{1}{2\sqrt{\pi}} \sum_{m\ge 0} \frac{(-2\lambda)^m}{m!} \cos\left(\frac{m\pi}{4}\right) \Gamma\left(\frac{m+2}{4}\right).$$

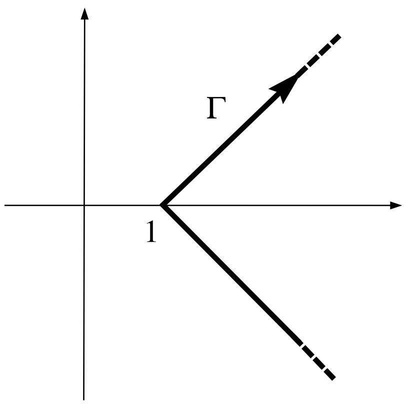


Fig. 5.13. Contour of integration Γ

Next let

$$N_{\mathrm{ISE}} = \sup\{y : \mu_{\mathrm{ISE}}((y,\infty)) > 0\}$$

denote the supremum of the support of the ISE. Its tail distribution function can be expressed by

$$\mathbb{P}\{N_{\mathrm{ISE}} > \lambda\} = \frac{12}{i\sqrt{\pi}} \int_{\Gamma} \frac{v^5 e^{v^4}}{\sinh^2(\lambda v)} dv$$

Observe that the integration contour Γ can be replaced by its translated version

$$\Gamma_0 = \{-re^{-i\pi/4}, r \in (-\infty, 0]\} \cup \{re^{i\pi/4}, r \in [0, \infty)\}.$$

This parametrisation of Γ_0 by r splits the integral into two real integrals, and one finds

$$\begin{aligned} G(\lambda) &= -\frac{12}{\sqrt{\pi}} \int_0^\infty \left(\frac{1}{\sinh^2(\lambda r e^{i\pi/4})} + \frac{1}{\sinh^2(\lambda r e^{-i\pi/4})} \right) r^5 e^{-r^4} dr \\ &= \frac{48}{\sqrt{\pi}} \int_0^\infty \frac{1 - \cos(\sqrt{2}\lambda r)\cosh(\sqrt{2}\lambda r)}{(\cosh(\sqrt{2}\lambda r) - \cos(\sqrt{2}\lambda r))^2} r^5 e^{-r^4} dr \qquad (5.14) \\ &= \frac{6}{\sqrt{\pi}\lambda^6} \int_0^\infty \frac{1 - \cos u \cosh u}{(\cosh u - \cos u)^2} u^5 e^{-u^4/(4\lambda^4)} du. \end{aligned}$$

The density is given by

$$\begin{aligned} f(\lambda) &= \frac{24}{i\sqrt{\pi}} \int_{\Gamma} \frac{\cosh(\lambda v) v^6 e^{v^4}}{\sinh^3(\lambda v)} dv \\ &= \frac{6}{\sqrt{\pi}\lambda^{11}} \int_0^\infty \frac{1 - \cos u \cosh u}{(\cosh u - \cos u)^2} u^5 (6\lambda^4 - u^4) e^{-u^4/(4\lambda^4)} du, \end{aligned}$$

where the contour Γ is given by (5.13).

The moments of N_{ISE} are explicit, for $\Re(r) > -4$ we have

$$\mathbb{E}(N_{\mathrm{ISE}}^r) = \frac{24\sqrt{\pi}\, \Gamma(r+1)\zeta(r-1)}{2^r \Gamma((r-2)/4)},$$

with analytic continuation at the points $-3, -2, -1, 2$. For example

$$\mathbb{E}\, N_{\mathrm{ISE}} = \frac{3\sqrt{\pi}}{2\Gamma(3/4)}, \quad \mathbb{E}(N_{\mathrm{ISE}}^2) = 3\sqrt{\pi}.$$

These moments were already obtained by Delmas [46] and rediscovered by [23].

Finally, we consider the density of the ISE f_{ISE}. The one dimensional distribution of $f_{\mathrm{ISE}}(\lambda)$ is characterised by the Laplace transform (for $|a| < 4/\sqrt{3}$)

$$\mathbb{E}\, e^{a f_{\mathrm{ISE}}(\lambda)} = 1 + \frac{48}{i\sqrt{\pi}} \int_{\Gamma} \frac{A(a/v^3)e^{-2\lambda v}}{(1 + A(a/v^3)e^{-2\lambda v})^2} v^5 e^{v^4} dv,$$

where $A(x) = A$ is the unique solution of

$$A = \frac{x}{24}\frac{(1+A)^3}{1-A} \tag{5.15}$$

satisfying $A(0) = 0$, and the integral is taken over the contour Γ given by (5.13).

The expected value is given by

$$\begin{aligned}\mathbb{E}\, f_{\mathrm{ISE}}(\lambda) &= (2\pi)^{-1/2}\int_0^\infty y^{1/2}\exp\left(-\frac{\lambda^2}{2y}-\frac{y^2}{2}\right)dy\\ &= \frac{2^{-1/4}}{\sqrt{\pi}}\sum_{m\ge 0}\frac{(-2^{3/4}|\lambda|)^m}{m!}\cos\frac{(m+1)\pi}{4}\Gamma\left(\frac{m+3}{4}\right).\end{aligned}$$

In particular, $f_{\mathrm{ISE}}(0)$ has the same distribution as $2^{1/4}3^{-1}T^{-1}$, where T is a positive 2/3-stable variable defined by its Laplace transform $\mathbb{E}\, e^{-tT} = e^{-t^{2/3}}$.

Hence $f_{\mathrm{ISE}}(0)$ has the moments

$$\mathbb{E}\, f_{\mathrm{ISE}}(0)^r = 2^{r/4}3^{-r}\frac{\Gamma(3r/4+1)}{\Gamma(r/2+1)}, \qquad -4/3 < r < \infty.$$

5.3 Combinatorics on Embedded Trees

The discussion in the previous section shows that there is considerable interest in labelled trees, where the labels of adjacent vertices are described by a fixed law (for example, that they differ at most by one).

We have also shown that well-labelled trees can be counted with help of generating functions (5.6) that have a *mysterious* representation in terms of $Z(t)$.

We will next discuss three different kinds of trees: two versions of embedded trees and binary trees (where the labels are not randomly chosen but correspond to the vertical position). Interestingly, the counting procedures are almost the same. We follow the work of Bousquet-Mélou [23].

The subsequent asymptotic analysis (see Section 5.4) provides one dimensional versions of Theorems 5.4 and 5.5. As already indicated, a full proof of these theorems will not be given in this book.

5.3.1 Embedded Trees with Increments ±1

We start by considering a family of embedded planted plane trees, where the root is labelled 0, and the labels of two adjacent nodes differ by ± 1 (see Figure 5.14), where the counting procedure is slightly easier; the case of increments 0 and ± 1 will be discussed in Section 5.3.2. Furthermore, trees with increments ± 1 appear as objects related to special triangulations as we have mentioned in Section 5.1.

The total number of trees with n edges of that kind is given by

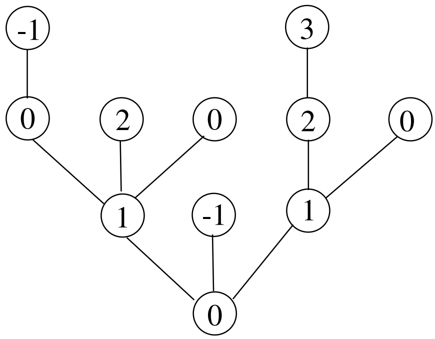


Fig. 5.14. Embedded tree with increments ±1

$$t_n = 2^n p_{n+1} = \frac{2^n}{n+1}\binom{2n}{n}.$$

Let $T(t)$ denote the (ordinary) generating function of these numbers.

For $j \in \mathbb{N}$, let $T_j(t)$ be the generating function of labelled trees in which all labels are less than or equal to j. Clearly, $T_j(t)$ converges to $T(t)$ (in the space of formal power series in t) as j goes to infinity. It is very easy to describe an infinite set of equations that completely defines the collection of series $T_j(t)$.

Lemma 5.9. *The series T satisfies*

$$T(t) = \frac{1}{1 - 2tT(t)}. \tag{5.16}$$

More generally, for $j \geq 0$,

$$T_j(t) = \frac{1}{1 - t(T_{j-1}(t) + T_{j+1}(t))}$$

while $T_j(t) = 0$ for $j < 0$.

Proof. First, replacing each label k by $j - k$ shows that $T_j(t)$ is also the generating function of trees *rooted at* j and having only non-negative labels (we say that a tree is rooted at j, if its root has label j). This means that we actually reduce the problem to a counting problem for specific well-labelled trees. Secondly, all subtrees of the root of such a tree are again trees of that kind, however, rooted at $j \pm 1$ (compare with the proof of Theorem 5.3). This leads to the recurrence

$$\begin{aligned} T_j(t) &= 1 + t(T_{j-1}(t) + T_{j+1}(t)) + (t(T_{j-1(t)} + T_{j+1}(t)))^2 + \cdots \\ &= \frac{1}{1 - t(T_{j-1}(t) + T_{j+1}(t))}. \end{aligned}$$

Finally, be letting (formally) $j \to \infty$ we get to (5.16).

The above lemma also shows that the series $T(t)$, counting labelled trees by edges, is algebraic. What is far less clear – but nevertheless true – is that each of the series $T_j(t)$ is algebraic too, as stated in the proposition below. These series will be expressed in terms of the series $T(t)$ and of the unique formal power series $Z(t)$, with constant term 0, satisfying

$$Z(t) + \frac{1}{Z(t)} = \frac{1}{tT(t)^2}. \tag{5.17}$$

Observe that $Z(t)$ and t are related by

$$Z(t) = t\,\frac{(1+Z(t))^4}{1+Z(t)^2} \tag{5.18}$$

and $Z(t)$ and $T(t)$ by

$$T(t) = \frac{(1+Z(t))^2}{1+Z(t)^2}. \tag{5.19}$$

Proposition 5.10. *Let $T_j(t)$ be the generating function of embedded trees with increments ±1 having no label greater than j. Then we have*

$$T_j(t) = T(t)\,\frac{\left(1-Z(t)^{j+1}\right)\left(1-Z(t)^{j+5}\right)}{\left(1-Z(t)^{j+2}\right)\left(1-Z(t)^{j+4}\right)}, \tag{5.20}$$

with $Z(t)$ given by (5.18)*. These functions are algebraic of degree* 2 *(at most). In particular, $T_0(t)$ satisfies the equation*

$$T_0(t) = 1 - 11\,t - t^2 + 4\,t\,(3+2\,t)\,T_0(t) - 16\,t^2 T_0(t)^2.$$

Proof. It is very easy to check, using (5.18)–(5.19), that the above values of $T_j(t)$ satisfy the recurrence relation of Lemma 5.9 and the initial condition $T_{-1}(t) = 0$ (compare also with the proof of Theorem 5.3).

The equation satisfied by $T_0(t)$ is obtained by eliminating $T(t)$ and $Z(t)$ from the case $j = 0$ of (5.20). Then an induction on j, based on Lemma 5.9, implies that each $T_j(t)$ is quadratic (at most) over $\mathbb{Q}(t)$.

Remark 5.11 *As we have mentioned in Section 5.1, trees counted by $T_0(t)$ (equivalently, the trees having only non-negative labels) are known to be in bijection with certain planar maps called Eulerian triangulations (see also* [26]*).*

Let us now turn to a bivariate counting problem. Let $S_j(t,u)$ be the generating function of embedded trees (with increments ±1), counted by the number of edges (variable t) and the number of nodes labelled j (variable u). Clearly, $S_j(t,1) = T(t)$ for all j. Moreover, an obvious symmetry shows that $S_j(t,u) = S_{-j}(t,u)$.

Lemma 5.12. *For $j \neq 0$,*

$$S_j(t,u) = \frac{1}{1 - t(S_{j-1}(t,u) + S_{j+1}(t,u))}, \tag{5.21}$$

while for $j = 0$,

$$S_0(t,u) = \frac{u}{1 - t(S_{-1}(t,u) + S_1(t,u))} = \frac{u}{1 - 2tS_1(t,u)}. \tag{5.22}$$

Proof. Observe that $S_j(t,u)$ is also the generating function of labelled trees rooted at j, counted by the number of edges and the number of nodes labelled 0. The decomposition of planted plane trees provides the lemma. The only difference between the cases $j = 0$ and $j \neq 0$ lies in the generating function of the tree reduced to a single node.

Again, the series $S_j(t,u)$ has a remarkable explicit representation (that shows that it is algebraic, too, for reasons that currently remain mysterious from the combinatorial point of view). They can be expressed in terms of the series $T(t)$ and $Z(t)$ given by (5.18)–(5.19). By symmetry we also $S_j(t,u) = S_{-j}(t,u)$. Thus, it is sufficient to consider the case $j \geq 0$.

Proposition 5.13. *For any $j \geq 0$, the generating function $S_j(t,u)$ that counts embedded trees (with increments ± 1) by the number of edges and the number of nodes with label j is given by*

$$S_j(t,u) = T(t)\,\frac{\left(1 + \mu(t,u)Z(t)^j\right)\left(1 + \mu(t,u)Z(t)^{j+4}\right)}{\left(1 + \mu(t,u)Z(t)^{j+1}\right)\left(1 + \mu(t,u)Z(t)^{j+3}\right)}, \tag{5.23}$$

where $Z(t)$ is given by (5.18) *and $\mu = \mu(t,u)$ is the unique formal power series in t satisfying*

$$\mu = (u-1)\frac{(1+Z(t)^2)(1+\mu Z(t))(1+\mu Z(t)^2)(1+\mu Z(t)^3)}{(1+Z(t))(1+Z(t)+Z(t)^2)(1-Z(t))^3(1-\mu Z(t)^2)}. \tag{5.24}$$

The series $\mu(t,u)$ has polynomial coefficients in u and satisfies $\mu(t,1) = 0$. In particular, $S_0(t,u)$ satisfies the equation

$$\frac{(T(t) - S_0(t,u))^2}{(u-1)^2} = 1 - \frac{2(1 - T(t)^2)}{2 + S_0(t,u) - S_0(t,u)T(t)}. \tag{5.25}$$

Proof. First, observe that the family of series $S_0(t,u), S_1(t,u), S_2(t,u), \ldots$ is completely determined by (5.21) (taken for $j > 0$) and the second part of (5.22). The fact that for any series $\mu(t,u) \in \mathbb{Q}(u)[[t]]$ the expression (5.23) satisfies (5.21) for all $j > 0$ is a straightforward verification, once t and $T(t)$ have been expressed in terms of $Z(t)$ (see (5.18) and (5.19)). In order to prove that (5.23) is the correct expression of $S_j(t,u)$, it remains to satisfy the second part of (5.22). This last condition provides a polynomial equation relating $\mu(t, u)$, $T(t)$, $Z(t)$, t and u. In this equation, replace t and $T(t)$ by their expressions in terms of $Z(t)$ (given by (5.18)–(5.19)). This exactly gives (5.24).

Equation (5.25) satisfied by S_0 is obtained by eliminating $\mu(t, u)$ and $Z(t)$ (using (5.24) and (5.19)) from the expression (5.23) of $S_0(t, u)$.

In the proof of Lemma 5.25 we will need a closed form expression for $\mu(t, u)$ in terms of $Z(t)$. Write

$$v = \frac{(u-1)Z(t)(1+Z(t)^2)}{(1+Z(t))(1+Z(t)+Z(t)^2)(1-Z(t))^3}.$$

Then $\mu(t, u)$ is given by

$$\mu(t,u) = \frac{1}{Z(t)^2}\left(\frac{2}{1+v(1-Z(t))^2/3+2/3\sqrt{3+v^2(1-Z(t))^4}\cos(\varphi/3)} - 1\right) \tag{5.26}$$

where

$$\varphi = \arccos\left(\frac{-9v(1+4Z(t)+Z(t)^2)+v^3(1-Z(t))^6}{(3+v^2(1-Z(t))^4)^{3/2}}\right).$$

The reason for that is that equation (5.24) which defines $\mu(t, u)$ can be rewritten to

$$\mu(t,u) = \frac{v}{Z(t)}\frac{(1+\mu(t,u)Z(t))(1+\mu(t,u)Z(t)^2)(1+\mu(t,u)Z(t)^3)}{1-\mu(t,u)Z(t)^2}.$$

Hence, $\mu(t, u)$ is the unique formal power series in v (with rational coefficients in $Z(t)$) that satisfies the above equation and equals 0 when v is 0. And it is not hard to check that the expression given in (5.26) satisfies these two conditions.

Let us finally study a third problem. Let $R_j(t, u)$ be the generating function of embedded trees (with increments ± 1), counted by the number of edges (variable t) and the number of nodes labelled at least j (variable u).

Lemma 5.14. *The set of series $R_0(t, u), R_1(t, u), R_2(t, u), \dots$ is completely determined by the following equations: for $j \geq 1$,*

$$R_j(t,u) = \frac{1}{1-t(R_{j-1}(t,u)+R_{j+1}(t,u))} \tag{5.27}$$

and

$$R_0(t,u) = uR_1(tu, 1/u). \tag{5.28}$$

More generally, for all $j \in \mathbb{Z}$, one has:

$$R_{-j}(t,u) = uR_{j+1}(tu, 1/u). \tag{5.29}$$

Proof. For all $j \in \mathbb{Z}$, the series $R_j(t,u)$ is also the generating function of trees rooted at j, counted by their number of edges and the number of nodes having a non-positive label. The equation satisfied by j, for $j \geq 0$, follows once again from the recursive structure.

It remains to prove the symmetry relation (5.29). For any tree T, let $n_{\leq 0}(T)$ denote the number of nodes of T having a non-positive label. We use similar notations for the number of nodes having label at most j, etc. Let $\mathcal{T}_{j,n}$ denote the set of trees rooted at j and having n edges. As observed above,

$$R_{-j}(t,u) = \sum_{n\geq 0} t^n \sum_{T\in\mathcal{T}_{-j,n}} u^{n_{\leq 0}(T)} = \sum_{n\geq 0} t^n \sum_{T\in\mathcal{T}_{-j,n}} u^{n+1-n_{>0}(T)},$$

because a tree with n edges has a total of $n+1$ nodes. A translation of all labels by -1 gives

$$R_{-j}(t,u) = u\sum_{n\geq 0}(tu)^n \sum_{T\in\mathcal{T}_{-j-1,n}} u^{-n_{\geq 0}(T)},$$

while replacing each label k by $-k$ finally gives

$$R_{-j}(t,u) = u\sum_{n\geq 0}(tu)^n \sum_{T\in\mathcal{T}_{j+1,n}} u^{-n_{\leq 0}(T)} = uR_{j+1}(tu, 1/u).$$

Remarkably, the series $R_j(t,u)$ admit a closed form expression in terms of $T(t)$ and $Z(t)$, too. Due to the symmetry (5.29) it is sufficient to consider the case $j \geq 0$.

Proposition 5.15. *Let $j \geq 0$. The generating function $R_j(t,u)$ that counts embedded trees (with increments ± 1) by the number of edges and the number of nodes with labels $\geq j$ is given by*

$$R_j(t,u) = T(t)\,\frac{\left(1+\nu(t,u)Z(t)^j\right)\left(1+\nu(t,u)Z(t)^{j+4}\right)}{\left(1+\nu(t,u)Z(t)^{j+1}\right)\left(1+\nu(t,u)Z(t)^{j+3}\right)}, \tag{5.30}$$

where $Z(t)$ is determined by (5.18) *and $\nu(t,u)$ is expressed by*

$$\nu(t,u) = \frac{P(t,u)}{Z(t)}\,\frac{1-P(t,u)(1+Z(t))-P(t,u)^2(1+Z(t)+Z(t)^2)}{1+Z(t)+Z(t)^2+P(t,u)Z(t)(1+Z(t))-P(t,u)^2Z(t)^2},$$

with

$$P(t,u) = (1+Z(t))\frac{1-V(t,u)-\sqrt{\Delta(t,u)}}{2V(t,u)Z(t)},$$

in which

$$V(t,u) = \frac{1-\sqrt{\frac{1-8tu}{1-8t}}}{4},$$

and

$$\Delta(t,u) = (1-V(t,u))^2 - \frac{4Z(t)V(t,u)^2}{(1+Z(t))^2}.$$

We note that the generating function $\nu(t,u)$ considered as a series in t has polynomial coefficients in u and the first terms in its expansion are

$$\nu(t,u) = (u-1)\Big(1+2\,ut+\big(7\,u+6\,u^2\big)\,t^2+\big(32\,u+36\,u^2+23\,u^3\big)\,t^3+O(t^4)\Big).$$

Proof. In the proof of Proposition 5.13 we have already checked, that for any formal power series ν in t, the series defined by (5.30) for $j \geq 0$ satisfy the recurrence relation (5.27) for $j \geq 1$. It remains to prove that one can choose ν so as to satisfy (5.28). For any formal power series A in t having rational coefficients in u, we denote by $\tilde{A}$ the series $\tilde{A}(t,u) = A(tu, 1/u)$. In particular $\tilde{T}(t,u) = T(tu)$. Observe that $\tilde{\tilde{A}} = A$. With this notation, if $R_j(t,u)$ is of the generic form (5.30), the relation (5.28) holds, if and only if

$$1+\nu = u\,\frac{\tilde{T}}{T}\,\frac{(1+\nu Z)(1+\nu Z^3)(1+\tilde{\nu}\tilde{Z})(1+\tilde{\nu}\tilde{Z}^5)}{(1+\nu Z^4)(1+\tilde{\nu}\tilde{Z}^2)(1+\tilde{\nu}\tilde{Z}^4)}, \tag{5.31}$$

where we use the abbreviations $T = T(t)$, $\tilde{T} = \tilde{T}(t,u)$, $Z = Z(t)$, $\tilde{Z} = \tilde{Z}(t,u) = Z(tu)$, $\nu = \nu(t,u)$, and $\tilde{\nu} = \tilde{\nu}(t,u) = \nu(tu, 1/u)$. Let $\mathbb{R}_m[u]$ denote the space of polynomials in u, with real coefficients, of degree at most m. Let $\mathbb{R}_n[u][[t]]$ denote the set of formal power series in t with polynomial coefficients in u such that for all $m \leq n$, the coefficient of t^m has degree at most m. Observe that this set of series is stable under the usual operations on series: sum, product, and quasi-inverse. Write $\nu = \sum_{n\geq 0} \nu_n(u)t^n$. We are going to prove, by induction on n, that (5.31) determines uniquely each coefficient $\nu_n(u)$, and that this coefficient belongs to $\mathbb{R}_{n+1}[u]$.

First, observe that for any formal power series ν, the right-hand-side of (5.31) is $u + O(t)$. This implies $\nu_0(u) = u - 1$. Now assume that our induction hypothesis holds for all $m < n$. Recall that Z is a multiple of t: this implies that νZ belongs to $\mathbb{R}_n[u][[t]]$. The induction hypothesis also implies that the coefficient of t^m in $u\tilde{\nu}$ belongs to $\mathbb{R}_{m+1}[u]$, for all $m < n$. Note that $\tilde{Z} = Z(tu) = tu + O(t^2)$ is a multiple of t and u and also belongs to $\mathbb{R}_n[u][[t]]$. This implies that $\tilde{\nu}\tilde{Z}$ belongs to $\mathbb{R}_n[u][[t]]$, too. The same is true for all the other series occurring in the right-hand-side of (5.31), namely $T, \tilde{T}, Z, \tilde{Z}$. Given the closure properties of the set $\mathbb{R}_n[u][[t]]$, we conclude that the right-hand-side of (5.31), divided by u, belongs to this set. Moreover, the fact that Z and $\tilde{Z}$ are multiples of t guarantees that the coefficient of t^n in this series only involves the $\nu_i(u)$ for $i < n$. By extracting the coefficient of t^n in (5.31), we conclude that $\nu_n(u)$ is uniquely determined and belongs to $u\mathbb{R}_n[u] \subset \mathbb{R}_{n+1}[u]$.

This completes the proof of the existence and uniqueness of the series ν satisfying (5.31). Also, setting $u = 1$ (that is, $\tilde{T} = T$ and $\tilde{Z} = Z$) in this equation shows that $\nu(t,1) = 0$.

Let us now replace t by tu and u by $1/u$ in (5.31). This gives

$$1+\tilde{\nu} = \frac{1}{u}\,\frac{T}{\tilde{T}}\,\frac{(1+\tilde{\nu}\tilde{Z})(1+\tilde{\nu}\tilde{Z}^3)(1+\nu Z)(1+\nu Z^5)}{(1+\tilde{\nu}\tilde{Z}^4)(1+\nu Z^2)(1+\nu Z^4)}. \tag{5.32}$$

In the above two equations, replace T by its expression (5.19) in terms of Z. Similarly, replace $\tilde{T}$ by its expression in terms of $\tilde{Z}$. Finally, it follows from (5.18) and from the fact that $\tilde{Z} = Z(tu)$ that

$$u = \frac{\tilde{Z}}{Z} \frac{(1+Z)^4(1+\tilde{Z}^2)}{(1+\tilde{Z})^4(1+Z^2)}. \tag{5.33}$$

Replace u by this expression in (5.31) and (5.32). Eliminate $\tilde{\nu}$ between the resulting two equations: this gives a polynomial equation of degree 2 in ν that relates ν, Z and $\tilde{Z}$. The elimination of $\tilde{Z}$ between this quadratic equation and (5.33) provides an equation of degree 4 in ν that relates $\nu = \nu(t,u)$ to $Z = Z(t)$ and u.

We do not make this equation explicit but it follows from (5.31) and (5.32). In this equation, we replace u in terms of

$$\delta = \delta(t,u) = 1 - 8(u-1)\frac{Z(t)(1+Z(t)^2)}{(1-Z(t))^4} = \frac{1-8tu}{1-8t},$$

and Z, that is, we set

$$u = 1 - \frac{1-\delta}{8} \frac{(1-Z)^4}{Z(1+Z^2)}.$$

Then we replace δ in terms of

$$V = \frac{1-\sqrt{\delta}}{4},$$

that is, we set $\delta = (1-4V)^2$. Interestingly, the resulting equation factors into two terms. Each of them is quadratic in ν. In order to decide which of these factors cancels, one uses the fact that when $u = 1$ (that is, $V = 0$), the series ν must be 0. It remains to solve a quadratic equation in ν. Its discriminant is found to be Δ, and it is convenient to introduce the series P and satisfies the equation

$$P = \frac{V}{1+Z}(1+P)(1+ZP).$$

5.3.2 Embedded Trees with Increments $0, \pm 1$

Next we consider the case of embedded trees with increments 0 and ± 1. As we have discussed in Section 5.1 this class of trees is of interest because of its relation to quadrangulations, but there are also relations to general planar maps [26, 35].

As above, let $T_j(t)$ be the generating function of labelled trees in which all labels are at most j, counted by their number of edges.[7] Let $S_j(t,u)$ be the

[7] We use the same notation as in Section 5.3.1 although they correspond to different trees. However, it will be always clear from the context which kind of trees is considered.

generating function of labelled trees, counted by the number of edges (variable t) and the number of nodes labelled j (variable u). Finally, let $R_j(t,u)$ be the generating function of labelled trees, counted by the number of edges and the number of nodes having label at least j. As above, it is easy to write an infinite system of equations defining any of the families $T_j(t)$, $S_j(t,u)$ or $R_j(t,u)$. The only difference to our first family of trees is that a third case arises in the decomposition of trees (compare also with the proof of Theorem 5.3). In particular, the generating function $T = T(t)$ counting these embedded trees now satisfies

$$T(t) = \frac{1}{1 - 3tT(t)},$$

while for $j \geq 0$,

$$T_j(t) = \frac{1}{1 - t(T_{j-1}(t) + T_j(t) + T_{j+1}(t))}. \tag{5.34}$$

The equations of Lemma 5.12 and Lemma 5.14 are modified in a similar way. The three infinite systems of equations obtained in that way can be solved using the same techniques as in Section 5.3.1. The solutions are expressed in terms of the above series $T(t)$ and the unique formal power series $Z(t)$, with constant term 0, satisfying

$$Z(t) + \frac{1}{Z(t)} + 1 = \frac{1}{tT(t)^2} \tag{5.35}$$

or equivalently

$$Z(t) = t\,\frac{(1 + 4Z(t) + Z(t)^2)^2}{1 + Z(t) + Z(t)^2}, \tag{5.36}$$

resp.

$$T(t) = \frac{1 + 4Z(t) + Z(t)^2}{1 + Z(t) + Z(t)^2}.$$

We state the counterparts of Propositions 5.10, 5.13 and 5.15 without proof.

Proposition 5.16. *Let $T_j(t)$ be the generating function of embedded trees (with increments 0 and ± 1) having no label greater than j. Then we have (for $j \geq -1$)*

$$T_j(t) = T(t)\,\frac{(1 - Z(t)^{j+1})(1 - Z(t)^{j+4})}{(1 - Z(t)^{j+2})(1 - Z(t)^{j+3})},$$

where $Z(t)$ is given by (5.36). These functions are algebraic of degree (at most) 2. In particular, $T_0(t)$ satisfies the equation

$$T_0(t) = 1 - 16\,t + 18\,tT_0(t) - 27\,t^2 T_0(t)^2.$$

Proposition 5.17. *For any $j \geq 0$, the generating function $S_j(t,u)$ that counts embedded trees (with increments 0 and ± 1) by the number of edges and the number of nodes labelled j is given by*

$$S_j(t,u) = T(t)\,\frac{(1+\mu(t,u)Z(t)^j)(1+\mu(t,u)Z(t)^{j+3})}{(1+\mu(t,u)Z(t)^{j+1})(1+\mu(t,u)Z(t)^{j+2})}, \tag{5.37}$$

where $Z(t)$ is given by (5.36) *and $\mu(t,u)$ is the unique formal power series in t satisfying*

$$\mu(t,u) = (u-1)\frac{(1+Z(t)+Z(t)^2)(1+\mu(t,u)Z(t))^2(1+\mu(t,u)Z(t)^2)^2}{(1+Z(t))^2(1-Z(t))^3(1-\mu(t,u)^2Z(t)^3)}.$$

The series $\mu(t,u)$ has polynomial coefficients in u and satisfies $\mu(t,1)=0$. In particular, $S_0(t,u)$ satisfies the equation

$$\frac{9T(t)^4(u-1)^2}{(T(t)-S_0(t,u))^2} = 9T(t)^2 - 2T(t)(T(t)-1)(2T(t)+1)S_0(t,u) + (T(t)-1)^2S_0(t,u)^2.$$

Recall that by symmetry $S_{-j}(t,u) = S_j(t,u)$. Thus (5.37) provides an explicit representation for all $j \in \mathbb{Z}$.

The functions $R_j(t,u)$ have a similar symmetry: $R_{-j}(t,u) = uR_{j+1}(tu,1/u)$. Hence, it is again sufficient to consider just the case $j \ge 0$.

Proposition 5.18. *Let $j \ge 0$. The generating function $R_j(t,u)$ that counts embedded trees (with increments* 0 *and* ± 1*) by the number of edges and the number of nodes labelled j or more is given by*

$$R_j(t,u) = T(t)\,\frac{(1+\nu(t,u)Z(t)^j)(1+\nu(t,u)Z(t)^{j+3})}{(1+\nu(t,u)Z(t)^{j+1})(1+\nu(t,u)Z(t)^{j+2})},$$

where $Z(t)$ is given by (5.36) *and $\nu(t,u)$ is a formal power series in t, with polynomial coefficients in u.*[8] *This series satisfies $\nu(t,1) = 0$ and the first terms in its expansion are*

$$\nu(t,u) = (u-1)\Big(1+3\,ut+\big(15\,u+14\,u^2\big)\,t^2+\big(104\,u+117\,u^2+83\,u^3\big)\,t^3+O(t^4)\Big).$$

5.3.3 Naturally Embedded Binary Trees

In this section we study incomplete binary trees. Such trees are either empty or have a root to which a left and right subtree (both possibly empty) are attached. Equivalently, we could consider complete binary trees and only take internal nodes into account.

A (minor) difference to the two previous families of trees is that the main enumeration parameter is now the number of nodes rather than the number of edges. The number of trees with n nodes is now $C_n = \frac{1}{n+1}\binom{2n}{n}$ (compare

[8] The function $\nu(t,u)$ is also algebraic of degree 2 over $\mathbb{Q}(Z(t),Z(tu))$, of degree 4 over $\mathbb{Q}(u,Z)$, and of degree 16 over $\mathbb{Q}(t,u)$ (see [23]).

with Section 3.1.1). For notational convenience we will again use the notation $T(t)$ for the generating function $T(t) = \left(1 - \sqrt{1-4t}\right)/(2t)$ of binary trees.

As explained in Section 5.2.2 we embed binary trees into the plane according to the position of the nodes (compare also with Figure 5.12). The position or natural label of a node is the number of right steps minus the number of left steps on the path from the root to this node

Let $T_j(t)$ be the generating function of (naturally labelled) binary trees in which all labels are at most j, counted by their number of nodes. Let $S_j(t,u)$ be the generating function of binary trees, counted by the number of nodes (variable t) and the number of nodes labelled j (variable u). Finally, let $R_j(t,u)$ be the generating function of binary trees, counted by the number of nodes and the number of nodes having label j at least. Again it is easy to write an infinite system of equations defining any of the families $T_j(t)$, $S_j(t,u)$ or $R_j(t,u)$. The decomposition of trees that was crucial in Section 5.3.1 is now replaced by the decomposition explained in Section 2.1. The generating function $T(t)$ counting naturally labelled binary trees satisfies

$$T(t) = 1 + tT(t)^2,$$

while for $j \ge 0$,

$$T_j(t) = 1 + tT_{j-1}(t)T_{j+1}(t). \tag{5.38}$$

The initial condition is now $T_{-1}(t) = 1$ (accounting for the empty tree).

The equations of Lemmas 5.12 and 5.14 respectively become:

$$S_j(t,u) = \begin{cases} 1 + tS_{j-1}(t,u)S_{j+1}(t,u) & \text{if } j \ne 0, \\ 1 + tuS_1(t,u)^2 & \text{if } j = 0, \end{cases} \tag{5.39}$$

while

$$R_j(t,u) = 1 + tR_{j-1}(t,u)R_{j+1}(t,u) \quad \text{for } j \ge 1, \tag{5.40}$$

and

$$R_{-j}(t,u) = R_{j+1}(tu, 1/u) \quad \text{for all } j \in \mathbb{Z}. \tag{5.41}$$

These three infinite systems of equations can be solved using the same techniques as in Section 5.3.1. The solutions are expressed in terms of the above series $T(t)$ and of the unique formal power series $Z(t)$, with constant term 0, satisfying

$$Z(t) + \frac{1}{Z(t)} - 1 = \frac{1}{tT(t)^2}. \tag{5.42}$$

Equivalently we have

$$Z(t) = t\,\frac{\left(1 + Z(t)^2\right)^2}{1 - Z(t) + Z(t)^2} \tag{5.43}$$

or

$$T(t) = \frac{1 + Z(t)^2}{1 - Z(t) + Z(t)^2}.$$

Again we state without proof the counterparts of Propositions 5.10, 5.13 and 5.15.

Proposition 5.19. *Let $T_j(t)$ be the generating function of (naturally embedded) binary trees having no label greater than j. Then we have for all $j \geq -1$,*

$$T_j(t) = T(t)\,\frac{(1 - Z(t)^{j+2})(1 - Z(t)^{j+7})}{(1 - Z(t)^{j+4})(1 - Z(t)^{j+5})},$$

where $Z(t)$ is given by (5.43). These functions are algebraic of degree (at most) 2. In particular, $T_0(t)$ is given by

$$T_0(t) = \frac{(1-4t)^{3/2} - 1 + 8t - 2t^2}{2t(1+t)}.$$

Recall that by symmetry we only have to discuss the case $j \geq 0$ in the next two propositions.

Proposition 5.20. *For any $j \geq 0$, the generating function $S_j(t,u)$ that counts (naturally embedded) binary trees by the number of nodes and the number of nodes labelled j is given by*

$$S_j(t,u) = T(t)\,\frac{(1 + \mu(t,u)Z(t)^j)(1 + \mu(t,u)Z(t)^{j+5})}{(1 + \mu(t,u)Z(t)^{j+2})(1 + \mu(t,u)Z(t)^{j+3})},$$

where $Z(t)$ is given by (5.43) and $\mu(t,u)$ is the unique formal power series in t satisfying

$$\mu(t,u) = (u-1)\frac{Z(t)(1 + \mu(t,u)Z(t))^2(1 + \mu(t,u)Z(t)^2)(1 + \mu(t,u)Z(t)^6)}{(1 + Z(t))^2(1 + Z(t) + Z(t)^2)(1 - Z(t))^3(1 - \mu(t,u)^2 Z(t)^5)}.$$

The series $\mu(t,u)$ has polynomial coefficients in u and satisfies $\mu(t,1) = 0$. In particular, $S_0 = S_0(t,u)$ satisfies the equation

$$\begin{aligned}\frac{T(t)^2(u-1)^2}{u(T(t) - S_0)^2} &= \frac{(T(t)-1)^4 S_0)^2 - 2T(t)S_0(T(t)-1)^2(3 - 9T(t) + 7T(t)^2)}{(T(t)-1)(S_0-1)(T(t)^2 + T(t)S_0) - S_0)^2} \\ &\quad + \frac{T(t)^2(T(t)^2 + T(t) - 1)^2}{(T(t)-1)(S_0-1)(T(t)^2 + T(t)S_0) - S_0)^2}.\end{aligned}$$

Proposition 5.21. *Let $j \geq 0$. The generating function $R_j(t,u)$ that counts (naturally embedded) binary trees by the number of nodes and the number of nodes labelled j or more is given by*

$$R_j(t,u) = T(t)\,\frac{\left(1 + \nu(t,u)Z(t)^j\right)\left(1 + \nu(t,u)Z(t)^{j+5}\right)}{(1 + \nu(t,u)Z(t)^{j+2})\,(1 + \nu(t,u)Z(t)^{j+3})},$$

where $Z(t)$ is given by (5.43) and $\nu(t,u)$ is a formal power series in t, with polynomial coefficients in u. This series satisfies $\nu(t,1) = 0$ and the first terms in its expansion are

$$\nu(t,u) = (u-1)\Big(t + (u+1)t^2 + (2u^2 + 3u + 3)t^3 + O(t^4)\Big).$$

5.4 Asymptotics on Embedded Trees

By Theorems 5.4–5.6 we already know the limiting behaviour of the vertical profile of embedded trees and also of naturally embedded binary trees. In what follows we show how the explicit formulas for the generating functions given in Section 5.3 can be used to derive one dimensional versions[9] of these properties. By this procedure we do not get the full result. On the other hand we get error terms and convergence of moments, too. We again follow Bousquet-Mélou [23].

Since the generating functions for the three classes of trees presented in Section 5.3 are principally very close to each other, we will only discuss embedded trees with increments ± 1 in detail.

5.4.1 Trees with Small Labels

Let $\mathcal{T}_0$ denote the set of well-labelled trees with increments ± 1, and let $\mathcal{T}_{0,n}$ denote the subset of $\mathcal{T}_0$ formed by trees having n edges. We endow $\mathcal{T}_{0,n}$ with the uniform distribution. In other words, any of its elements occurs with probability

$$\frac{1}{\frac{2^n}{n+1}\binom{2n}{n}}.$$

Let M_n denote the random variable equal to the largest label occurring in a random tree of $\mathcal{T}_{0,n}$. By definition the distribution of M_n is related to the series $T_j(t)$ studied in Proposition 5.10:

$$\mathbb{P}\left\{M_n \le j\right\} = \frac{[t^n]T_j(t)}{\frac{2^n}{n+1}\binom{2n}{n}}.$$

Theorem 5.22. *Let*

$$N_{\mathrm{ISE}} = \sup\{y : \mu_{\mathrm{ISE}}((y,\infty)) > 0\}.$$

Then, as $n \to \infty$*,*

$$\frac{M_n}{n^{1/4}} \xrightarrow{d} N_{\mathrm{ISE}}.$$

Furthermore, the moments of $M_n/n^{1/4}$ *converge to the moments of* N_{ISE}.

Set $N_n = M_n/n^{1/4}$ and let $\lambda \ge 0$ and $j = \lfloor \lambda n^{1/4} \rfloor$. We are interested in the probability

$$\mathbb{P}\{N_n > \lambda\} = \mathbb{P}\{M_n > \lambda n^{1/4}\} = \mathbb{P}\{M_n > j\} = \frac{[t^n]U_j(t)}{\frac{2^n}{n+1}\binom{2n}{n}}, \tag{5.44}$$

[9] In this context *one dimensional* means a one dimensional projection $X_n(t_0) \xrightarrow{d} X(t_0)$ of a functional limit theorem $(X_n(t), -\infty < t < \infty) \xrightarrow{d} (X(t), -\infty < t < \infty)$.

where

$$U_j(t) = T(t) - T_j(t) \tag{5.45}$$
$$= \frac{(1+Z(t))^2 Z(t)^{j+1} \left(1+Z(t)+Z(t)^2\right)(1-Z(t))^2}{\left(1+Z(t)^2\right)\left(1-Z(t)^{j+2}\right)\left(1-Z(t)^{j+4}\right)}$$

is the generating function of trees having at least one label greater than j. This algebraic series has a positive radius of convergence, and by Cauchy's formula

$$[t^n]U_j(t) = \frac{1}{2i\pi}\int_\gamma U_j(t)\frac{dt}{t^{n+1}}$$
$$= \frac{1}{2i\pi}\int_\gamma \frac{(1+Z)^2 Z^{j+1}\left(1+Z+Z^2\right)(1-Z)^2}{(1+Z^2)(1-Z^{j+2})(1-Z^{j+4})}\frac{dt}{t^{n+1}} \tag{5.46}$$

for any contour γ included in the analyticity domain of $U_j(t)$ and enclosing positively the origin.

This leads us to study the singularities of $U_j(t)$, and therefore those of $Z(t)$. The following lemma provides a complete picture of the (surprisingly simple) singular structure of $Z(t)$.

Lemma 5.23. *Let $Z(t)$ be the unique formal power series in t with constant term 0 satisfying (5.18). This series has non-negative integer coefficients. It has radius of convergence $1/8$, and can be continued analytically on the domain $D = \mathbb{C} \setminus [1/8, +\infty)$. In the neighbourhood of $t = 1/8$, one has*

$$Z(t) = 1 - 2(1-8t)^{1/4} + O(\sqrt{1-8t}). \tag{5.47}$$

Moreover, $|Z(t)| < 1$ on the domain D. More precisely, the only roots of unity that are accumulation points of the set $Z(D)$ are 1 and -1, and they are only approached by $Z(t)$ when t tends to $1/8$ and when $|t|$ tends to ∞, respectively.

Proof. In order to establish the first statement, we observe that

$$Z(t) = W(t)(1+Z(t))^2$$

where $W(t)$ is the only formal power series in t with constant term zero satisfying

$$W(t) = t + 2W(t)^2. \tag{5.48}$$

These equations imply that both $W(t)$ and $Z(t)$ have non-negative integer coefficients.

The polynomial equation defining $Z(t)$ has the leading coefficient t and the discriminant $4(1-8t)^3$, so that the only possible singularity of $Z(t)$ is $1/8$. Alternatively, one can exploit the following closed form expression

$$Z(t) = \frac{\sqrt{1-4t+\sqrt{1-8t}}\left(\sqrt{1-4t+\sqrt{1-8t}} - \sqrt{2}(1-8t)^{1/4}\right)}{4t}, \quad (5.49)$$

which can also be used to obtain (5.47).

Next we prove that $|Z(t)|$ never reaches 1 on the domain D. Assume $Z(t) = e^{i\vartheta}$, with $\vartheta \in [-\pi, \pi]$. From (5.18), one has

$$t = t_\vartheta \quad \text{where} \quad t_\vartheta = \frac{\cos\vartheta}{8\cos^4(\vartheta/2)} \quad \text{and} \quad \vartheta \in (-\pi, \pi).$$

This shows that t is real and belongs to $(-\infty, 1/8)$. But the expression (5.49) of $Z(t)$ shows that $Z(t)$ is real, which contradicts the hypothesis $Z(t) = e^{i\vartheta}$, unless $\vartheta = 0$. But then $t = 1/8$ which does not belong to the domain D. Hence the modulus of Z never reaches 1 on D.

Finally, if a sequence t_n of D is such that $Z(t_n) \to e^{i\vartheta}$ as $n \to \infty$, with $\vartheta \in (-\pi, \pi]$, then either $\vartheta = \pi$ and, by (5.18), the sequence $|t_n|$ tends to ∞ or $\vartheta \in (-\pi, \pi)$ and t_n converges to t_ϑ. But then by continuity, $Z(t_n)$ actually converges to $Z(t_\vartheta)$, which, as argued above, only coincides with $e^{i\vartheta}$ when $\vartheta = 0$, that is, $t_\vartheta = 1/8$. In this case, $Z(t_n) \to 1$.

Let us now go back to the evaluation of the tail distribution function of $N_n = M_n/n^{1/4}$ via the integral (5.46). We use the Hankel like contour $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, where

$$\gamma_1 = \left\{ t = \frac{1}{8}\left(1 + \frac{-i + (\log n)^2 - w}{n}\right) : 0 \le w \le (\log n)^2 \right\},$$

$$\gamma_2 = \left\{ t = \frac{1}{8}\left(1 - \frac{1}{n}e^{-i\varphi}\right) : -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2} \right\},$$

$$\gamma_3 = \left\{ t = \frac{1}{8}\left(1 + \frac{i + w}{n}\right) : 0 \le w \le (\log n)^2 \right\},$$

and γ_4 is a circular arc centred at the origin and making γ a closed curve, that is, if we set $t = \frac{1}{8}\left(1 + \frac{z}{n}\right)$ then for $t \in \gamma_1 \cup \gamma_2 \cup \gamma_3$ the (new) variable z varies over a finite section H_n of a Hankel contour H (see Figure 2.5).

First observe that for $t \in \gamma$

$$1 - |Z(t)| \ge \frac{1}{2}n^{-1/4}, \quad (5.50)$$

and consequently for all $j \ge 1$

$$|1 - Z(t)^j| \ge 1 - |Z(t)|^j \ge 1 - |Z(t)| \ge \frac{1}{2}n^{-1/4}.$$

By Lemma 5.23, the quantity

$$\frac{(1 + Z(t))^2 Z(t)^{j+1}\left(1 + Z(t) + Z(t)^2\right)(1 - Z(t))^2}{1 + Z(t)^2}$$

is uniformly bounded on γ. Therefore the contribution of the integral over γ_4 is bounded by

$$\int_{\gamma_4} U_j(t)\frac{dt}{t^{n+1}} = O\left(8^n n^{1/2} e^{-(\log n)^2}\right). \tag{5.51}$$

The dominant part of the integral comes from the Hankel-like contour $\gamma_1 \cup \gamma_2 \cup \gamma_3$. Recall that $t = \frac{1}{8}\left(1+\frac{z}{n}\right)$ and that $|z| \le \log^2 n$. We have the following approximations as $n \to \infty$ and $j = \lfloor \lambda n^{1/4} \rfloor$:

$$\begin{aligned}
Z(t) &= 1 - 2(-z)^{1/4} n^{-1/4} + O\left(n^{-1/2}\log n\right) \\
1 - Z(t) &= 2(-z)^{1/4} n^{-1/4}\left(1 + O(n^{-1/4}\sqrt{\log n})\right) \\
Z(t)^j &= \exp(-2\lambda(-z)^{1/4})\left(1 + O(n^{-1/4}\log n)\right) \\
t^{-n-1} &= 8^{n+1} e^{-z}\left(1 + O((\log n)^4/n)\right).
\end{aligned}$$

Observe that, for $z \in H_n$, the real part of $(-z)^{1/4}$ is bounded from below by a positive constant c_0. Hence,

$$|\exp(-2\lambda(-z)^{1/4})| = \exp(-2\lambda\Re(-z)^{1/4}) \le \exp(-2\lambda c_0),$$

so that $\exp(-2\lambda(-z)^{1/4})$ does not approach 1. This allows us to write

$$\frac{1}{1-Z^{j+2}} = \frac{1}{1-\exp(-2\lambda(-z)^{1/4})}\left(1 + O(n^{-1/4}\log n)\right).$$

Hence, uniformly in $t \in \gamma_1 \cup \gamma_2 \cup \gamma_3$, we have

$$\begin{aligned}
U_j(t)t^{-n-1} &= \frac{(1+Z)^2 Z^{j+1}\left(1+Z+Z^2\right)(1-Z)^2}{(1+Z^2)(1-Z^{j+2})(1-Z^{j+4})} t^{-n-1} \\
&= \frac{6\cdot 8^{n+1}}{n^{1/2}} \frac{\sqrt{-z}e^{-z}}{\sinh^2(\lambda(-z)^{1/4})}(1 + O(n^{-1/4}\log n))
\end{aligned}$$

and consequently

$$\begin{aligned}
\int_{\gamma_1\cup\gamma_2\cup\gamma_3} U_j(t)\frac{dt}{t^{n+1}} &= \frac{6\cdot 8^n}{n^{3/2}} \int_{H_n} \frac{\sqrt{-z}e^{-z}(1+O(n^{-1/4}\log n))}{\sinh^2(\lambda(-z)^{1/4})} dz \\
&= \frac{6\cdot 8^n}{n^{3/2}}\left(\int_H \frac{\sqrt{-z}e^{-z}}{\sinh^2(\lambda(-z)^{1/4})}dz + o(1)\right).
\end{aligned}$$

Hence,

$$[t^n]U_j(t) = \frac{6\cdot 8^n n^{-3/2}}{2i\pi}\left(\int_H \frac{\sqrt{-z}e^{-z}}{\sinh^2(\lambda(-z)^{1/4})}dz + o(1)\right)$$

and by using the estimate $\frac{1}{n+1}\binom{2n}{n} \sim 4^n n^{-3/2}/\sqrt{\pi}$, this gives

$$\mathbb{P}\{N_n > \lambda\} \to \frac{3}{i\sqrt{\pi}} \int_H \frac{\sqrt{-z}e^{-z}}{\sinh^2(\lambda(-z)^{1/4})} dz = G(\lambda).$$

If we use the substitution $v = (-z)^{1/4}$ in the above integral, the variable v runs to a contour that is close to the contour Γ that is defined by (5.13). Due to analyticity we can shift the line of integration to Γ and also obtain

$$G(\lambda) = \frac{12}{i\sqrt{\pi}} \int_\Gamma \frac{v^5 e^{v^4}}{\sinh^2(\lambda v)}\, dv.$$

Thus we have proved weak convergence of N_n to some random variable $\tilde{N}$ which has tail distribution function $G(\lambda)$. By using the property that the discrete Brownian snake process (which is precisely the process of traversing an embedded tree and looking at the labels) converges (after scaling) to the Brownian snake it also follows that the maximum label converges (after scaling) to N_{ISE}. Hence, $\tilde{N} \overset{\mathrm{d}}{=} N_{\mathrm{ISE}}$. This completes the first part of the proof of Theorem 5.22.

Now recall that the series $U_j(t)$, given by (5.45), counts the trees that contain at least one label larger than j. Hence, $U_{j-1}(t) - U_j(t)$ counts the trees having maximal label j. Also, note that

$$U_j(t) = V(Z(t)^j, t) - V(Z(t)^{j+2}, t), \tag{5.52}$$

where

$$V(x,t) = \frac{xZ(t)(1+Z(t))(1-Z(t)^3)}{(1+Z(t)^2)(1-xZ(t)^2)}.$$

Consequently, for $k \ge 1$,

$$\begin{aligned} \mathbb{E}\, M_n^k &= \frac{1}{\frac{2^n}{n+1}\binom{2n}{n}} \sum_{j\ge 1} j^k [t^n](U_{j-1}(t) - U_j(t)) \\ &= \frac{1}{\frac{2^n}{n+1}\binom{2n}{n}} [t^n] \sum_{j \ge 0} \left((j+1)^k - j^k\right) U_j(t). \end{aligned} \tag{5.53}$$

For example, for $k = 1$ this gives

$$\begin{aligned} \mathbb{E}\, M_n &= \frac{1}{\frac{2^n}{n+1}\binom{2n}{n}} [t^n] \sum_{j\ge 0} \left(V(Z(t)^j, t) - V(Z(t)^{j+2}, t)\right) \\ &= \frac{1}{\frac{2^n}{n+1}\binom{2n}{n}} [t^n] \left(V(1,t) + V(Z(t),t)\right) \\ &= \frac{1}{\frac{2^n}{n+1}\binom{2n}{n}} [t^n] \frac{Z(t)(1+2Z(t)+2Z(t)^2)}{1+Z(t)^2}. \end{aligned}$$

Since

$$\frac{Z(t)(1+2Z(t)+2Z(t)^2)}{1+Z(t)^2} = \frac{5}{6} - 6(1-8t)^{1/4} + O(\sqrt{1-8t}),$$

we have

$$[t^n]\frac{Z(t)(1+2Z(t)+2Z(t)^2)}{1+Z(t)^2} \sim -6\frac{8^n n^{-5/4}}{\Gamma(-1/4)} = \frac{3}{2}\frac{8^n n^{-5/4}}{\Gamma(3/4)}$$

and consequently

$$\mathbb{E}\, M_n \sim \frac{3\sqrt{\pi}}{2\Gamma(3/4)}\, n^{1/4}$$

which corresponds to the first moment of N_{ISE}.

Now, by combining the expression (5.52) of $U_j(t)$ and (5.53), one obtains, for $k \ge 2$,

$$\sum_{j\ge 0} \big((j+1)^k - j^k\big)U_j(t) = V(1,t) + (2^k-1)V(Z(t),t) \qquad (5.54)$$

$$+ \sum_{j\ge 2} \big((j+1)^k - j^k - (j-1)^k + (j-2)^k\big)\, V(Z(t)^j,t).$$

Observe that

$$\begin{aligned}(j+1)^k - j^k - (j-1)^k + (j-2)^k &\sim 2k(k-1)j^{k-2} \\ &\sim 2k(k-1)(j+2)^{k-2}.\end{aligned}$$

Hence, the dominating part of this sum will be of the form

$$\begin{aligned} A_k(z) &= 2k(k-1)\sum_{j\ge -1}(j+2)V(Z(t)^j,t) \\ &= 2k(k-1)\frac{(1+Z(t))(1-Z(t)^3)}{Z(t)(1+Z(t)^2)}\sum_{j\ge -1}(j+2)^{k-2}\frac{Z(t)^{j+2}}{1-Z(t)^{j+2}} \\ &= 2k(k-1)\frac{(1+Z(t))(1-Z(t)^3)}{Z(t)(1+Z(t)^2)}D_{k-2}(Z(t)), \end{aligned}$$

where $D_R(z) = \sum_{k\ge 1} k^R z^k/(1-z^k)$ was discussed in Lemma 4.37. In particular it follows that

$$D_{k-2}(Z(t)) \sim \begin{cases} \dfrac{1}{8(1-8t)^{1/4}}\log\dfrac{1}{1-8t} & \text{for } k=2, \\ \dfrac{(k-2)!\zeta(k-1)}{2^{k-1}(1-8t)^{\frac{k-1}{2}}} & \text{for } k>2, \end{cases}$$

and consequently

$$A_k(t) \sim \begin{cases} 3\log\dfrac{1}{1-8t} & \text{for } k=2, \\ \dfrac{24\, k!\zeta(k-1)}{2^{k-1}(1-8t)^{\frac{k-2}{4}}} & \text{for } k>2. \end{cases}$$

Hence, after applying Lemma 2.14 and normalising by $\frac{2^n}{n+1}\binom{2n}{n}$, this gives, as expected:

$$\mathbb{E}\, M_n^k \sim \begin{cases} 3\sqrt{\pi}\, n^{1/4} & \text{if } k = 2, \\ \dfrac{24\sqrt{\pi}k!\zeta(k-1)}{2^k\Gamma((k-2)/4)}\, n^{k/4} & \text{if } k \geq 3. \end{cases}$$

This completes the proof of Theorem 5.22.

5.4.2 The Number of Nodes of Given Label

Next let $X_n(j)$ denote the (random) number of nodes having label j in a random tree of $\mathcal{T}_{0,n}$. This quantity is related to the series $S_j(t,u)$ studied in Proposition 5.13. In particular,

$$\mathbb{E}\left(e^{aX_n(j)}\right) = \frac{[t^n]S_j(t,e^a)}{\frac{2^n}{n+1}\binom{2n}{n}}.$$

Let us denote the normalised version of $X_n(j)$ by

$$Y_n(j) = \frac{X_n(j)}{n^{3/4}}.$$

The aim of this section is to prove a weak limit theorem for $Y_n(\lfloor \lambda n^{1/4} \rfloor)$, where $\lambda \in \mathbb{R}$ is arbitrary but fixed. Actually this is the one dimensional version of Theorem 5.5 applied to planted plane trees and η uniformly distributed on $\{-1,1\}$.

Theorem 5.24. *For every $\lambda \in \mathbb{R}$ and for all $|a| < 4/\sqrt{3}$ we have*

$$\mathbb{E}\, e^{aY_n(\lfloor \lambda n^{1/4} \rfloor)} \to \mathbb{E}\, e^{a\sqrt{2}f_{\mathrm{ISE}}(\sqrt{2}\,\lambda)}.$$

Hence,

$$Y_n(\lfloor \lambda n^{1/4} \rfloor) \xrightarrow{d} \sqrt{2} f_{\mathrm{ISE}}(\sqrt{2}\,\lambda).$$

and all moments of Y_n converge to the corresponding moments of f_{ISE}.

An interesting special case of Theorem 5.24 is $\lambda = 0$. Here $X_n(0)$ is the number of nodes labelled 0 in a tree rooted at 0 and we obtain

$$\frac{3X_n(0)}{\sqrt{2}n^{3/4}} \xrightarrow{d} T^{-1/2},$$

where T follows a unilateral stable law of parameter $2/3$ (that can be defined by its Laplace transform $\mathbb{E}(e^{-aT}) = e^{-a^{2/3}}$ for $a \geq 0$).

In what follows we will show that (for $|a| < 4/\sqrt{3}$)

$$\mathbb{E}\, e^{aY_n(\lfloor \lambda n^{1/4} \rfloor)} \to 1 + \frac{48}{i\sqrt{\pi}} \int_\Gamma \frac{A(a/v^3)e^{-2\lambda v}}{(1+A(a/v^3)e^{-2\lambda v})^2} v^5 e^{v^4}\, dv, \tag{5.55}$$

where $A(x)$ is the unique solution of

$$A = \frac{x}{24} \frac{(1+A)^3}{1-A},$$

satisfying $A(0) = 0$, and the integral is taken over the contour Γ given by (5.13). By Theorem 5.5 we already know that $Y_n(\lfloor \lambda n^{1/4} \rfloor) \xrightarrow{d} \sqrt{2} f_{\mathrm{ISE}}(\sqrt{2}\,\lambda)$. Thus, the Laplace transform of $\sqrt{2} f_{\mathrm{ISE}}(\sqrt{2}\,\lambda)$ is given by right hand side in (5.55). Moreover, convergence of Laplace transforms implies not only convergence in distribution but also convergence of all moments (see [71, Thm. 9.8.2]). This proves Theorem 5.24. Thus, we just have to concentrate on the limit relation (5.55).

Note that the series $A(x)$ admits the following closed form expression:

$$A(x) = \frac{2}{1 + \frac{2}{\sqrt{3}} \cos\left(\frac{\arccos(-x\sqrt{3}/4)}{3}\right)} - 1. \tag{5.56}$$

This can be checked by proving that this expression satisfies the defining relations for $A(x)$.

Let $\lambda \ge 0$ and $j = \lfloor \lambda n^{1/4} \rfloor$. The Laplace transform of $Y_n(j)$ is related to the generating functions $S_j(t,u)$ of Proposition 5.13:

$$\mathbb{E}\left(e^{aY_n(j)}\right) = \mathbb{E}\left(e^{an^{-3/4} X_n(j)}\right) = \frac{[t^n] S_j(t, e^{an^{-3/4}})}{\frac{2^n}{n+1}\binom{2n}{n}}. \tag{5.57}$$

We will evaluate this Laplace transform by contour integration over $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, the same contour that we used in the proof of Theorem 5.22. This requires to prove that $S_j(t,u)$ is analytic in a neighbourhood of this contour (for n large and $u = e^{an^{-3/4}}$) which is guaranteed by the following lemma. For convenience we denote by I_n the part of the complex plane enclosed by γ (including γ itself; note that γ depends on n).

Lemma 5.25. *Let a be a real number such that $|a| < 4/\sqrt{3}$. Then there exists $\varepsilon > 0$ such that for n large enough the series $\mu(t, u_n)$ with $u_n = e^{an^{-3/4}}$ is analytic in the domain*

$$E_n = \{t : |t - 1/8| > 1/((8+\varepsilon)n)\} \setminus [1/8, +\infty).$$

In particular, $\mu(t,u_n)$ is analytic in a neighbourhood of I_n. Its modulus in I_n is smaller than α, for some $\alpha < 1$ independent of a and n. The series $S_j(t,u_n)$ is also analytic in a neighbourhood of I_n.

Proof. The lemma is clear, if $a = 0$; in this case $u_n = 1$, the series $\mu(t,u_n)$ vanishes, and the series $S_j(t,u)$ reduces to the size generating function of labelled trees, namely $T(t)$, which is analytic in $\mathbb{C} \setminus [1/8, \infty)$. We now assume that $a \ne 0$ and $|a| < 4/\sqrt{3}$. This guarantees that $A(a)$ is well-defined.

Let us first study the singularities of the series $\bar{\mu} = \bar{\mu}(z, u)$ defined as the unique formal power series in z, satisfying

$$\bar{\mu} = (u-1)\frac{(1+z^2)(1+\bar{\mu}z)(1+\bar{\mu}z^2)(1+\bar{\mu}z^3)}{(1+z)(1+z+z^2)(1-z)^3(1-\bar{\mu}z^2)}.$$

Note that $\bar{\mu}$ has polynomial coefficients in u, and vanishes when $u = 1$. Assume that u is a fixed real number close to, but different from, 1. Recall that, as all algebraic formal power series, $\bar{\mu}(t, u)$ has a positive radius of convergence. We perform a classical analysis to detect its possible singularities. These singularities are found in the union of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$:

1. $\mathcal{S}_1$ is the set of non-zero roots of the dominant coefficient of the equation defining $\bar{\mu}$, that is, $\mathcal{S}_1 = \{\pm i\}$.
2. $\mathcal{S}_2$ is the set of the roots of the discriminant of the equation defining $\bar{\mu}$. For $u = 1 + x$ and x small, these roots are found to be

   $$z = \pm 1, \qquad z = -1 + O(x),$$
   $$z = e^{\pm 2i\pi/3} + O(x), \qquad z = 1 + \omega 12^{1/6} x^{1/3} + O(|x|^{2/3}),$$

   where ω satisfies $\omega^6 = 1$. (The term ω allows us to write $x^{1/3}$ without saying which of the cubic root we take.)
   Observe that the moduli of all these "candidates for singularities" go to 1 as x goes to 0.

Now the series $\mu = \mu(t, u)$ involved in the expression (5.23) of $S_j(t, u)$ satisfies

$$\mu(t, u) = \bar{\mu}(Z(t), u)$$

where $Z(t)$ is defined by (5.18). In other words, we could have defined the series $\bar{\mu}$ by

$$\bar{\mu}(z, u) = \mu\left(\frac{z(1+z^2)}{(1+z)^4}, u\right).$$

Recall that $Z(t)$ is analytic in the domain $D = \mathbb{C} \setminus [1/8, \infty)$. Take $u = u_n = e^{an^{-3/4}} = 1 + x$, with $x = an^{-3/4}(1 + o(1))$. By Lemma 5.23, for n large, the only values of $\mathcal{S}_1 \cup \mathcal{S}_2$ that may be reached by $Z(t)$, for $t \in D$, are of the form

$$z = 1 + \omega 12^{1/6} a^{1/3} n^{-1/4} + O(n^{-1/2}).$$

In view of (5.18), these values of $Z(t)$ are reached for

$$t = \frac{1}{8} - \frac{\omega^4 (12)^{2/3}}{128}\frac{a^{4/3}}{n} + O(n^{-5/4}).$$

Since $|a| < 4/\sqrt{3}$, these values of t are at distance less than $1/((8+\varepsilon)n)$ of $1/8$, for some $\varepsilon > 0$, and hence outside the domain E_n. Consequently, $\mu(t, u_n)$ is analytic inside E_n.

We now want to bound $\mu(t, u_n)$ inside I_n. Let $t_n \in I_n$ be such that

$$|\mu(t_n, u_n)| = \max_{t \in I_n} |\mu(t, u_n)|.$$

In particular, $|\mu(t_n, u_n)| \geq |\mu((1-1/n)/8, u_n)|$. In order to evaluate the latter quantity, note that $Z((1-1/n)/8) = 1-2n^{-1/4}+O(n^{-1/2})$. Due to the closed form expression of μ given in (5.26), and to the expression (5.56) of the series A, we see that $\mu((1-1/n)/8, u_n) \to A(a)$. Since $a \neq 0$, $A(a) \neq 0$, and for n large enough,

$$|\mu(t_n, u_n)| \geq |\mu((1-1/n)/8, u_n)| = |A(a)| + o(1) > 0. \tag{5.58}$$

Recall that all the sets I_n are included in a ball of finite radius centred at the origin. Let α be an accumulation point of the sequence t_n. Then $|\alpha| \leq 1/8$.

Assume first that $\alpha \neq 1/8$. Then there exists N_0 such that α is in E_n for all $n \geq N_0$, that is, in the analyticity domain of $\mu(\cdot, u_n)$. Let $t_{n_1}, t_{n_2}, \ldots$ converge to α. By continuity of μ in t and u, we have

$$\mu(t_{n_i}, u_{n_i}) \to \mu(\alpha, 1) = 0.$$

This contradicts (5.58). Hence the only accumulation point of t_n is 1/8, and t_n converges to 1/8. Let us thus write

$$t_n = \frac{1}{8}\left(1 - \frac{x_n}{n}\right).$$

We have $x_n = o(n)$, but also $|x_n| > 1$, since t_n belongs to I_n. We wish to estimate $\mu(t_n, u_n)$. From the singular behaviour of $Z(t)$ (Lemma 5.23), we derive

$$Z(t_n) = 1 - 2\left(\frac{x_n}{n}\right)^{1/4} + O\left(\left(\frac{x_n}{n}\right)^{1/2}\right).$$

Moreover,

$$u_n - 1 = an^{-3/4}\left(1 + O(n^{-3/4})\right),$$

which gives

$$\frac{u_n - 1}{(1-Z)^3} = \frac{a}{8x_n^{3/4}}\left(1 + O\left(\left(\frac{x_n}{n}\right)^{1/4}\right)\right).$$

If the sequence x_n was unbounded, then there would exist a subsequence x_{n_i} converging to infinity. Then $(u_{n_i} - 1)/(1 - Z(t))^3$ would tend to 0. The closed form expression of μ given in (5.26) implies that $\mu(t_{n_i}, u_{n_i})$ would tend to 0, contradicting (5.58). Hence, the sequence x_n is bounded, and one derives from the explicit expressions of μ and A that

$$\mu(t_n, u_n) = A(ax_n^{-3/4}) + o(1).$$

Since A is bounded by $2 - \sqrt{3}$ inside its disk of convergence, $|\mu(t_n, u_n)|$ is certainly smaller than some α for $\alpha < 1$ and n large enough. This concludes the proof of the second statement of Lemma 5.25.

By continuity of $\mu(t, u_n)$, this function in t is still bounded by 1 (in modulus) in a neighbourhood of I_n. Recall also that the modulus $|Z(t)|$ never reaches 1 for $t \in \mathbb{C} \setminus [1/8, \infty)$. The form (5.23) then implies that $S_j(t, u_n)$ is an analytic function of t in a neighbourhood of I_n.

Let us now go back to the expression (5.57) of the Laplace transform of $Y_n(j)$. We use Cauchy's formula to extract the coefficient of t^n in $S_j(t, u_n)$. From (5.23) we obtain the representation

$$S_j(t,u) = T(t) + T(t)\frac{(1-Z(t))^2(1+Z(t)+Z(t)^2)\mu(t,u)Z(t)^j}{(1+\mu(t,u)Z(t)^{j+1})(1+\mu(t,u)Z(t)^{j+3})}.$$

Thus

$$[t^n]S_j(t,u_n) = \frac{2^n}{n+1}\binom{2n}{n} + \frac{1}{2i\pi}\int_\gamma T\frac{(1-Z)^2(1+Z+Z^2)\mu Z^j}{(1+\mu Z^{j+1})(1+\mu Z^{j+3})}\frac{dt}{t^{n+1}}.$$

We split the contour γ into two parts $\gamma_1 \cup \gamma_2 \cup \gamma_3$ and γ_4. As in the proof of Theorem 5.22, the contribution of γ_4 is easily seen to be $o(8^n/n^m)$ for all $m > 0$, thanks to the results of Lemmas 5.23 and 5.25. On $\gamma_1 \cup \gamma_2 \cup \gamma_3$, one has

$$t = \frac{1}{8}\left(1 + \frac{z}{n}\right)$$

where z lies in the truncated Hankel contour H_n. Conversely, let $z \in H$. Then $z \in H_n$ for n large enough, and, in addition to the estimates already used in the proof of Theorem 5.22, one finds

$$\mu(t,u_n) = A(a(-z)^{-3/4})(1+o(1)), \tag{5.59}$$

where $A(x)$ is the series defined by (5.15). After a few reductions, one finally obtains

$$\begin{aligned}[t^n]S_j(t,u_n) &= \frac{2^n}{n+1}\binom{2n}{n} \\ &+ \frac{12\cdot 8^n n^{-3/2}}{i\pi}\int_H \frac{A(a(-z)^{-3/4})\exp(-2\lambda(-z)^{1/4})\sqrt{-z}e^{-z}}{(1+A(a(-z)^{-3/4})\exp(-2\lambda(-z)^{1/4}))^2}dz + o(8^n n^{-3/2}).\end{aligned}$$

It remains to normalise by $\frac{2^n}{n+1}\binom{2n}{n} \sim 8^n n^{-3/2}/\sqrt{\pi}$, and then to set $v = (-z)^{1/4}$ to obtain the expected expression for the limit of the Laplace transform of $Y_n(j)$, with $j = \lfloor \lambda n^{1/4} \rfloor$. This completes the proof of Theorem 5.24.

5.4.3 The Number of Nodes of Large Labels

Let $X_n^+(j)$ denote the (random) number of nodes having label at least j in a random tree of $\mathcal{T}_{0,n}$. We denote the normalised version of $X_n^+(j)$ by

$$Y_n^+(j) = \frac{X_n^+(j)}{n}.$$

These quantities are related to the series $R_j(t,u)$ studied in Proposition 5.15. In particular,

$$\mathbb{E}\left(e^{aY_n^+(j)}\right) = \mathbb{E}\left(e^{an^{-1}X_n^+(j)}\right) = \frac{[t^n]R_j(t,e^{a/n})}{\frac{2^n}{n+1}\binom{2n}{n}}.$$

In the present context it is convenient to extend the definition of X_n^+ and Y_n^+ to real values by setting $X_n^+(t) = X_n^+(\lceil t\rceil)$ and $Y_n^+(t) = Y_n^+(\lceil t\rceil)$. Our next aim is to prove that $Y_n^+(\lambda n^{1/4})$ (for some $\lambda \ge 0$) converges weakly and in terms of all moments to the (random) tail distribution function

$$G^+(\lambda) = \mu_{\mathrm{ISE}}((\lambda,\infty))$$

of the ISE. This is the one dimensional version of Theorem 5.4 applied to planted plane trees with increments η that are uniformly distributed on $\{-1,1\}$.

Theorem 5.26. *For every $\lambda \in \mathbb{R}$ and for all $|a|<1$ we have*

$$\mathbb{E}\, e^{aY_n^+(\lfloor \lambda n^{1/4}\rfloor)} \to \mathbb{E}\, e^{aG^+(\lambda/\sqrt{2})}.$$

Hence,

$$Y_n^+(\lfloor \lambda n^{1/4}\rfloor) \xrightarrow{d} G^+(\lambda/\sqrt{2}).$$

and all moments of Y_n^+ converge to the corresponding moments of G^+.

One interesting consequence of Theorem 5.26 is the limit law for $X_n^+(0)$, the number of nodes having a non-negative label in a tree rooted at 0:

$$\frac{X_n^+(0)}{n} \xrightarrow{d} U,$$

where U is uniformly distributed on $[0,1]$.

Similarly to the proof of Theorem 5.24 we just show that

$$\mathbb{E}\, e^{aY_n^+(\lfloor \lambda n^{1/4}\rfloor)} \to 1 + \frac{48}{i\sqrt{\pi}} \int_\Gamma \frac{B(a/v^4)e^{-2\lambda v}}{(1+B(a/v^4)e^{-2\lambda v})^2} v^5 e^{v^4}\, dv,$$

where

$$B(x) = -\frac{(1-D)(1-2D)}{(1+D)(1+2D)}, \qquad D = \sqrt{\frac{1+\sqrt{1-x}}{2}}.$$

By Theorem 5.4 the limiting distribution of $Y_n^+(\lfloor \lambda n^{1/4}\rfloor)$ coincides with G^+.

For the proof of Theorem 5.26 let $j = \lceil \lambda n^{1/4}\rceil$. Since the product forms of the series $S_j(t,u)$, and $R_j(t,u)$ are very similar, it is not surprising that the analysis is similar, too. We start from

$$\mathbb{E}(e^{aY_n^+(j)}) = \mathbb{E}(u_n^{X_n^+(j)}) = \frac{[t^n]R_j(t,u_n)}{\frac{2^n}{n+1}\binom{2n}{n}},$$

with $u_n = e^{a/n}$. For technical reasons, we choose to modify slightly the integration contour γ to $\overline{\gamma} = \overline{\gamma}_1 \cup \overline{\gamma}_2 \cup \overline{\gamma}_3 \cup \overline{\gamma}_4$ given by

$$\overline{\gamma}_1 = \left\{ t = \frac{1}{8}\left(1 + \frac{4}{3}\,\frac{-i + (\log n)^2 - w}{n}\right) : 0 \le w \le (\log n)^2 \right\},$$
$$\overline{\gamma}_2 = \left\{ t = \frac{1}{8}\left(1 - \frac{4}{3}\,\frac{1}{n}e^{-i\varphi}\right) : -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2} \right\},$$
$$\overline{\gamma}_3 = \left\{ t = \frac{1}{8}\left(1 + \frac{4}{3}\,\frac{i + w}{n}\right) : 0 \le w \le (\log n)^2 \right\},$$

and $\overline{\gamma}_4$ is a circular arc centred at the origin and making γ a closed curve. Informally, we are now considering a Hankel contour centred around $1/8$ at distance $1/(6n)$ to the real axis.

We first need to prove that the series $R_j(t,u_n)$ is analytic in a neighbourhood of $\overline{I}_n$, the region lying inside the integration contour $\overline{\gamma}$. The following lemma is the counterpart of Lemma 5.25.

Lemma 5.27. *Let a be a real number such that $|a| < 1$. Then $\nu(t,u_n)$ is analytic in a neighbourhood of $\overline{I}_n$. Its modulus in $\overline{I}_n$ is smaller than α, for some $\alpha < 1$ independent of a and n. The series $R_j(t,u_n)$ is also analytic in a neighbourhood of $\overline{I}_n$.*

Proof. The lemma is obvious if $a = 0$. We thus assume $a \neq 0$ and $|a| < 1$.

Let us first study the singularities of the series $\bar{\nu} = \bar{\nu}(z,u)$ defined by

$$\bar{\nu}(z,u) = \nu\left(\frac{z(1+z^2)}{(1+z)^4}, u\right).$$

According to Proposition 5.15, $\bar{\nu}$ is a formal power series in z with polynomial coefficients in u, and by (5.18), one has:

$$\nu(t,u) = \bar{\nu}(Z(t),u).$$

In the proof of Proposition 5.15, we have obtained a polynomial equation $P(\nu, Z, u) = 0$ of degree 4 in ν, relating $\nu(t,u)$, $Z(t)$ and the variable u which can be obtained by using the expression of ν given in Proposition 5.15. By definition of $\bar{\nu}$, we have $P(\bar{\nu}, z, u) = 0$.

Assume that u is a fixed real number close to 1. That is, $u = 1 + x$, with x being small. In order to study the singularities of $\bar{\nu}$, we look again at the zeros of the leading coefficient of P, and at the zeros of its discriminant. This gives several candidates for singularities of $\bar{\nu}(z,u)$, which we classify in three series according to their behaviour when x is small. First, some candidates tend to a limit that is different from 1,

$$z = -1, \quad z = \pm i, \quad z = e^{\pm 2i\pi/3}, \quad z = e^{\pm 2i\pi/3} + O(x), \quad z = -1 + O(x).$$

Then, some candidates tend to 1 and lie at distance at most $|x|^{1/4}$ of 1 (up to a multiplicative constant):

$$z = 1 + \omega(cx)^{1/4} + O(\sqrt{|x|}),$$

where ω is a fourth root of unity and c is in the set $\{0, 16, 64/3, -16/3\}$. Finally, some candidates tend to 1 but lie further away from 1 (more precisely, at distance $|x|^{1/6}$):

$$z = 1 + 2e^{i\pi/6}\omega' x^{1/6} + O(|x|^{1/3}),$$

where ω' is a sixth root of unity.

Let us now consider $\nu(t,u) = \bar{\nu}(Z(t),u)$ with $u = u_n = e^{a/n} = 1+x$, where $x = a/n(1+o(1))$. Recall that Z is analytic in $\mathbb{C} \setminus [1/8, \infty)$. By Lemma 5.23, the series $Z(t)$ never approaches any root of unity different from 1. Hence for n being large enough, $Z(t)$ never reaches any of the candidates z of the first series.

The candidates of the second series are of the form

$$z = 1 + \omega(ac/n)^{1/4} + O(n^{-1/2})$$

for some constant c, with $|c| \leq 64/3$ depending on the candidate. By (5.18), $Z(t)$ may only reach these values for

$$t = \frac{1}{8} - \frac{ac}{128n} + O(n^{-5/4}).$$

Since $|a| < 1$, there exists $\varepsilon > 0$ such that these values lie at distance less than $1/((6+\varepsilon)n)$ of $1/8$, that is, outside a neighbourhood of the domain $\overline{I}_n$.

The candidates of the third series are more worrying: $Z(t)$ may reach them for

$$t = \frac{1}{8} - \frac{\omega''}{8}\left(\frac{a}{n}\right)^{2/3} + O(n^{-5/6}), \tag{5.60}$$

where ω'' is a cubic root of unity, and these values may lie inside $\overline{I}_n$. If $a > 0$ and $\omega'' = e^{\pm 2i\pi/3}$, or if $a < 0$ and $\omega'' = e^{2i\pi/3}$, the modulus of the above value of t is found to be $1/8(1+cn^{-2/3}+o(n^{-2/3}))$, for some positive constant c: this is larger than the radius of the contour $\overline{\gamma}$, which implies that t lies outside a neighbourhood of $\overline{I}_n$. However, if $a > 0$ and $\omega'' = 1$, or if $a < 0$ and $\omega'' = 1$ or $e^{-2i\pi/3}$, the above value of t definitely lies inside $\overline{I}_n$. Its modulus is $1/8(1 - cn^{-2/3} + o(n^{-2/3}))$, for some positive constant c.

In order to rule out the possibility that $\nu(t, u_n)$ has such a singularity, we are going to prove, by having a close look at the expression of ν given in Proposition 5.15, that the radius of convergence of $\nu(t, u_n)$ is at least $1/8 - O(1/n)$. Below we use the notation of Proposition 5.15.

Clearly, the series $V(t, u_n)$ has radius of convergence $\min(1/8, 1/(8u_n))$. In particular, this radius is at least $\rho_n = 1/(8(1 + |x|))$ with $u_n = 1 + x$. Moreover, the series V admits the following expansion

$$V(t, 1 + x) = \frac{1}{4}\left(1 - \sqrt{1 - \frac{8tx}{1 - 8t}}\right) = \frac{1}{2}\sum_{n\geq 1} C_{n-1}\left(\frac{2tx}{1 - 8t}\right)^n,$$

where $C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th Catalan number. This shows that $V(t, 1+|x|)$ is a series in t with *positive* coefficients and that for all t such that $|t| \leq \rho_n$,

$$|V(t, 1 + x)| \leq V(|t|, 1 + |x|) \leq V\left(\frac{1}{8(1 + |x|)}, 1 + |x|\right) = \frac{1}{4}.$$

The next step is to prove that $\varDelta(t, u_n)$ never vanishes for $|t| \leq \rho_n$. Indeed,

$$\varDelta = (1 - V)^2 - 4WV^2,$$

where $W = W(t)$ is the formal power series in t defined by (5.48). This series has radius 1/8, and non-negative coefficients. Hence, for all t such that $|t| \leq 1/8$, one has $|W(t)| \leq W(1/8) = 1/4$. Consequently, for $|t| \leq \rho_n$,

$$|\varDelta(t, 1+x)| \geq (1-|V(t, 1+x)|)^2 - 4|W(t)||V(t, 1+x)|^2 \geq \left(1 - \frac{1}{4}\right)^2 - \frac{1}{16} = \frac{1}{2}.$$

Hence, $\varDelta(t, u_n)$ does not vanish in the centred disk of radius ρ_n. It follows that the series $P(t, u_n)$ is analytic inside this disk.

According to the expression of ν given in Proposition 5.15, the series $\nu(t, u_n)$ is meromorphic for $|t| \leq \rho_n$. The final question we have to answer is whether ν has poles in this disk, and where. Returning to the polynomial P such that $P(\nu, Z, u) = 0$ shows that this can only happen, if the coefficient of ν^4 in this polynomial vanishes. But this can only occur, if $z = Z(t)$ has one of the following forms:

$$z = \pm i, \quad z = e^{\pm 2i\pi/3} + O(x), \quad z = -1 + O(x), \quad z = 1 + \omega(64x/3)^{1/4} + O(x^{1/2}).$$

As argued above, only the last value of z is likely to be reached by $Z(t)$, and this may only occur, if

$$t = \frac{1}{8} - \frac{a}{6n} + O(n^{-5/4}).$$

Consequently, the radius of $\nu(t, u_n)$ is at least $1/8 - O(1/n)$, and this proves that the values (5.60) that have been shown to lie in the centred disk of radius 1/8, are not, after all, singularities of $\nu(t, u_n)$. This completes our proof that $\nu(t, u_n)$ is analytic in a neighbourhood of $\overline{I}_n$.

We now want to bound $\nu(t, u_n)$ inside $\overline{I}_n$. From now on, we can walk safely along the steps of the proof of Lemma 5.25. Let $t_n \in \overline{I}_n$ be such that

$$|\nu(t_n, u_n)| = \max_{t \in \overline{I}_n} |\nu(t, u_n)|.$$

We first give a lower bound for this quantity, by estimating $\nu(t, u_n)$ for $t = 1/8 - 1/(6n)$. This is easily done by combining the closed form expressions of ν (Proposition 5.15) and $B(x)$ (see (5.12)). One obtains:

$$|\nu(t_n, u_n)| \geq |\mu(1/8 - 1/(6n), u_n)| = |B(3a/4)| + o(1) > 0.$$

This lower bound is then used to rule out the possibility that the sequence t_n has an accumulation point different from 1/8. Thus t_n converges to 1/8, and one can write

$$t_n = \frac{1}{8}\left(1 - \frac{x_n}{n}\right).$$

We have $x_n = o(n)$, but also $|x_n| > 4/3$, since t_n belongs to $\overline{I}_n$. We want to estimate $\nu(t_n, u_n)$. Since

$$Z(t_n) = 1 - 2\left(\frac{x_n}{n}\right)^{1/4} + O\left(\left(\frac{x_n}{n}\right)^{1/2}\right)$$

and

$$u_n - 1 = a/n\,(1 + O(1/n)),$$

one has

$$\frac{u-1}{(1-Z(t))^4} = \frac{a}{16x_n}\left(1 + O\left(\left(\frac{x_n}{n}\right)^{1/4}\right)\right).$$

The closed form expressions of ν and B imply that the sequence x_n is bounded and

$$\nu(t_n, u_n) = B(a/x_n) + o(1).$$

Since B is bounded by 0.12 inside its disk of convergence, $|\nu(t_n, u_n)|$ is certainly smaller than some α for $\alpha < 1$ and n large enough. This concludes the proof of the second statement of Lemma 5.27.

By continuity of $\nu(t, u_n)$, this function in t is still bounded by 1 (in modulus) in a neighbourhood of $\overline{I}_n$. Recall also that the modulus of $Z(t)$ never reaches 1 for $t \in \mathbb{C} \setminus [1/8, \infty)$. The form (5.30) then implies that $R_j(t, u_n)$ is an analytic function of t in a neighbourhood of $\overline{I}_n$.

The rest of the proof of Theorem 5.26 is precisely the same as the corresponding part of the proof of Theorem 5.24, with S_j, μ and A, respectively, replaced by R_j, ν and B. The counterpart of (5.59) is

$$\nu(t, u_n) = B(-a/z)(1 + o(1)).$$

Recall that the Hankel part of the contour $\overline{\gamma}$ is now at distance $1/(6n)$ of the real axis. Hence, when n goes to infinity, one finds

$$[t^n]R_j(t,u_n) = \frac{2^n}{n+1}\binom{2n}{n}$$
$$+ \frac{12\cdot 8^n n^{-3/2}}{i\pi}\int_{4/3H} \frac{B(-a/z)\exp(-2\lambda(-z)^{1/4})\sqrt{-z}e^{-z}}{(1+B(-a/z)\exp(-2\lambda(-z)^{1/4}))^2}dz + o(8^n n^{-3/2}).$$

After normalising by $\frac{2^n}{n+1}\binom{2n}{n}$ and setting $v=(-z)^{1/4}$, this gives

$$\mathbb{E}\, e^{aY_n^+(j)} \to 1 + \frac{48}{i\sqrt{\pi}}\int_{(4/3)^{1/4}\Gamma} \frac{B(a/v^4)e^{-2\lambda v}}{(1+B(a/v^4)e^{-2\lambda v})^2}v^5 e^{v^4}\,dv,$$

but the analyticity properties of the integrand allow us to replace the integration contour by Γ.

Remark 5.28 *It is probably possible to prove finite dimensional convergence and tightness by using the above combinatorial and analytic tools (see Lemma 5.8 and its application). This would provide an alternative proof of Theorems 5.4 and 5.5 in these special cases. However, it is not clear whether the above methods will generalise to arbitrary Galton-Watson trees or even to Pólya trees. It would be necessary to study analytic properties of infinite systems of functional equations and so forth. This seems to be out of reach at the moment.*

5.4.4 Embedded Trees with Increments 0 and ±1

If we consider embedded trees with increments 0 and ±1 then we can obtain precisely the same results as presented in Theorems 5.22, 5.24 and 5.26. We just comment the slight modifications that have to be made.

We consider the set of labelled trees having n edges with the uniform distribution, and look at the same random variables as for our first family of trees: M_n, the largest label, $X_n(j)$, the number of nodes having label j, and finally $X_n^+(j)$, the number of nodes having at least label j.

Again, we can prove that $M_n n^{-1/4} \xrightarrow{d} N_{\mathrm{ISE}}/\sqrt{3}$, where N_{ISE} is the supremum of the support of the ISE, and that for all $\lambda\in\mathbb{R}$, $X_n(\lfloor \lambda n^{1/4}\rfloor)n^{-3/4} \xrightarrow{d} \sqrt{3}f_{\mathrm{ISE}}(\sqrt{3}\lambda)$ and $X_n^+(\lambda n^{1/4})/n \xrightarrow{d} G^+(\sqrt{3}\lambda)$. In all three cases, the convergence of the moments holds as well.

5.4.5 Naturally Embedded Binary Trees

For binary trees with n nodes (that are considered to be equally likely) we use the labels corresponding to the natural embedding and consider the same random variables as above: M_n, the largest label, $X_n(j)$, the number of nodes having label j, and $X_n^+(j)$, the number of nodes having label at least j. Here we get that $M_n n^{-1/4} \xrightarrow{d} N_{\mathrm{ISE}}$ and that for all $\lambda\in\mathbb{R}$, $X_n(\lfloor\lambda n^{1/4}\rfloor)n^{-3/4} \xrightarrow{d} f_{\mathrm{ISE}}(\lambda)$ and $X_n^+(\lambda n^{1/4})/n \xrightarrow{d} G^+(\lambda)$, that is, we just have to replace the constant $\sqrt{3}$ by 1.

6

Recursive Trees and Binary Search Trees

Recursive trees and binary search trees can be considered as the result of a growth process and although they are of different structure (in particular, concerning their degree distribution) they have many properties in common.

First, one observes that both kinds of trees are closely related to permutations so that their generating functions are almost the same. But even the finer structure is comparable. In particular we will discuss profile and height. Although the limiting behaviour is not as universal as it is for Galton-Watson trees there are remarkable similarities. The limiting object of the normalised profile is now a stochastic process of random analytic functions and the height distribution is highly concentrated around its mean that is of order $\log n$.

We also discuss d-ary recursive trees, generalised plane oriented trees and (fringe balanced) m-ary search trees, where we observe analogue properties.

The proof methods are leaning very much on the analytic side. Generating functions are very well suited for the above mentioned classes of random trees. Nevertheless, we also make use of sophisticated probabilistic tools like proper L_2 settings for martingales or the contraction method for recursively defined random variables. We completely skip the use of Pólya urn model, since there is a recent monograph [147] on that subject. For early literature and complementary results we refer to the excellent survey by Smythe and Mahmoud [192] and to the article by Bergeron, Flajolet and Salvy [12].

6.1 Permutations and Trees

There are different ways to represent a permutation $\pi \in \mathfrak{S}_n$. One standard way is to use a 2-row matrix

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{pmatrix},$$

or the cycle decomposition.[1] For example, the permutation

$$\pi = \begin{pmatrix} 1\,2\,3\,4\,5\,6\,7\,8 \\ 4\,6\,3\,5\,1\,8\,2\,7 \end{pmatrix} \in \mathfrak{S}_8$$

has the cycle decomposition

$$\pi = (1,4,5)(2,6,8,7)(3)$$

which can also be depicted graphically (see Figure 6.1). Note that the order of the cycles and also the cyclic order inside a cycle are irrelevant.

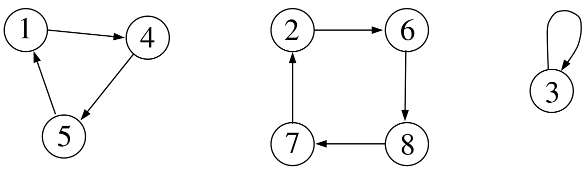


Fig. 6.1. Cycle decomposition of a permutation

The cycle decomposition of permutations can be used to define an evolution process of permutations. Suppose that we have a permutation $\pi \in \mathfrak{S}_n$ represented in its cycle decomposition. We can uniquely reduce π to a permutation $\tilde{\pi} \in \mathfrak{S}_{n-1}$ just by *deleting* n in the cycle decomposition of π. More precisely, if n forms a cycle of length 1, that is, n is a fixed point of π, then we just remove this cycle. However, if n is contained in a cycle of length greater than 1 then there exist i and j with $\pi(i) = n$ and $\pi(n) = j$. Deleting n means we set $\tilde{\pi}(i) = j$. For example, if we remove $n = 8$ in our previous example then we obtain

[1] A cycle C (of length k) of a permutation π is a sequence of the kind $C = (j, \pi(j), \pi^2(j), \cdots, \pi^{k-1}(j))$ with $\pi^k(j) = j$, where k is the smallest positive integer with this property. Every permutation π induces a partition of the underlying set $\{1, 2, \ldots,\}$ of cycles, which is called the cycle decomposition of π. We also say that j is a fixed point of π if $\pi(j) = j$, that is, (j) is a singleton in the corresponding cycle decomposition. A permutation is called cyclic, if the cycle decomposition consists just of one cycle.

$$\tilde{\pi} = \begin{pmatrix} 1\ 2\ 3\ 4\ 5\ 6\ 7 \\ 4\ 6\ 3\ 5\ 1\ 7\ 2 \end{pmatrix} \in \mathfrak{S}_7$$

with its cycle decomposition

$$\tilde{\pi} = (1,4,5)(2,6,7)(3)$$

(compare also with Figure 6.2).

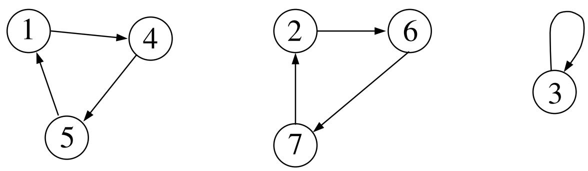


Fig. 6.2. Cycle decomposition of a reduced permutation $\tilde{\pi}$

It is also possible to reverse this process. Starting with $\tilde{\pi} \in \mathfrak{S}_{n-1}$ given through its cycle decomposition we can include n either by defining a new cycle of length 1, that is, we set $\pi(n) = n$, or we include it somewhere in the existing cycles. Since there are precisely k possible ways to include a new element into a cycle of length k, there are $n-1$ possible ways to include an n element into the already existing cycles of $\tilde{\pi}$. Hence, in total there are $n = n-1+1$ possible ways to extend $\tilde{\pi} \in \mathfrak{S}_{n-1}$ to a permutation $\pi \in \mathfrak{S}_n$ consistently with the cycle decomposition.

This can be, of course, extended to a process that generates all $n!$ permutations. Starting with $(1) \in \mathfrak{S}_1$ there are two possible ways to extend (1) to a permutation in $\mathfrak{S}_2$. Next, for every permutation in $\mathfrak{S}_2$ there are three different ways to extend it to a permutation in $\mathfrak{S}_3$, and so on. Since we can uniquely go back, every permutation is uniquely constructed in this way (and also provides another proof of $|\mathfrak{S}_n| = n!$).

6.1.1 Permutations and Recursive Trees

We recall that recursive trees are rooted labelled trees, where the root is labelled by 1 and the labels of all successors of any node v are larger than the label of v. The labels can also be seen as the result of an evolution process where at the j-th step the node with label j is attached. Thus, the labels encode the history of the process. Using this interpretation it is easy to see that there are exactly $(n-1)!$ different recursive trees of size n. For example, Figure 6.3 depicts all 6 recursive trees of size 4.

In order to show that there is a close relation to permutations we slightly change the labels by starting with label 0 at the root. Then there are $n!$ recursive trees with labels $0, 1, \ldots, n$.

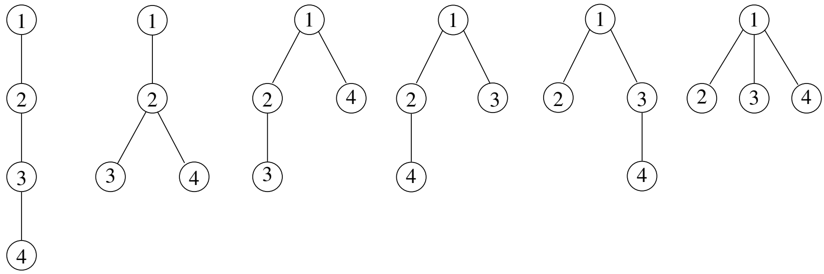


Fig. 6.3. All recursive trees of size 4

We now present a bijection between the evolution process of recursive trees (with labels $\{0, 1, \ldots, n\}$) and permutations (in $\mathfrak{S}_n$) in their cycle decomposition. We start with the recursive tree consisting just of the root that is labelled by 0 at one hand and with the empty permutation on the other hand. The first step is to attach a node with label 1 to the root and to consider the permutation $1 \in \mathfrak{S}_1$. Now we proceed inductively in the following way. If the j-node is attached to the root node then modify the corresponding cycle decomposition by appending a fixed point that maps j to j, that is, a new cycle of length 1. On the other hand, if the j-th node is attached to node $i > 0$ (that is different from the root) then we consider the cycle C that contains i. If i is followed by k in this cycle then we modify it by including j in-between i and k, that is, in the new cycle i is followed by j and j is followed by k. Figure 6.4 shows this procedure applied to the example permutation π. This kind of process is also called *Chinese restaurant process*. The cycles in the permutations can be interpreted as (circular) tables in a Chinese restaurant and when a new guest arrives he/she either opens a new table or joins an already used table.

It is easy to check that this procedure provides a bijection. Furthermore the uniform random model on recursive trees corresponds to the uniform random model on permutations for every fixed n. The bijection also transfers some shape characteristics. For example, the root degree of the recursive tree corresponds to the number of cycles in the cycle decomposition and the subtree sizes of the root correspond to the sizes of the cycles.

Let us use this correspondence in two ways to characterise the limiting behaviour of the root degree of a random recursive tree resp. the number of cycles in a random permutation.

The Stirling numbers of the first kind $s_{n,k}$ are defined by

$$\sum_{k=0}^{n} s_{n,k} u^k = u(u-1)\cdots(u-n+1).$$

The first few values are given in Figure 6.5.

Equivalently, they can be defined by the initial values $s_{0,k} = \delta_{0,k}$, $s_{n,0} = \delta_{n,0}$ and the recurrence relation

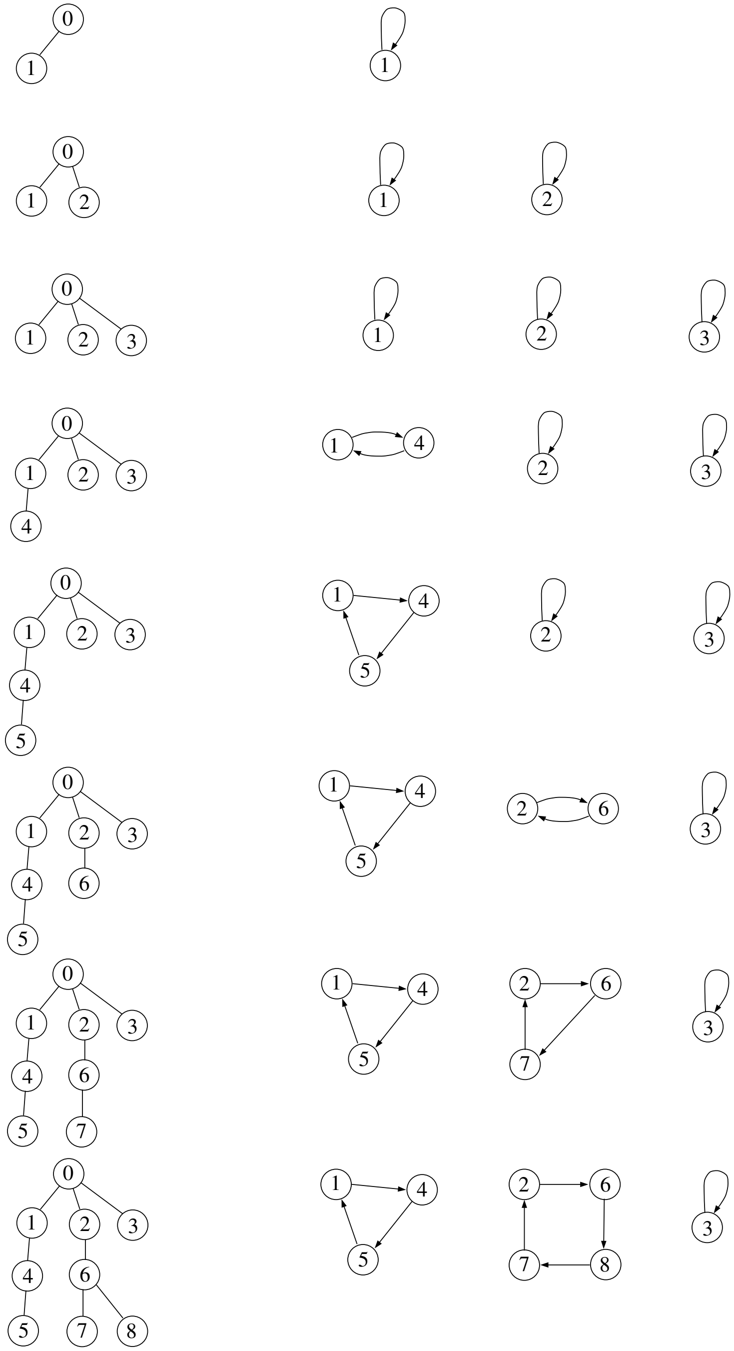


Fig. 6.4. Bijection between recursive trees and permutations

$s_{n,k}$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$n=0$	1								
$n=1$	0	1							
$n=2$	0	-1	1						
$n=3$	0	2	-3	1					
$n=4$	0	-6	11	-6	1				
$n=5$	0	24	-50	35	-10	1			
$n=6$	0	-120	274	-225	85	-15	1		
$n=7$	0	720	-1764	1624	-735	175	-21	1	
$n=8$	0	-5040	13068	-13132	6769	-1960	322	-28	1

Fig. 6.5. Stirling numbers of the first kind

$$s_{n+1,k} = s_{n,k-1} - n s_{n,k}.$$

It is easy to check that the Stirling numbers have alternating signs. Hence, $|s_{n,k}| = (-1)^{n+k} s_{n,k}$. These absolute values can, thus, be defined by

$$\sum_{k=0}^{n} |s_{n,k}| u^k = u(u+1)\cdots(u+n-1)$$

or by the recurrence

$$|s_{n+1,k}| = |s_{n,k-1}| + n|s_{n,k}|. \tag{6.1}$$

It is well known that the absolute value $|s_{n,k}|$ coincides with the number of permutations $a_{n,k}$ in $\mathfrak{S}_n$ where the cycle decomposition consists of k cycles. There are various ways to prove this assertion. For example, by using the above evolution process of permutations, it follows that the numbers $a_{n,k}$ satisfy the same recurrence as (6.1) and they have the same initial values.

Permutations can certainly be viewed as labelled objects with the exponential generating function

$$a(x) = \sum_{n\ge 0} n! \frac{x^n}{n!} = \frac{1}{1-x}.$$

Similarly the generating function of cyclic permutations is given by

$$c(x) = \sum_{n\ge 1} (n-1)! \frac{x^n}{n!} = \sum_{n\ge 1} \frac{x^n}{n} = \log\frac{1}{1-x}.$$

Indeed, this corresponds with the cycle decomposition: every permutation is a set of cycles:

$$a(x) = e^{c(x)} = e^{\log\frac{1}{1-x}} = \frac{1}{1-x}.$$

Consequently, the double generating function for the number $a_{n,k}$ is given by

$$a(x,u) = \sum_{n,k} a_{n,k} \frac{x^n}{n!} u^k = e^{u \log \frac{1}{1-x}} = \frac{1}{(1-x)^u}. \tag{6.2}$$

If we now consider the random variable C_n that counts the number of cycles in random permutations in $\mathfrak{S}_n$, that is,

$$\mathbb{P}\{C_n = k\} = \frac{a_{n,k}}{n!},$$

we obtain from (6.2) and (2.10)

$$\begin{aligned} \mathbb{E}\, u^{C_n} &= \frac{n!}{n!} [x^n] a(x,u) \\ &= (-1)^n \binom{-u}{n} \\ &= \frac{n^{u-1}}{\Gamma(u)} + O(n^{\Re(u)-2}) \\ &= \frac{1}{\Gamma(u)} e^{(u-1)\log n} \left(1 + O\left(\frac{1}{n}\right)\right), \end{aligned}$$

provided that u is contained in a (small) complex neighbourhood of $u = 1$.

Now a direct application of the Quasi Power Theorem 2.22 implies that C_n satisfies a central limit theorem with $\mathbb{E}\, C_n = \log n + O(1)$ and $\mathbb{V}\mathrm{ar}\, C_n = \log n + O(1)$:

$$\frac{C_n - \log n}{\sqrt{\log n}} \xrightarrow{d} N(0,1). \tag{6.3}$$

This also shows that the root degree R_{n+1} of random recursive trees (of size $n+1$) has the same behaviour.

We complete this combinatorial approach to the root degree with a probabilistic one. Looking at the evolution process of recursive trees it follows that R_{n+1} can be represented as

$$R_{n+1} = \sum_{j=1}^{n} \xi_j,$$

where $\xi_j = 1$ if the j-th node is attached to the root, and $\xi_j = 0$ if it is attached somewhere else. By definition the random variables ξ_j $(1 \le j \le n)$ are independent and have the probability distribution $\mathbb{P}\{\xi_j = 1\} = 1/j$, that is, they are Bernoulli distributions with parameter $1/j$. Hence we have

$$\mathbb{E}\, R_{n+1} = \sum_{j=1}^{n} \frac{1}{j} = \log n + O(1),$$

and

$$\mathbb{V}\mathrm{ar}\, R_{n+1} = \sum_{j=1}^{n} \left(\frac{1}{j} - \frac{1}{j^2} \right) = \log n + O(1).$$

Finally, the classical central limit theorem (for independent but not necessarily identically distributed random variables) can be applied and we recover (6.3).

We close this section by providing a bivariate asymptotic expansion for Stirling numbers of the first kind that is valid for $k = O(\log n)$.

Lemma 6.1. *We have uniformly for* $c \log n \le k \le C \log n$

$$|s_{n,k}| = \frac{(n-1)!(\log n)^k}{k!\Gamma\left(\frac{k}{\log n}\right)} \left(1 + O\left(\frac{1}{\sqrt{\log n}}\right)\right),$$

where $c < C$ *are arbitrary positive constants.*

Proof. By definition we have

$$\sum_{k\ge 0} |s_{n,k}| u^k = (-1)^n \binom{-u}{n} n!.$$

Hence, by Lemma 2.14 we have uniformly for $c \le |u| \le C$

$$\begin{aligned}\sum_{k\ge 0} |s_{n,k}| u^k &= n! \frac{n^{u-1}}{\Gamma(u)} \left(1 + O\left(\frac{1}{n}\right)\right) \\ &= (n-1)! \frac{e^{u \log n}}{\Gamma(u)} \left(1 + O\left(\frac{1}{n}\right)\right)\end{aligned}$$

and consequently by Cauchy's formula

$$|s_{n,k}| = \frac{(n-1)!}{2\pi i} \int_\gamma \frac{e^{u \log n}}{\Gamma(u)} \left(1 + O\left(\frac{1}{n}\right)\right) \frac{du}{u^{k+1}}.$$

We will choose γ to be a circle $|u| = r$ with radius $r = k/\log n$, since the dominating part of the integrand

$$e^{u \log n} u^{-k} = e^{u \log n - k \log u}$$

has a saddle point at $u = k/\log n$. In fact, a standard saddle point method yields the result.

We use the substitution $u = re^{it}$ $(-\pi < t \le \pi)$ with $r = k/\log n$. Then by using the approximation $e^{it} = 1 + it - \frac{1}{2}t^2 + O(t^3)$ we get

$$e^{u \log n} u^{-k} = e^k r^{-k} e^{-\frac{k}{2}t^2 + O(kt^3)}.$$

We split up the integral into two parts:

$$\begin{aligned}\gamma_1 &= \{u = re^{it} : |t| \le k^{-1/3}\}, \\ \gamma_2 &= \{u = re^{it} : k^{-1/3} < |t| \le \pi\}.\end{aligned}$$

For the integral over γ_1 we have

$$\begin{aligned}
&\frac{(n-1)!}{2\pi i}\int_{\gamma_1}\frac{e^{u\log n}}{\Gamma(u)}\left(1+O\left(\frac{1}{n}\right)\right)\frac{du}{u^{k+1}}\\
&\quad=\frac{(n-1)!}{2\pi}\int\limits_{|t|\le k^{-1/3}}\frac{e^k r^{-k}}{\Gamma(r)+O(t)}e^{-\frac{k}{2}t^2+O(kt^3)}\left(1+O\left(\frac{1}{n}\right)\right)dt\\
&\quad=\frac{e^k r^{-k}}{\Gamma(r)}\frac{(n-1)!}{2\pi}\int\limits_{|t|\le k^{-1/3}}e^{-\frac{k}{2}t^2}\left(1+O(t)+O(kt^3)+O\left(\frac{1}{n}\right)\right)\\
&\quad=\frac{e^k r^{-k}}{\Gamma(r)}\frac{(n-1)!}{2\pi}\left(\sqrt{\frac{2\pi}{k}}+O\left(e^{-\frac{1}{2}k^{1/3}}\right)+O\left(\frac{1}{k}\right)+O\left(\frac{1}{n}\right)\right)\\
&\quad=\frac{(n-1)!(\log n)^k}{k!\Gamma(r)}\left(1+O\left(\frac{1}{\sqrt{k}}\right)\right).
\end{aligned}$$

The integral over γ_2 is negligible. This can be checked by the following observations. By continuity it follows that there exists a constant $c'>0$ such that for all $|t|\le\pi$ and uniformly for $c\le r\le C$

$$\left|e^{u\log n}u^{-k}\right|\le e^k r^{-k}e^{-c'kt^2}.$$

Furthermore we have

$$\int\limits_{k^{-1/3}\le|t|\le\pi}e^{-c'kt^2}\,dt=O\left(e^{-c'k^{1/3}}\right).$$

This completes the proof of the lemma.

Remark 6.2 *Lemma 6.1 holds even uniformly for $1\le k\le C\log n$. This can be proved, for example, by using the series expansion*

$$\sum_{n\ge 0}|s_{n,k}|\frac{x^n}{n!}=\frac{1}{k!}\left(\log\frac{1}{1-x}\right)^k$$

and a Hankel like contour integration applied to Cauchy's formula (similarly to the proof of Lemma 2.14).

Using this asymptotic expansion we obtain even more precise information on the local behaviour of the distribution of the number of cycles in permutations (resp. of the root degree of recursive trees). Since $\mathbb{P}\{C_n=k\}=|s_{n,k}|/n!$, we have (uniformly for $c\log n\le k\le C\log n$)

$$\mathbb{P}\{C_n=k\}=\frac{(\log n)^k}{n\,k!\Gamma\left(\frac{k}{\log n}\right)}\left(1+O\left(\frac{1}{\sqrt{\log n}}\right)\right).$$

In particular, if k is close to $\log n$, that is, if $|k - \log n| = O((\log n)^{2/3})$ then we also get

$$\mathbb{P}\{C_n = k\} = \frac{1}{\sqrt{2\pi \log n}} e^{-\frac{(k-\log n)^2}{2\log n}} \left(1 + O\left(\frac{|k - \log n|^3}{(\log n)^2}\right) + O\left(\frac{1}{\sqrt{\log n}}\right)\right),$$

which is a local central limit theorem.

6.1.2 Permutations and Binary Search Trees

In Section 1.4.1 we have introduced binary search trees as binary trees that are generated by input keys that can be viewed as a permutation π. The construction starts by putting $\pi(1)$ to the root and ends with a tree where all nodes that are put to the left of an internal node v have a smaller value and all nodes that are put to the right have a larger value. For example the permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 6 & 3 & 5 & 1 & 8 & 2 & 7 \end{pmatrix} \in \mathfrak{S}_8$$

resp. the list $(4, 6, 3, 5, 1, 8, 2, 7)$ induces the binary tree that is depicted on the left-hand-side of Figure 6.6.

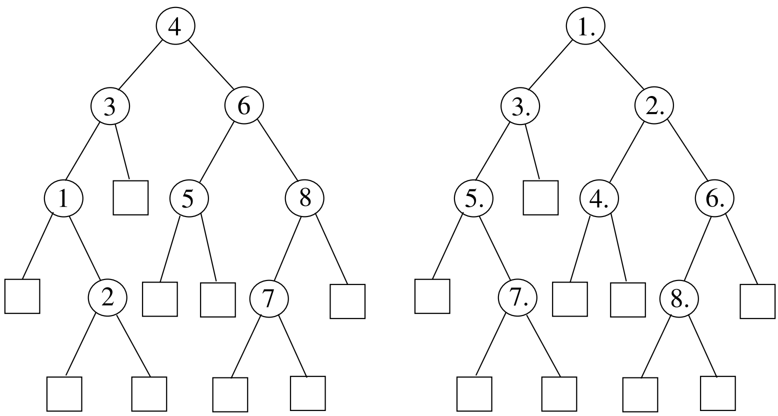
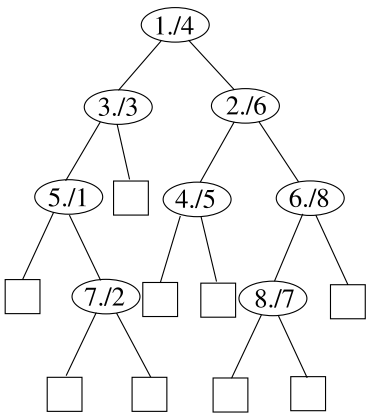


Fig. 6.6. Binary search tree and its evolution with input list $(4, 6, 3, 5, 1, 8, 2, 7)$

Note, however, that this mapping is not one-to-one. For example, the list $(4, 3, 6, 5, 1, 8, 7, 2)$ produces exactly the same (labelled) tree. Nevertheless, if we take care of the insertion order of the evolution process then the situation is different. For example, the right-hand-side in Figure 6.6 shows the order of insertion for this particular example. However, in this way we just recover binary recursive trees (see Section 1.3.3, where we have discussed ternary recursive trees in more detail). The input list encodes in which order we have to put the nodes in the evolution process.

Interestingly, this is actually a bijection between permutations $\pi \in \mathfrak{S}_n$ and binary recursive trees with n labelled internal nodes. We already discussed how a permutation leads to a binary recursive tree. Conversely, if a binary recursive tree is given, where the labels encode the evolution of the tree, then we use the fact that the values of nodes of the left subtree of any node v are smaller than the value of v to recover the permutation π. In our example the left subtree of the root has 3 internal nodes. Hence, the root corresponds to the value 4. In this way we can continue. If we have the correspondence between the values of π and the order of the insertion we get π (see Figure 6.7).



$$\longrightarrow \quad \pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 6 & 3 & 5 & 1 & 8 & 2 & 7 \end{pmatrix}$$

Fig. 6.7. Bijection between binary recursive trees and permutations

We will not exploit this bijection further. Nevertheless, the study of random binary search trees (or random binary recursive trees) is always related to the study of random permutations.

6.2 Generating Functions and Basic Statistics

Since recursive trees and binary recursive trees are labelled objects, it is natural to use exponential generating functions. Nevertheless there is some flexibility in the case of recursive trees as we will see.

We will now show that recursive trees have a similar recursive structure as simply generated trees. Note that the definition of recursive trees is a *bottom-up*-approach. A recursive tree of size $n+1$ is obtained by attaching a new node to a recursive tree of size n. However, it is also possible to use a *top-down*-view by separating the root and by studying subtrees that can be considered as (relabelled) recursive trees (of a smaller size, of course).

6.2.1 Generating Functions for Recursive Trees

We already know that there are $(n-1)!$ recursive trees of size n. Hence, the exponential generating function $y(x)$ is given by

$$y(x) = \sum_{n\ge 1} (n-1)! \frac{x^n}{n!} = \sum_{n\ge 1} \frac{x^n}{n} = \log \frac{1}{1-x}. \tag{6.4}$$

We first show how we could have gotten this relation by using the recursive structure of recursive trees.

Lemma 6.3. *The exponential generating function $y(x)$ of recursive trees satisfies the differential equation*

$$y'(x) = e^{y(x)}, \quad y(0) = 0. \tag{6.5}$$

Proof. Instead of considering a recursive tree with labels $\{1, 2, \ldots, n\}$ we look at a recursive tree with labels $\{0, 1, 2, \ldots, n\}$, where the root gets label 0. Then the generating function of this structure is given by

$$Y(x) = y'(x).$$

Equivalently we can say that we disregard the root. Then the subtrees of the root represent an unordered labelled sequence of recursive trees of total size n and have, thus, the generating function $e^{y(x)}$ (compare also with the graphical representation of the recursive structure given in Figure 6.8). This completes the proof of the lemma.

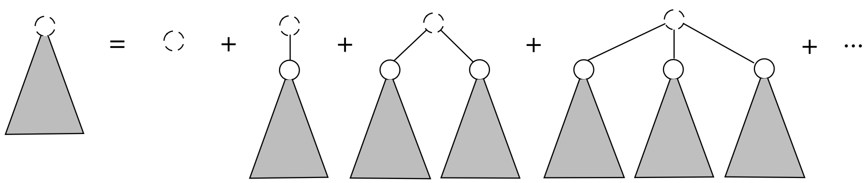


Fig. 6.8. Recursive structure of recursive trees

The differential equation (6.5) is easy to solve and we, thus, recover the explicit representation (6.4).

In the proof of Lemma 6.3 we made use of recursive trees, where the root is disregarded (or labelled with 0) and has the generating function

$$Y(x) = y'(x) = \frac{1}{1-x}. \tag{6.6}$$

Alternatively to the recurrence used in Lemma 6.3 we can also use the following recurrence (compare also with the alternate recurrence for planted plane trees depicted in Figure 3.4). We divide the recursive tree (where the root is labelled by 0) into two trees, the subtree that is rooted at the node with label 1 and the remaining tree (see Figure 6.9). Formally we do that by disregarding the root and the vertex with label 1, which corresponds to the generating function $Y'(x) = y''(x)$. Due to this decomposition we get the relation

$$Y'(x) = Y(x)^2, \quad Y(0) = 1. \tag{6.7}$$

This also follows from (6.6).

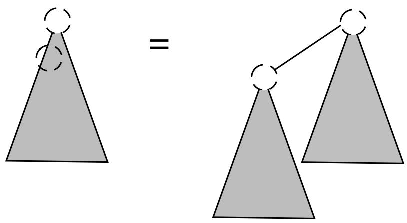


Fig. 6.9. Alternate recursive structure of recursive trees

6.2.2 Generating Functions for Binary Search Trees

The usual probabilistic model for binary search trees is the random permutation model, that is, one assumes that every permutation $\pi \in \mathfrak{S}_n$ is equally likely. This naturally induces a probability distribution for binary search trees with n internal nodes. Equivalently we can consider binary recursive trees. Due to the bijective relation to permutations the random permutation model corresponds to the uniform distribution on binary recursive trees of given size.

Let us work with binary recursive trees. Since there are $n!$ different trees of size n, the exponential generating function $y(x)$ is given by

$$y(x) = \sum_{n\ge 1} n!\,\frac{x^n}{n!} = \frac{x}{1-x}.$$

As in the case of recursive trees it is also possible to use a *top-down-view* by separating the root and by studying the two subtrees. This again leads to a differential equation.

Lemma 6.4. *The exponential generating function $y(x)$ of binary recursive trees satisfies the differential equation*

$$y'(x) = (1 + y(x))^2, \quad y(0) = 0. \tag{6.8}$$

Proof. Let $y'(x)$ denote the (exponential) generating function of binary recursive trees, where the root is disregarded. Then the two subtrees of the root represent a labelled product of the union of the empty tree and a binary recursive tree, which has the (exponential) generating function $(1+y(x))^2$ (compare also with Figure 6.10).

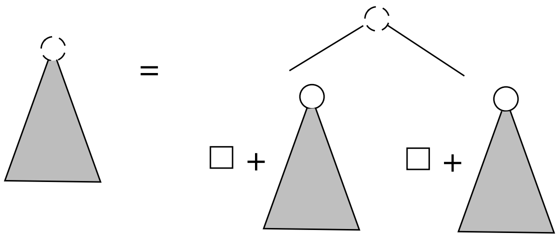


Fig. 6.10. Recursive structure of binary recursive trees

The relations (6.7) and (6.8) again show that recursive trees and binary recursive trees (resp. binary search trees) are very similar. Nevertheless the actual shape characteristics are not necessarily the same (see again Figures 6.9 and 6.10).

Occasionally it will be convenient to replace $y(x)$ by $\tilde{y}(x) = y(x)+1$, that is we change the convention for recursive trees of size 0. Then (6.8) rewrites to

$$\tilde{y}'(x) = \tilde{y}(x)^2.$$

Equivalently we can think of binary search trees, where it is usual to include the case of size 0. Thus, the generating function for binary search trees is $y(x) = 1/(1-x)$.

Lemma 6.4 extends directly to d-ary recursive trees (compare with Figures 1.10 and 1.11).

Lemma 6.5. *The exponential generating function $y(x)$ of d-ary recursive trees (with an integer $d \ge 2$) satisfies the differential equation*

$$y'(x) = (1+y(x))^d, \quad y(0) = 0, \tag{6.9}$$

and is explicitly given by

$$y(x) = (1-(d-1)x)^{-\frac{1}{d-1}} - 1.$$

Consequently, the number of different d-ary trees of size n is given by

$$\begin{aligned} y_n &= \prod_{j=1}^{n-1} (jd - (j-1)) \\ &= n!(-1)^n (d-1)^n \binom{-\frac{1}{d-1}}{n} \\ &\sim n!(d-1)^n \frac{n^{\frac{2-d}{d-1}}}{\Gamma\left(\frac{1}{d-1}\right)}. \end{aligned}$$

Again it is also convenient to consider instead

$$\tilde{y}(x) = y(x) + 1 = (1 - (d-1)x)^{-\frac{1}{d-1}}$$

that satisfies the differential equation $\tilde{y}'(x) = \tilde{y}(x)^d$.

6.2.3 Generating Functions for Plane Oriented Recursive Trees

Plane oriented recursive trees are similarly defined to recursive trees with the only difference that the left-to-right-order of the successors is taken into account. This means that we are considering plane trees. There are again different points of view.

The corresponding evolution process works as follows. As usual we start with the root that gets label 0. Now if a node already has out-degree k (where the descendants are ordered), then there are $k+1$ possible ways to attach a new node. Hence, if a plane tree already has $j-1$ nodes then there are precisely $2j-3$ possibilities to attach the j-th node and to generate a plane oriented recursive tree of size j. We already observed in Section 1.3.2 that this consideration implies that there are

$$1 \cdot 3 \cdot \ldots \cdot (2n-3) = (2n-3)!! = \frac{1}{2^{n-1}} \frac{(2(n-1))!}{(n-1)!}$$

different plane oriented recursive trees.

We can also use a generating function approach.

Lemma 6.6. *The exponential generating function $y(x)$ of plane oriented recursive trees satisfies the differential equation*

$$y'(x) = \frac{1}{1-y(x)}, \quad y(0) = 0 \tag{6.10}$$

and is explicitly given by

$$y(x) = 1 - \sqrt{1-2x} = \sum_{n \ge 1} (2n-3)!! \, \frac{x^n}{n!}.$$

Proof. The only difference to the proof of Lemma 6.3 is that the subtrees of the root are now ordered and, thus, constitute a sequence of plane oriented trees that is counted by $1/(1-y(x))$.

The natural probabilistic model is to assume that every plane oriented recursive tree of size n is equally likely. Since the evolution process discussed above represents every tree in a unique way, this uniform model is obtained in the following way, too.

The process starts with the root that is labelled with 1. Then inductively at step j a new node (with label j) is attached to any previous node of out-degree k with a probability proportional to $k+1$. More precisely, the probability of choosing a node of out-degree k equals $(k+1)/(2j-3)$.

If we think in this way it is not necessary anymore to consider plane trees. It is now a non-uniform evolution process on (usual) recursive trees, where the local out-degrees of the nodes are taken into account.

Now we use a slightly more general model. We fix a parameter $r>0$ and randomly generate recursive trees in the following way: We start with the root that is labelled with 1. Then inductively at step j a new node (with label j) is attached to any previous node of out-degree k with probability proportional to $k+r$; more precisely the probability equals $(k+r)/((r+1)j-(r+2))$.

If $r=1$, we get plane oriented recursive trees. It is further possible to describe these kinds of trees by using a recursive procedure that leads to a differential equation of the form

$$y'(x) = \sum_{m\ge 0} (-1)^m \binom{-r}{m} y(x)^m = \frac{1}{(1-y(x))^r}$$

with solution

$$y(x) = \sum_{n\ge 1} y_n \frac{x^n}{n!} = 1 - (1-(r+1)x)^{\frac{1}{r+1}} .$$

The n-th coefficient

$$\begin{aligned} y_n &= (-1)^{n-1} n! (r+1)^n \binom{\frac{1}{r+1}}{n} \\ &\sim -n!(r+1)^n \frac{n^{-\frac{r+2}{r+1}}}{\Gamma\left(-\frac{1}{r+1}\right)} \end{aligned} \tag{6.11}$$

has to be interpreted as the sum of weights of trees of size n, where the weights are proportional to the probabilities that come from the evolution process. Anyway, it is a special class of increasing trees.[2]

[2] We have discussed this correspondence more precisely in Section 1.3.3.

6.2.4 The Degree Distribution of Recursive Trees

Next we will discuss the distribution of the number of nodes of given degree in a random recursive tree of size n, where we use one of the above evolution processes to define the probability distribution. First of all note that all discussed trees (recursive trees, d-ary recursive trees,[3] variants of plane oriented recursive trees) can be considered as special kinds of increasing trees (see Section 1.3.3) where the corresponding exponential generating function $y(x)$ satisfies a differential equation of the kind

$$y'(x) = \Phi(y(x)),$$

where $\Phi(w) = \phi_0 + \phi_1 w + \phi_2 w^2 + \cdots$ is a generating series of weights. As described in Section 1.3.3 this kind of equation can be interpreted in a combinatorial way. The left-hand-side $y'(x)$ corresponds to those objects, where we disregard the root resp. where the root (formally) gets the label 0. The right-hand-side is then the weighted union of (labelled) n-tuples of trees, that is

$$\phi_0 + \phi_1 y(x) + \phi_2 y(x)^2 + \cdots .$$

This is very similar to the combinatorial recursive description of simply generated trees resp. of Galton-Watson trees.

Now let $X_n^{(k)}$ denote the (random) number of nodes of out-degree k in a class of random recursive trees of size n, where the probability distribution is defined in terms of the weights ϕ_j. We can extend the above combinatorial description to characterise the distribution of $X_n^{(k)}$.

Lemma 6.7. *The bivariate generating function*

$$y_k(x,u) = \sum_{n\ge 1} y_n \mathbb{E} u^{X_n^{(k)}} \frac{x^n}{n!} = \sum_{n,m} y_n \mathbb{P}\{X_n^{(k)} = m\} \frac{x^n}{n!} u^m$$

satisfies the differential equation

$$\frac{\partial y_k(x,u)}{\partial x} = \Phi(y_k(x,u)) + \phi_k\,(u-1)\, y_k(x,u)^k, \quad y(0,u) = 0. \tag{6.12}$$

Moreover, we have

$$\left[\frac{\partial y_k(x,u)}{\partial u}\right]_{u=1} = \sum_{n\ge 1} y_n \mathbb{E}\, X_n^{(k)} \frac{x^n}{n!} = \phi_k y'(x) \int_0^x \frac{y(t)^k}{y'(t)}\, dt. \tag{6.13}$$

Proof. Equation (6.12) follows from the combinatorial decomposition, since the number of nodes of out-degree k equals the sum of those nodes in the

[3] In d-ary recursive trees the node degree is defined by the number of internal descendent nodes.

subtrees of the root with the only exception that we have to add 1, if the out-degree of the root equals k.

Next set

$$S(x) = \left[\frac{\partial y(x,u)}{\partial u}\right]_{u=1}.$$

Then from (6.12) we derive the linear differential equation

$$S'(x) = \Phi'(y(x))S(x) + \phi_k y(x)^k, \qquad S(0) = 0$$

which has the solution (6.13).

Before we present general results on the distribution of $X_n^{(k)}$ we comment on a special case, namely on the number of leaves $L_n = X_n^{(0)}$ of recursive trees. In this case we have $\Phi(x) = e^x$ and $y_n = (n-1)!$. Here we explicitly obtain

$$\mathbb{P}\{L_n = k\} = \frac{1}{(n-1)!}\left\langle {n-1 \atop k-1} \right\rangle, \tag{6.14}$$

where $\left\langle {n \atop k} \right\rangle$ denotes the Eulerian numbers. These numbers satisfy a recurrence that looks similar to that of Stirling numbers:

$$\left\langle {n \atop k} \right\rangle = (k+1)\left\langle {n-1 \atop k} \right\rangle + (n-k)\left\langle {n-1 \atop k-1} \right\rangle.$$

The first few values are listed in Figure 6.11.

There is an easy proof for (6.14). It is immediate from the definition of the evolution process of recursive trees that the number of recursive trees of size n with k leaves satisfies the same recurrence as the numbers $\left\langle {n-1 \atop k-1} \right\rangle$ (and have the same initial values). Consequently they coincide.

$\left\langle {n \atop k} \right\rangle$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$n=0$	1								
$n=1$	1	1							
$n=2$	1	4	1						
$n=3$	1	11	11	1					
$n=4$	1	26	66	26	1				
$n=5$	1	57	302	302	57	1			
$n=6$	1	120	1191	4216	1191	120	1		
$n=7$	1	247	4293	15619	15619	4293	247	1	
$n=8$	1	502	14608	88234	156190	88234	14608	502	1

Fig. 6.11. Eulerian numbers

Note that the differential equation

$$\frac{\partial y(x,u)}{\partial x} = u + e^{y(x,u)} - 1$$

for the generating function $y(x,u) = \sum_{n\ge 1} \mathbb{E}\, u^{L_n}\, x^n/n$ has an explicit solution

$$y(x,u) = \log\left(\frac{u-1}{ue^{-x(u-1)}-1}\right),$$

which can be used to obtain explicit formulas for mean value and variance:

$$\mathbb{E}\, L_n = \frac{n}{2}, \qquad \mathbb{V}\mathrm{ar}\, L_n = \frac{7}{12}n + \frac{1}{3}.$$

Moreover, a central limit theorem for L_n holds:

$$\frac{L_n - \frac{n}{2}}{\sqrt{\frac{7}{12}n}} \xrightarrow{d} N(0,1).$$

In fact we will extend this kind of limiting result to $X_n^{(k)}$ for all k and for all classes of recursive trees we have discussed. In order to make this more transparent we split it into three statements.

Theorem 6.8. *Let $X_n^{(k)}$ denote the (random) number of nodes of out-degree k in a random recursive tree of size n in the uniform model. Then for every $k \ge 0$, $X_n^{(k)}$ satisfies a central limit theorem with mean value and variance asymptotically proportional to n. In particular we have*

$$\mathbb{E}\, X_n^{(k)} = \frac{n}{2^{k+1}} + O((\log n)^{k+1}).$$

Theorem 6.9. *Let $X_n^{(k)}$ denote the (random) number of (internal) nodes with k internal successors in a random d-ary recursive tree of size n, where $d \ge 2$ is a fixed integer. Then for every $k \ge 0$, $X_n^{(k)}$ satisfies a central limit theorem with mean value and variance asymptotically proportional to n. In particular we have for $0 \le k \le d$*

$$\mathbb{E}\, X_n^{(k)} = \frac{(d-1)d!(2d-k-2)!}{(2d-1)!(d-k)!}n + O\left(n^{\min\{\frac{d-2}{d-1},\frac{k-1}{d-1}\}}\right).$$

Theorem 6.10. *Let $X_n^{(k)}$ denote the (random) number of nodes of out-degree k of a random increasing tree of size n, defined by the generating series $\Phi(x) = (1-x)^{-r}$, where $r > 0$. Then for every $k \ge 0$, $X_n^{(k)}$ satisfies a central limit theorem with mean value and variance asymptotically proportional to n. In particular we have*

$$\mathbb{E}\, X_n^{(k)} = n\frac{(r+1)\Gamma(2r+1)\Gamma(r+k)}{\Gamma(r)\Gamma(2r+k+2)} + O\left(n^{\frac{1}{r+1}}\right).$$

Theorem 6.10 is of special interest, since the corresponding (asymptotic) degree distribution

$$d_k = \frac{(r+1)\Gamma(2r+1)\Gamma(r+k)}{\Gamma(r)\Gamma(2r+k+2)}$$

has a polynomial tail of the form

$$\frac{(r+1)\Gamma(2r+1)}{\Gamma(r)}\, k^{-2-r}.$$

Such distributions are also called *scale-free* and appear in several *real networks* like the internet or social networks that have an evolution in time. In particular there is substantial interest in random graph models where vertices are added to the graph successively and are connected to several already existing nodes according to some given law. The so-called Albert-Barabási model (see [1]) joins a new node to an existing one with probability proportional to the degree. Recall that the random recursive trees that are considered in Theorem 6.10 are constructed exactly by this principle: a new node is attached to an already existing one (with out-degree k) with probability proportional to $k+r$. Thus, these models confirm the observation that networks that follow the Albert-Barabási principle have asymptotically a scale-free degree distribution.

The proofs of Theorems 6.8–6.10 are very similar. So we group them together.

Proof. Let us start with recursive trees. The differential equation ($y = y_k(x,u)$)

$$\frac{\partial y}{\partial x} = e^y + \frac{y^k}{k!}(u-1) \tag{6.15}$$

can be rewritten to

$$\int_0^y \frac{1}{e^v + \frac{v^k}{k!}(u-1)}\, dv = x. \tag{6.16}$$

We split up the integral on the left-hand-side into two parts:

$$\begin{aligned}\int_0^y \frac{1}{e^v + \frac{v^k}{k!}(u-1)}\, dv &= \int_0^\infty \frac{1}{e^v + \frac{v^k}{k!}(u-1)}\, dv \\ &\quad - \int_y^\infty \frac{1}{e^v + \frac{v^k}{k!}(u-1)}\, dv \\ &= C(u) - D(y,u).\end{aligned}$$

Observe that $D(y,u)$ is given asymptotically by

$$D(y,u) = e^{-y}\left(1 + O\left(e^{-y}y^k|u-1|\right)\right).$$

Hence (6.16) translates to

$$C(u) - x = e^{-y}\left(1 + O\left(e^{-y}y^k|u-1|\right)\right)$$

which can be inverted to

$$y_k(x,u) = \log\frac{1}{C(u)-x} + O\left(|u-1|\,|C(u)-x|\left(\log\frac{1}{C(u)-x}\right)^{k+1}\right).$$

Consequently it follows from Theorem 2.25 (combined with the Remark following Corollary 2.16) that $X_n^{(k)}$ satisfies a central limit theorem with mean value

$$\mathbb{E}\,X_n^{(k)} = -\frac{C'(1)}{C(1)}\,n + O\left((\log n)^{k+1}\right)$$

and variance

$$\mathbb{V}\mathrm{ar}\,X_n^{(k)} = \left(-\frac{C''(1)}{C(1)} - \frac{C'(1)}{C(1)} + \frac{C'(1)^2}{C(1)^2}\right)n + O\left((\log n)^{k+1}\right).$$

It is now an easy exercise to compute

$$C(1) = 1, \quad C'(1) = -\frac{1}{2^{k+1}}, \quad C''(1) = \frac{2}{3^{2k+1}}\frac{(2k)!}{(k!)^2}.$$

This completes the proof for recursive trees.

In the d-ary case we can work in completely the same way. The solution of the corresponding differential equation rewrites to $x - C(u) = D(y,u)$, where (for $k < d$)

$$C(u) = \int_0^\infty \frac{1}{(1+v)^d + \binom{d}{k}v^k(u-1)}\,dv$$

and

$$\begin{aligned} D(y,u) &= \int_y^\infty \frac{1}{(1+v)^d + \binom{d}{k}v^k(u-1)}\,dv \\ &= \frac{1}{d-1}\,\frac{1}{(1+y)^{d-1}}\left(1 + O\left(\frac{|u-1|}{y^{d-k}}\right)\right). \end{aligned}$$

This leads to

$$y_k(x,u) = ((d-1)(C(u)-x))^{-\frac{1}{d-1}}\left(1 + O\left(|u-1|\,|C(u)-x|^{\frac{d-k}{d-1}}\right)\right) - 1$$

which again implies a central limit theorem with expected value and variance that is asymptotically proportional to n. In particular we have

$$C(1) = \frac{1}{d-1}, \quad C'(1) = -\frac{d!(2d-k-2)!}{(2d-1)!(d-k)!}, \quad \text{and}$$
$$C''(1) = 2\binom{d}{k}^2\frac{(2k)!(3d-2k-2)!}{(3d-1)!}.$$

Since $X_n^{(d)} = n - \sum_{k<d} X_n^{(k)}$, the case $k = d$ can be reduced (more or less) to the previous cases by considering the differential equation

$$\frac{\partial y}{\partial x} = (y+1)^d + (u-1)\left((y+1)^d - y^d\right).$$

The proof in the final case, where $\Phi(w) = (1-w)^{-r}$, is slightly different, since $y(x_0)$ is finite at the radius of convergence. Here the solution $y = y_k(x,u)$ of the corresponding differential equation is determined by $x - C(u) = D(y,u)$, where

$$C(u) = \int_0^1 \frac{1}{(1-v)^{-r} + \binom{r+k-1}{k} v^k (u-1)}\, dv$$

and

$$\begin{aligned} D(y,u) &= \int_0^y \frac{1}{(1-v)^{-r} + \binom{r+k-1}{k} v^k (u-1)}\, dv \\ &= \frac{1}{r+1}(1-y)^{r+1}\left(1 + O\left(|u-1|\,|1-y|^r\right)\right). \end{aligned}$$

Hence, we get

$$y_k(x,u) = 1 - ((r+1)(C(u)-x))^{\frac{1}{r+1}} + O\left(|u-1|\;|C(u)-x|\right).$$

Consequently it follows from Theorem 2.25 that $X_n^{(k)}$ satisfies a central limit theorem with mean and variance asymptotically proportional to n. It is an easy exercise to compute

$$C(1) = \frac{1}{r+1}, \quad C'(1) = -\frac{k!\Gamma(r+k)\Gamma(2r+1)}{\Gamma(r)\Gamma(2r+k+2)} \quad \text{and}$$
$$C''(1) = \frac{(2k)!\Gamma(r+k)^2\Gamma(3r+1)}{(k!)^2\Gamma(r)^2\Gamma(3r+2k+2)}$$

which completes the proof of Theorems 6.8–6.10.

Note that in the previous discussion we have not distinguished nodes by their label. Obviously the out-degree of node n is always 0, whereas the root degree is usually much larger than the average degree. We will not go much into detail here (for a detailed analysis see Kuba and Panholzer [134]). We will only focus on the root degree R_n. By construction it follows that the derivative with respect to x of the function

$$y(x,u) = \sum_{n\ge 1} y_n \mathbb{E}\, u^{R_n} \frac{x^n}{n!}$$

is given by

$$\frac{\partial y(x,u)}{\partial x} = \Phi(uy(x)).$$

Note that the analysis is relatively simple, since

$$\frac{\partial y(x,u)}{\partial x} = \sum_{n\ge 1} y_n \mathbb{E}\, u^{R_n} \frac{x^{n-1}}{(n-1)!}.$$

We have already observed that the root degree of recursive trees is related directly to Stirling numbers of the first kind and satisfies a central limit theorem with mean value and variance asymptotically proportional to $\log n$. Thus, we just present a corresponding result for plane oriented recursive trees (and their variants defined by $\Phi(w) = 1/(1-w)^r$). For the d-ary case we get concentration at d.

Theorem 6.11. *Let R_n denote the root degree of a random increasing tree of size n, defined by the generating series $\Phi(w) = (1-w)^{-r}$, where $r > 0$. Then for every $\kappa > 0$ we have*

$$\mathbb{P}\{R_n = \lfloor \kappa n^{\frac{1}{r+1}} \rfloor\} \sim \frac{\Gamma\left(\frac{r}{r+1}\right)}{\Gamma(r)} \frac{\kappa^{r-1}}{n^{\frac{1}{r+1}}} \frac{1}{2\pi i} \int_H e^{-\kappa\cdot(-z)^{\frac{1}{r+1}} - z}\, dz, \tag{6.17}$$

where H denotes a Hankel contour, and

$$\mathbb{E}\, R_n \sim r\Gamma\left(\frac{r}{r+1}\right) n^{\frac{1}{r+1}}. \tag{6.18}$$

In particular for $r = 1$ we have

$$\mathbb{P}\{R_n = k\} = \frac{(2n-3-k)!}{2^{n-1-k}(n-1-k)!} \sim \sqrt{\frac{2}{\pi n}} e^{-k^2/(4n)}$$

and $\mathbb{E}\, R_n = \sqrt{\pi n} + O(1)$.

Proof. Since $\frac{\partial y(x,u)}{\partial x} = (1-uy(x))^{-r}$, it follows that

$$\mathbb{P}\{R_n = k\} = \frac{(n-1)!}{y_n} \binom{r+k-1}{k} [x^{n-1}] y(x)^k,$$

where $y(x) = \sum_{n\ge 1} y_n\, x^n/n! = 1-(1-(r+1)x)^{\frac{1}{r+1}}$. By using an approximate Hankel contour and the substitution $x = \frac{1}{r+1}\left(1+\frac{z}{n}\right)$ one obtains (if k is proportional to $n^{\frac{1}{r+1}}$)

$$\begin{aligned} [x^{n-1}]y(x)^k &= \frac{1}{2\pi i} \int_\gamma \left(1-(1-(r+1)x)^{\frac{1}{r+1}}\right)^k \frac{dx}{x^n} \\ &\sim \frac{(r+1)^{n-1}}{n} \frac{1}{2\pi i} \int_H e^{-kn^{-\frac{1}{r+1}}\cdot(-z)^{\frac{1}{r+1}} - z}\, dz. \end{aligned}$$

By using the asymptotic relations $\binom{r+k-1}{k} \sim k^{r-1}/\Gamma(r)$ and (6.11), we obtain (6.17).

The generating function for the expected value is $r\,y(x)(1-y(x))^{-2r}$. Thus, (6.18) follows easily.

For the case $r = 1$ we have $y(x) = 1 - \sqrt{1-2x}$, which is the solution of the equation

$$y = \frac{x}{1 - \frac{y}{2}}.$$

Hence, by Lagrange's inversion formula, we obtain explicit formulas for y_n and $[x^{n-1}]y(x)^k$, and consequently for $\mathbb{P}\{R_n = k\}$.

Another interesting parameter is the maximum degree Δ_n. For recursive trees the following precise result due to Goh and Schmutz [94] is known. (It extends previous results by Szymánski [198] and Devroye and Jiang Lu [53].)

Theorem 6.12. *The maximum degree Δ_n in random recursive trees of size n satisfies*

$$\mathbb{P}\{\Delta_n \le d\} = \exp\left(-2^{-(d-\log_2 n+1)}\right) + o(1).$$

This says that the distribution of Δ_n is highly concentrated around the value $\log_2 n$ and the distribution's behaviour is related to the extreme value distribution (or Gumbel distribution) with distribution function $F(t) = e^{-e^{-t}}$. Note that the root degree is approximately $\log n$ which is considerably smaller than the maximum degree.

The proof of Theorem 6.12 is an analytic *tour de force* which we do not present here. Nevertheless we want to give some indications why the maximum degree is concentrated around $\log_2 n$. More precisely we derive an upper bound for the expected maximum degree and indicate that a corresponding lower bound can be obtained, too.

The method we use here relies on the so-called *first moment method.*

Lemma 6.13. *Let X be a discrete random variable on non-negative integers. Then*

$$\mathbb{P}\{X > 0\} \le \min\{1, \mathbb{E}\,X\}.$$

Proof. We only have to observe that

$$\mathbb{E}\,X = \sum_{k\ge 0} k\,\mathbb{P}\{X = k\} \ge \sum_{k\ge 1} \mathbb{P}\{X = k\} = \mathbb{P}\{X > 0\}.$$

We apply this principle to the random variable $X_n^{(>k)}$ that counts the number of nodes in recursive trees with out-degree $> k$. Obviously we have

$$\Delta_n > k \iff X_n^{(>k)} > 0$$

and consequently

$$\mathbb{E}\,\Delta_n = \sum_{k\ge 0} \mathbb{P}\{\Delta_n > k\}$$

$$= \sum_{k\ge 0} \mathbb{P}\{X_n^{(>k)} > 0\}$$
$$\le \sum_{k\ge 0} \min\{1, \mathbb{E}\, X_n^{(>k)}\}.$$

Thus, if we have some (uniform) information on the expected values

$$\mathbb{E}\, X_n^{(>k)} = \sum_{j>k} \mathbb{E}\, X_n^{(j)}$$

then we obtain an upper bound for $\mathbb{E}\, \Delta_n$.

Lemma 6.14. *For random recursive trees of size n we have uniformly in $k \ge 0$*

$$\mathbb{E}\, X_n^{(k)} = \frac{n}{2^{k+1}} + O\left(\frac{1}{n}\frac{(\log n)^k}{k!}\right). \tag{6.19}$$

Proof. By Lemma 6.7 we have

$$\sum_{n\ge 1} \mathbb{E}\, X_n^{(k)} \frac{x^n}{n} = \frac{1}{k!(1-x)} \int_0^x (1-t)\left(\log\frac{1}{1-t}\right)^k dt$$
$$= \frac{1}{2^{k+1}}\frac{1}{1-x} + (x-1)\sum_{j=0}^{k} \frac{1}{j!2^{k+1-j}}\left(\log\frac{1}{1-x}\right)^j.$$

It is now an easy exercise to use the methods of Lemma 2.12 to obtain the uniform estimate

$$[x^n](x-1)\left(\log\frac{1}{1-x}\right)^k \le \frac{C}{n^2}(\log n)^k$$

for some constant $C > 0$. Hence, (6.19) follows.

Suppose that $k \ge \log_2 n$. Then

$$\sum_{j>k} \frac{(\log n)^j}{j!} = O\left(\frac{(\log n)^{k+1}}{(k+1)!}\right).$$

Consequently we get (for $k \ge \log_2 n$)

$$\mathbb{E}\, X_n^{(>k)} = \frac{n}{2^{k+1}} + O\left(\frac{1}{n}\frac{(\log n)^{k+1}}{(k+1)!}\right)$$

and also

$$\sum_{k\ge \log_2 n} \mathbb{E}\, X_n^{(>k)} = O(1).$$

Summing up we, thus, obtain

$$\begin{aligned}\mathbb{E}\,\Delta_n &\le \sum_{k\ge 0} \min\{1, \mathbb{E}\,X_n^{(>k)}\}\\ &\le \sum_{k<\log_2 n} 1 + \sum_{k\ge \log_2 n} \mathbb{E}\,X_n^{(>k)}\\ &= \log_2 n + O(1).\end{aligned}$$

In order to obtain a corresponding lower bound one can use the so-called *second moment method.*

Lemma 6.15. *Suppose that X is a non-negative random variable which is not identically zero and has finite second moment. Then*

$$\mathbb{P}\{X>0\} \ge \frac{(\mathbb{E}\,X)^2}{\mathbb{E}\,(X^2)}.$$

Proof. This follows from an application of the Cauchy-Schwarz inequality:

$$\mathbb{E}\,X = \mathbb{E}\,\left(X\cdot \mathbf{1}_{[X>0]}\right) \le \sqrt{\mathbb{E}\,(X^2)}\sqrt{\mathbb{E}\,(\mathbf{1}^2_{[X>0]})} = \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{P}\{X>0\}}.$$

In order to apply the second moment method to obtain a lower bound for $\mathbb{E}\,\Delta_n$ one needs estimates for $\mathbb{E}\,(X_n^{(>k)})^2$ which can be derived in a similar way as above. In particular it follows that

$$\mathbb{E}\,\Delta_n \ge \log_2 n + O(1)$$

and, hence, $\mathbb{E}\,\Delta_n = \log_2 n + O(1)$. However, we do not work out the details.

Nevertheless, remark that the threshold $k \approx \log_2 n$ is determined by the equation

$$\mathbb{E}\,X_n^{(>k)} \approx 1 \quad \text{resp. by} \quad \mathbb{E}\,X_n^{(k)} \approx 1.$$

Thus, in order to get a first impression where the average maximum degree is situated one should look at this equation. A rigorous treatment can then be realised with the help of the first and second moment method.

6.2.5 The Insertion Depth

The insertion depth D_n of a recursive tree is the distance of the n-th node to the root. Note that the distribution of D_n depends on the whole evolution process until step $n-1$. Nevertheless, there is a very elegant way to describe the distribution of D_n in relation to the profiles of the occurring random trees. Let $(T_1, T_2, \ldots)$ denote the sequence of random recursive trees, where in each step a new node is attached to one of the existing ones. The profile $X_{n,k}$ of T_n is the number of nodes at distance k to the root. Now the conditional probability $\mathbb{P}\{D_n = k | T_{n-1}\}$ is given by

$$\mathbb{P}\{D_n = k | T_{n-1}\} = \frac{X_{n-1,k-1}}{n-1},$$

because $D_n = k$ if the n-th node is attached to a node at distance $k-1$ in T_{n-1}. Hence, by taking expectations we obtain

$$\mathbb{P}\{D_n = k\} = \frac{\mathbb{E}\, X_{n-1,k-1}}{n-1}.$$

Therefore, the distribution of D_n is determined by the expected profile that we will calculate in the next lemma (that is due to Dondajewski and Szymánski [56], see also Meir and Moon [151]).

Lemma 6.16. *The expected profile in random recursive trees is given by*

$$\mathbb{E}\, X_{n,k} = \frac{|s_{n,k+1}|}{(n-1)!},$$

where $s_{n,k}$ denotes the of the first kind.

Proof. We present two different proofs of this property. The first proof uses the evolution process $(T_1, T_2, \ldots)$ and conditional probabilities resp. expectations. The insertion of the n-th node has the following (conditional) effect on the profile:

$$\begin{aligned}\mathbb{E}(X_{n,k}|T_{n-1}) &= \left(1 - \frac{X_{n-1,k-1}}{n-1}\right) X_{n-1,k} + \frac{X_{n-1,k-1}}{n-1}(X_{n-1,k}+1)\\ &= X_{n-1,k} + \frac{X_{n-1,k-1}}{n-1}.\end{aligned}$$

Taking the expectation this leads to the recurrence relation

$$\mathbb{E}\, X_{n,k} = \mathbb{E}\, X_{n-1,k} + \frac{\mathbb{E}\, X_{n-1,k-1}}{n-1}. \tag{6.20}$$

Hence, by using the series

$$e_n(u) = \sum_{k\ge 0} \mathbb{E}\, X_{n,k}\, u^k,$$

we derive the recurrence

$$e_n(u) = e_{n-1}(u)\left(1 + \frac{u}{n-1}\right).$$

Since $e_1(u) = 1$, we this leads to an explicit representation

$$e_n(u) = \left(1 + \frac{u}{n-1}\right)\left(1 + \frac{u}{n-2}\right)\cdots\left(1 + \frac{u}{1}\right)$$

or to

$$(n-1)!\, e_n(u) = (u+1)\cdots(u+n-1).$$

Hence, by definition

$$(n-1)!\mathbb{E}\,X_{n,k} = |s_{n,k+1}|.$$

The second proof uses the fact that a random recursive tree can be partitioned into two random recursive trees (compare with Figure 6.9). This shows that the double sequence $(X_{n,k})$ satisfies

$$X_{n,k} \stackrel{\mathrm{d}}{=} X_{I_n,k-1} + X^*_{n-I_n,k}, \tag{6.21}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution and where $(X_{n,k})$, $(X^*_{n,k})$, and (I_n) are independent, $X_{n,k} \stackrel{\mathrm{d}}{=} X^*_{n,k}$, and I_n is uniformly distributed over $\{1,2,\ldots,n-1\}$. Taking expectations this relation implies

$$\mathbb{E}\,X_{n,k} = \frac{1}{n-1}\sum_{j=1}^{n-1}\left(\mathbb{E}\,X_{j,k-1} + \mathbb{E}\,X_{j,k}\right). \tag{6.22}$$

Setting

$$F(x,u) = \sum_{n,k}\mathbb{E}\,X_{n,k}\,x^n\,u^k$$

the equation (6.22) rewrites to

$$x\frac{\partial F}{\partial x} - F = (1+u)\frac{xF}{1-x}$$

with initial condition $F(0,u)=0$. Its unique solution is given by

$$F(x,u) = x(1-x)^{-1-u} = \sum_{n\ge 1}\binom{n+u-1}{n-1}x^n = \sum_{n,k}\frac{|s_{n,k+1}|}{(n-1)!}\,x^n\,u^k$$

which again proves the formula $\mathbb{E}\,X_{n,k} = \frac{|s_{n,k+1}|}{(n-1)!}$.

We can now use the fact that Stirling numbers approximate a Gaussian distribution (compare with (6.3)) or we apply the Quasi-Power Theorem to derive the following theorem for the insertion depth.

Theorem 6.17. *The insertion depth D_n in random recursive trees satisfies a central limit theorem of the form*

$$\frac{D_n - \log n}{\sqrt{\log n}} \xrightarrow{d} N(0,1)$$

and we have

$$\mathbb{E}\,D_n = \log n + O(1) \quad \textit{and} \quad \mathbb{V}\mathrm{ar}\,D_n = \log n + O(1).$$

This theorem has a natural probabilistic interpretation.[4] Fix n and let $\xi_j = 1$ if node j appears on the path from the root to node n, and $\xi_j = 0$ if not. Evidently

$$D_n = \sum_{j=1}^{n-1} \xi_j.$$

It is a nice exercise to show that ξ_j, $1 \le j \le n-1$, are independent Bernoulli random variables with $\mathbb{P}\{\xi_j = 1\} = 1/j$ (see [54]). This implies the central limit theorem and is also consistent with the explicit relation between the probability distribution of D_n and the Stirling numbers $s_{n,k}$.

6.3 The Profile of Recursive Trees

We already observed that the expected profile of random recursive trees is given by $\mathbb{E}\,X_{n,k} = \frac{|s_{n,k+1}|}{(n-1)!}$. Due to the asymptotic properties of Stirling numbers (see Lemma 6.1) we have for $k = \alpha \log n$

$$\mathbb{E}\,X_{n,k} \sim \frac{(\log n)^k}{k!\Gamma\left(\frac{k}{\log n}+1\right)} \sim \frac{n^{\alpha(1-\log\alpha)}}{\Gamma(\alpha+1)\sqrt{2\pi\alpha\log n}} \tag{6.23}$$

and if k is close to $\log n$

$$\mathbb{E}\,X_{n,k} \sim \frac{n}{\sqrt{2\pi\log n}} e^{-\frac{(k-\log n)^2}{2\log n}}.$$

These asymptotic expansions show that almost all nodes are concentrated around the level $\log n$. This concentration property is not present for conditioned Galton-Watson trees, where the profile was of order $\sqrt{n}$ in the whole range. Therefore we cannot expect that the profile of recursive trees is approximated by a process like the local time of the Brownian excursion. Nevertheless there is a natural limiting structure if we consider the normalised profile

$$\frac{X_{n,k}}{\mathbb{E}\,X_{n,k}}.$$

Theorem 6.18. *For $0 \le \alpha \le e$ there exists a random variable $X(\alpha)$ that satisfies the equation*

$$X(\alpha) \stackrel{\mathrm{d}}{=} \alpha U^\alpha X(\alpha) + (1-U)^\alpha X(\alpha)^*$$

with $\mathbb{E}\,X(\alpha) = 1$, where $X(\alpha)$, $X(\alpha)^*$, U are independent, $X(\alpha) \stackrel{\mathrm{d}}{=} X(\alpha)^*$, and U is uniformly distributed on $[0,1]$, such that

[4] This fact was communicated to the author by Svante Janson.

$$\frac{X_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \xrightarrow{d} X(\alpha). \tag{6.24}$$

Furthermore, for $0 \le \alpha \le m^{1/(m-1)}$ *we also have convergence up to the m-th moment.*

Remark 6.19 *It is also possible to define* $(X(\alpha), 0 \le \alpha \le e)$ *as a stochastic process, where the elements* $X(\alpha)$ *are not only continuous (a.s.) but even analytic and have, thus, an analytic continuation to a proper region of the complex plane. Theorem 6.18 then extends to a weak limit law of the form*

$$\left(\frac{X_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} 0 \le \alpha \le e \right) \xrightarrow{d} (X(\alpha), 0 \le \alpha \le e)$$

(compare with [68]).

There are at least three different possible approaches to prove Theorem 6.18, namely a martingale approach, a moment approach, and a contraction method approach. All these three methods have advantages and drawbacks. In what follows we will describe all these methods, however, we mainly focus on the martingale approach (although we do not give a full treatment for the whole range of Theorem 6.18) and comment on the underlying ideas of the other two methods.

6.3.1 The Martingale Method

The martingale method is in fact the strongest method (of the three mentioned ones), since we also obtain almost sure convergence (instead of weak convergence). In what follows we prove that

$$\sup_{\frac{k}{\log n} \in K} \left| \frac{X_{n,k}}{\mathbb{E}\, X_{n,k}} - X\left(\frac{k}{\log n} \right) \right| \to 0 \quad a.s. \tag{6.25}$$

as $n \to \infty$, for all compact sets K that are contained in the open interval $((3-\sqrt{2})/2, 1.58)$. The reason why we restrict to this interval is that we apply L_2 methods. Since the limit $X(\alpha)$ has finite second moments only for $\alpha \in (0,2)$, we might extend the interval $((3-\sqrt{2})/2, 1.58)$ to the interval $(0,2)$ but we will not cover the whole interval $(0,e)$. It is possible to extend (6.25) to all compact sets $K \subseteq (0,e)$. For this purpose we would have to introduce proper continuous embeddings of recursive trees (Yule processes) which is much beyond the scope of this book (for details we refer to [38, 39] and [189]).

The martingale method makes use of certain random polynomials, the so-called *profile polynomials*

$$W_n(u) = \sum_{k \ge 0} X_{n,k}\, u^k \qquad (u \in \mathbb{C}).$$

From the proof of Lemma 6.16 we already know that the expected values are given by

$$\mathbb{E}\, W_n(u) = \frac{(u+1)\cdots(u+n-1)}{(n-1)!} = (-1)^{n-1}\binom{-u-1}{n-1}. \tag{6.26}$$

The main observation is the following martingale property. In the (similar) context of binary search trees this observation is due to Jabbour-Hattab [108].

Lemma 6.20. *The normalised profile polynomial*

$$M_n(u) = \frac{W_n(u)}{\mathbb{E}\, W_n(u)}$$

is a martingale (with respect to the natural filtration related to the tree evolution process).

Proof. We use the identity

$$\mathbb{E}(X_{n,k}|T_{n-1}) = X_{n-1,k} + \frac{X_{n-1,k-1}}{n-1}$$

from the proof of Lemma 6.16 to deduce that

$$\mathbb{E}\left(W_n(u)|T_{n-1}\right) = W_{n-1}(u)\left(1 + \frac{u}{n-1}\right).$$

Alternatively we can use the recurrence

$$W_n(u) = W_{n-1}(u) + u^{D_n}$$

and the observation

$$\mathbb{E}\left(u^{D_n}|T_{n-1}\right) = \sum_{k\ge 1} \frac{X_{n-1,k-1}}{n-1} u^k = \frac{uW_{n-1}(u)}{n-1}. \tag{6.27}$$

Thus, normalising by the expected profile (6.26) leads to

$$\mathbb{E}\left(M_n(u)|T_{n-1}\right) = M_{n-1}(u),$$

which proves the martingale property.

If u is real and non-negative, then $M_n(u)$ is a non-negative martingale. Hence, there exits $M(u)$ which is the a.s. limit:

$$M_n(u) \to M(u) \quad a.s.$$

The idea of the martingale method is to first show that this convergence also holds in a certain complex domain and then to use Cauchy's formula to derive a limit relation for $X_{n,k}$ from the limit relation $W_n(u) = M_n(u)\,\mathbb{E}\, W_n(u) \sim M(u)\,\mathbb{E}\, W_n(u)$.

Accordingly our first goal is to extend the convergence property of the martingale $M_n(u)$ to complex u.

Proposition 6.21. *For any compact set $C \subseteq \{u \in \mathbb{C} : |u-1| < 1\}$ the martingale $M_n(u)$ converges a.s. uniformly to its limit $M(u)$.*

Note that $M_n(u)$ can be considered as a random analytic function. Hence, the limit $M(u)$ is a random analytic function, too.

We start by establishing an explicit formula for the covariance function of $(W_n(u_1), W_n(u_2))$ which is valid for all $u_1, u_2 \in \mathbb{C}$.

Lemma 6.22. *For all $u_1, u_2 \in \mathbb{C}$:*

$$\mathbb{E}\,(W_n(u_1)W_n(u_2)) = \sum_{j=1}^{n-1}\left(\beta_j(u_1,u_2)\prod_{k=j+1}^{n-1}\alpha_k(u_1,u_2)\right) + \prod_{j=1}^{n-1}\alpha_j(u_1,u_2),$$

where

$$\alpha_k(u_1,u_2) = 1 + \frac{u_1+u_2}{k} \tag{6.28}$$

and

$$\beta_k(u_1,u_2) = u_1u_2\frac{\mathbb{E}\,W_k(u_1u_2)}{k}. \tag{6.29}$$

Proof. By using the relation $W_n(u) = W_{n-1}(u) + u^{D_n}$ and (6.27) we get the linear recurrence

$$\mathbb{E}\,(W_n(u_1)W_n(u_2)) = \alpha_{n-1}(u_1,u_2)\mathbb{E}\,(W_{n-1}(u_1)W_{n-1}(u_2)) + \beta_{n-1}(u_1,u_2). \tag{6.30}$$

Now, the explicit formula for $\mathbb{E}\,(W_n(u_1)W_n(u_2))$ follows from (6.30) (and $\mathbb{E}\,(W_1(u_1)W_1(u_2)) = 1$).

Using Lemma 6.22 we establish regularity of the covariance function of M over $\mathcal{U}\times\mathcal{U}$, where $\mathcal{U} = \{u \in \mathbb{C} : |u-1| < 1\}$.

Corollary 6.23 $(M_n(u))_{n\in\mathbb{N}}$ *is bounded in L_2, if and only if $|u-1| < 1$. Hence, there exists a random variable $M(u) \in L_2$ such that $M_n(u) \to M(u)$ almost surely and in L_2 for $u \in \mathcal{U} = \{u \in \mathbb{C} : |u-1| < 1\}$. Furthermore, the covariance function $\Gamma(u_1,u_2) = \mathbb{E}\,(M(u_1)M(u_2))$ is holomorphic in $\mathcal{U}\times\mathcal{U}$.*

Proof. By (6.28) we have

$$\prod_{k=j+1}^{n-1}\alpha_k(u_1,u_2) = \left(\frac{n}{j}\right)^{u_1+u_2}\left(1 + O\left(\frac{1}{j}\right)\right)$$

and consequently

$$\mathbb{E}\,(W_n(u_1)W_n(u_2)) = u_1u_2\sum_{j=1}^{n-1}\frac{\mathbb{E}\,W_j(u_1u_2)}{j}\prod_{k=j+1}^{n-1}\alpha_k(u_1,u_2) + \prod_{j=1}^{n-1}\alpha_j(u_1,u_2)$$

$$= O\left(\sum_{j=1}^{n-1} j^{\Re(u_1 u_2)-1}\left(\frac{n}{j}\right)^{\Re(u_1+u_2)} + n^{2\Re(u_1+u_2-1)}\right)$$

$$= O\left(n^{\Re(u_1+u_2)}\sum_{j=1}^{n-1} j^{-\Re(u_1+u_2-u_1u_2+1)}\right).$$

Thus,

$$\begin{aligned}\Gamma_n(u_1,u_2) &:= \mathbb{E}\,(M_{n+1}(u_1)M_{n+1}(u_2)) \\ &= \frac{\mathbb{E}\,(W_{n+1}(u_1)W_{n+1}(u_2))}{\mathbb{E}\,W_{n+1}(u_1)\cdot\mathbb{E}\,W_{n+1}(u_2)} \\ &= O\left(\sum_{j=1}^{n-1} j^{-\Re(u_1+u_2-u_1u_2+1)}\right).\end{aligned}$$

Obviously, we have the same lower bound. Hence, $(M_n(u))_{n\in\mathbb{N}}$ is bounded in L_2, if and only if $2\Re u - |u|^2 > 0$, respectively if and only if $u \in \mathcal{U}$.

Now, if $2\Re u_1 - |u_1|^2 > 0$ and $2\Re u_2 - |u_2|^2 > 0$ then we also have $\Re(u_1+u_2-u_1u_2) > 0$. Thus, $\Gamma_n(u_1,u_2) \to \Gamma(u_1,u_2)$ is uniformly over the compact sets of $\mathcal{U}\times\mathcal{U}$. Since for any n, Γ_n is holomorphic we conclude that Γ is holomorphic in $\mathcal{U}\times\mathcal{U}$.

The holomorphy of Γ will give us (with the help of the Kolmogoroff criterion) continuity of $M(u)$ over any parametrised arc $\gamma \subseteq \mathcal{U}$. However, Kolmogoroff's criterion is not sufficient to establish directly a continuity of M as a complex function.

Lemma 6.24. *Set $I' := (0,2)$. Then $(M(t))_{t\in I'}$ has a continuous modification $\tilde{M}$ such that, for any compact interval $K \subseteq I'$,*

$$\mathbb{E}\left(\sup_{t\in K}|\tilde{M}(t)|^2\right) < +\infty.$$

More generally, if $\gamma : \mathbb{R} \to \mathcal{U}$ is continuously differentiable, then there is s a modification $\tilde{M}_\gamma$ of $(M_n(\gamma(t)))_{t\in\mathbb{R}}$ such that, for any compact set K of $\mathbb{R}$,

$$\mathbb{E}\left(\sup_{t\in K}|\tilde{M}_\gamma(t)|^2\right) < +\infty.$$

Proof. Observe that, as $M_n(u)$ is a rational function with real coefficients which implies $\overline{M_n(u)} = M_n(\overline{u})$. Thus for all $u_1, u_2 \in \mathcal{U}$

$$\mathbb{E}\left(|M(u_1) - M(u_2)|^2\right) = \Gamma(u_1,\overline{u_1}) + \Gamma(u_2,\overline{u_2}) - 2\Re\left(\Gamma(u_1,\overline{u_2})\right). \quad (6.31)$$

Let C be a compact set of $\mathcal{U}$; since Γ is holomorphic, a local expansion of Γ up to order 2 yields

$$\Gamma(u_1, \overline{u_1}) + \Gamma(u_2, \overline{u_2}) - 2\Re\left(\Gamma(u_1, \overline{u_2})\right) \le C|u_1 - u_2|^2 \tag{6.32}$$

for some constant $C > 0$ and for all $u_1, u_2 \in K$. Hence, by (6.31) and (6.32)

$$\mathbb{E}\left(|M(u_1) - M(u_2)|^2\right) \le C|u_1 - u_2|^2 \tag{6.33}$$

for all $u_1, u_2 \in K$. Hence by Kolmogoroff's criterion (cf. [183, p. 25]), a continuous modification $\tilde{M}$ exists and

$$\mathbb{E}\left[\left(\sup_{s,t\in K} \frac{|\tilde{M}_t - \tilde{M}_s|}{|t-s|^\alpha}\right)^2\right] < +\infty$$

for all $\alpha \in (0, \frac{1}{2})$. Consequently, for all compact set $K \subseteq (0,2)$, we have

$$\mathbb{E}\left(\sup_{t\in K} |\tilde{M}(t)|^2\right) < +\infty.$$

Now let $\gamma : \mathbb{R} \to \mathcal{U}$ be continuously differentiable. We can do the same as before with the martingales $(M_{n,\gamma}(t))_{t\in\mathbb{R}}$ for $M_{n,\gamma}(t) = M_n(\gamma(t))$. Equation (6.33) becomes

$$\mathbb{E}\left(|M_\gamma(t_1) - M_\gamma(t_2)|^2\right) \le K|\gamma(t_1) - \gamma(t_2)|^2 \le C'|t_1 - t_2|^2$$

for some constant $C' > 0$ depending on the compact interval $K \subseteq \mathbb{R}$ under consideration. Thus, $(M_\gamma(t))_{t\in\mathbb{R}}$ has a continuous modification $\tilde{M}_\gamma$ such that $\mathbb{E}\left(\sup_{t\in K} |\tilde{M}_\gamma(t)|^2\right) < +\infty$ for all compact set $K \subseteq \mathbb{R}$.

Now uniform convergence of (M_n) follows from a theorem of vectorial martingales. (We proceed as in [118].)

Lemma 6.25. *For any compact set $K \subseteq (0,2)$, we have a.s.*

$$M_n \to M \qquad \textit{uniformly over } K$$

and

$$\lim_{n\to\infty} \mathbb{E}\left(\sup_{t\in K} |M_n(t) - M(t)|^2\right) = 0.$$

More generally, let $\gamma : \mathbb{R} \to \mathcal{U}$ be continuously differentiable and set $M_{n,\gamma}(t) = M_n(\gamma(t))$ and $M_\gamma(t) = M(\gamma(t))$. Then the same result holds for $(M_{n,\gamma})$.

Proof. Let $[a,b] \subseteq (0,2)$. The modification $\tilde{M}$ of the previous proposition is a random variable taking its values in the separable Banach space $E = \mathrm{C}_\mathbb{C}([a,b])$. Let $\mathcal{E}$ be the Borelian σ-field of E and $\mathcal{F}_\infty = \sigma(T_n, n \ge 1)$, and $\tilde{M}$ is $\mathcal{E}|\mathcal{F}_\infty$-measurable and is in $L_2(E) = L_2(\Omega, E)$.

We will show that $\mathbb{E}(\tilde{M}|T_n)$ can be identified as $M_{n|[a,b]}$ if $M_{n|[a,b]}$ is understood as a random variable taking its values in E.
Observe that

$$\varphi_t : \mathrm{C}_{\mathbb{C}}([a,b]) \to \mathbb{C},\ X \mapsto X(t)$$

is a continuous linear form over $\mathrm{C}_{\mathbb{C}}([a,b])$, hence $\mathbb{E}\,(\varphi_t(\tilde{M})|T_n) = \varphi_t(\mathbb{E}\,(\tilde{M}|T_n))$ almost surely. Saying that $\tilde{M}$ is a modification of M means that for all $t \in [a,b]$, $\varphi_t(\tilde{M}) = M(t)$ (a.s.). Hence it follows that $M_n(t) = \mathbb{E}\,(\tilde{M}|T_n)(t)$ a.s., so that $M_n = \mathbb{E}\,(\tilde{M}|T_n)$ a.s.

We can now apply the theorem of vectorial martingales (see [164, Proposition V-2-6, p104]), which yields

$$M_n = \mathbb{E}\left(\tilde{M}|T_n\right) \to \mathbb{E}\left(\tilde{M}|\mathcal{F}_\infty\right) \quad \text{a.s. and in } L_2(E).$$

Since $\tilde{M}$ is $\mathcal{F}_\infty$-measurable we get, as $n \to \infty$,

$$\sup_{t\in[a,b]} |M_n(t) - \tilde{M}(t)| \to 0 \quad \text{a.s.} \tag{6.34}$$

and

$$\mathbb{E}\left(\sup_{t\in[a,b]} |M_n(t) - \tilde{M}(t)|^2\right) \to 0. \tag{6.35}$$

Hence, the first part of the lemma is proved since (6.34) implies that, almost surely, $M(t)$ exists for all $t \in [a,b]$ and is equal to $\tilde{M}(t)$.

By the previous proposition we can proceed for $(M_{n,\gamma})$ along the same lines as for (M_n). This completes the proof.

Now, since M_n is holomorphic, for all n and any $\rho < 1$, the uniform convergence of M_n over the arc $\gamma(t) = 1+\rho e^{it}$, implies (via Cauchy's formula) uniform convergence of M_n and all its derivatives over compact subsets of $\mathcal{U}$.

In order to apply Cauchy's formula to obtain information on the profile $X_{n,k}$ it is, however, not sufficient to know the behaviour of $M_n(u)$ near the real axis. We need more precise information about the limiting behaviour of $M_n(u)$ for $0 < |u| < 2$, resp. of $W_n(u)$. The next proposition provides a.s. estimates for $u \in \mathbb{C}$ also away from the real axis.

Proposition 6.26. *For any $K > 0$ there exists $\delta > 0$ such that a.s.*

$$W_n(u) = O\left(\frac{n^{|u|}}{(\log n)^K}\right)$$

uniformly for $u \in \mathbb{C}$ with $(3-\sqrt{2})/2 + \epsilon \le |u| \le 1.58$, $|u-1| \ge 1-\delta$ as $n \to \infty$.

Before we prove Proposition 6.26 we state a corollary.

Corollary 6.27 *For any $K > 0$ and $\epsilon > 0$ we have a.s. that there exists n_0 such that for all $n \ge n_0$*

$$|W_n(u)| \le \frac{\mathbb{E}\,W_n(|u|)}{(\log n)^K}$$

for all $u \in \mathbb{C}$ with $(3-\sqrt{2})/2 + \epsilon \le |u| \le 1.58$ and $(\log n)^{-\frac{1}{2}+\epsilon} \le |\arg u| \le \pi$ as $n \to \infty$.

Proof. By Proposition 6.26 this estimate is true for $u \in \mathbb{C}$ with $(3-\sqrt{2})/2+\epsilon \le |u| \le 1.58$ and $|u-1| \ge 1-\delta$. Moreover, for $u \in \mathbb{C}$ with $|u-1| \le 1-\delta$ we know that $M_n(u)$ is a.s. bounded. Furthermore we have uniformly in n and t for $(\log n)^\epsilon/\sqrt{\log n} \le |t| \le \pi$

$$|\mathbb{E}\, W_n(u_0 e^{it})| \le \mathbb{E}\, W_n(u_0) e^{-ct^2 \log n}$$

for some constant $c > 0$ (depending continuously on u_0). A combination of these two estimates proves the corollary.

For the proof of Proposition 6.26 we need an estimate for $\mathbb{E}\,|W_n(u)|^2$.

Lemma 6.28. *For every $\delta > 0$ we have uniformly for u with $|u-1| \le 1-\delta$*

$$\mathbb{E}\,|W_n(u)|^2 = O(n^{2\Re u})$$

and for u with $1-\delta \le |u-1| \le 1$

$$\mathbb{E}\,|W_n(u)|^2 = O(n^{2\Re u}\log n)$$

as $n \to \infty$. Let K be a compact set in the complex plane such that $|u-1| \ge 1$ for all $u \in C$. Then

$$\mathbb{E}\,|W_n(u)|^2 = O(n^{|u|^2}\log n)$$

as $n \to \infty$, uniformly for $u \in K$.

Proof. We recall that

$$\mathbb{E}\,|M_n(u)|^2 = O\left(n^{2\Re u}\sum_{j=1}^{n-1} j^{-(2\Re u - |u|^2+1)}\right).$$

We also have $2\Re u - |u|^2 > 0$ for $|u-1| < 1$ and $2\Re u - |u|^2 < 0$ for $|u-1| > 1$. Thus, for $|u-1| \le 1-\delta$, there exists $\delta' > 0$ such that

$$\mathbb{E}\,|W_n(u)|^2 = O\left(n^{2\Re u}\sum_{j=1}^{n} j^{-1-\delta'}\right) = O\left(n^{2\Re u}\right),$$

and for $1-\delta \le |u-1| \le 1$

$$\mathbb{E}\,|W_n(u)|^2 = O\left(n^{2\Re u}\sum_{j=1}^{n} j^{-1}\right) = O\left(n^{2\Re u}\log n\right),$$

which prove the first part of the lemma.

Finally, for u with $|u-1| > 1$ we obtain

$$\mathbb{E}\,|W_n(u)|^2 = O\left(n^{|u|^2}\log n\right).$$

This completes the proof of the lemma.

We will also use an a.s. estimate for the derivative of $W_n(u)$.

Lemma 6.29. *We have for all u with $|u| < 2$*

$$|W_n'(u)| \le W_n'(|u|) = O\left(|u|^{-1} n^{|u|} \log n\right) \quad a.s.$$

Proof. Obviously, we have $|W_n'(u)| \le W_n'(|u|)$. It is also known that $H_n \sim e \log n$ a.s. (compare with Section 6.4 and [175]). Hence, $X_{n,k} = 0$ a.s. if $k > (e+1)\log n$. This implies a.s.

$$W_n'(|u|) = \sum_{k\ge 0} k X_{n,k} |u|^{k-1} \le (e{+}1)\log n \sum_{k\ge 0} X_{n,k}|u|^{k-1} = (c{+}1)\log n \frac{W_n(|u|)}{|u|}.$$

Since $W_n(|u|) \sim M_n(|u|)\mathbb{E}\, W_n(|u|) = O\left(n^{|u|}\right)$, this proves the lemma.

We are now ready to work out the Proof of Proposition 6.26. By Lemma 6.28 we have for $(3-\sqrt{2})/2 < |u| \le 1.58$ and $|u-1| \ge 1-\delta$ (for some sufficiently small $\delta > 0$)

$$\mathbb{P}\left\{|W_n(u)| \ge \frac{n^{|u|}}{(\log n)^K}\right\} \le \frac{\mathbb{E}\,|W_n(u)|^2}{(n^{|u|}/(\log n)^K)^2} = O\left(\frac{(\log n)^{2K+1}}{n^{2|u|-|u|^2}}\right).$$

First, let us consider the range $R_1 := \{u \in \mathbb{C} : |u-1| \ge 1-\delta, 1 \le |u| \le 1.58\}$. Set $\alpha = 1/(1.581(2-1.581))$. Then we have

$$\alpha|u|(2-|u|) > 1 \quad \text{and} \quad \alpha - 1 + \alpha e \log |u| < \alpha |u|$$

for all u with $1 \le |u| \le 1.58$.

We now use $O((\log n)^{2K+2})$ points $u_{n,j}$ covering R_1 with maximal distance $(\log n)^{-K-1}$. Observe that for $u \in R_1$ the series

$$\sum_{n\ge 1} \frac{(\log[n^\alpha])^{2K+1}}{[n^\alpha]^{2|u|-|u|^2}}$$

converges uniformly. Hence, by the Borel-Cantelli-Lemma we have a.s.

$$\sup_j |W_{[n^\alpha]}(u_{[n^\alpha],j})| \le \frac{[n^\alpha]^{|u|}}{(\log[n^\alpha])^K}$$

for all but finitely many n. By using Lemma 6.29 we obtain a.s.

$$|W_{[n^\alpha]}(u) - W_{[n^\alpha]}(v)| = O\left([n^\alpha]^{\max\{|u|,|v|\}} \log n\right)$$

which implies that we can interpolate between $u_{[n^\alpha],j}$ and obtain a.s.

$$\sup_{u\in R_1} |W_{[n^\alpha]}(u)| \le \frac{[n^\alpha]^{|u|}}{(\log[n^\alpha])^K}$$

for all but finitely many n.

Finally, we have to observe that a.s., uniformly for $n^\alpha \le k \le (n+1)^\alpha$

$$|W_k(u) - W_{[n^\alpha]}(u)| = O\left(\frac{[n^\alpha]^{2|u|-1}}{(\log[n^\alpha])^K}\right). \qquad (6.36)$$

Since

$$W_{n+k}(u) - W_n(u) = \sum_{l=n+1}^{n+k} u^{D_l},$$

we have to estimate u^{D_l}. We know that a.s. $H_n \le e \log n$ (for sufficiently large n). Hence it follows that a.s.

$$\max_{n^\alpha \le l \le (n+1)^\alpha} D_l \le e \cdot \log(n^\alpha).$$

So we only have to check that

$$|u| \left([(n+1)^\alpha] - [n^\alpha]\right) |u|^{e \log(n^\alpha)} = O\left(\frac{[n^\alpha]^{|u|}}{(\log[n^\alpha])^K}\right).$$

Alternatively, it suffices to show that there exists $\eta > 0$ such that

$$\alpha - 1 + \alpha e \log|u| \le \alpha|u| - \eta$$

for $1 \le |u| \le 1.58$. However, this is true for the above choice of α. Hence, (6.36) follows, which completes the proof for $u \in R_1$.

For $u \in R_2 := \{u \in \mathbb{C} : |u-1| \ge 1/-\delta, (3-\sqrt{2})/2 + \epsilon \le |u| \le 1\}$ we have to do almost the same. Again we use a subsequence $[n^\beta]$ to apply the Borel-Cantelli-Lemma. We set $\beta = (1+\sqrt{5})/2$ and observe that

$$\beta|u|(2-|u|) > 1 \quad \text{and} \quad \beta - 1 < \beta|u|$$

for all u with $(3-\sqrt{2})/2 + \epsilon \le |u| \le 1$. This completes the proof of Proposition 6.26.

The proof of (6.25) can now be completed in the following way. In particular we show that the limit $X(\alpha)$ coincides with the limit $M(\alpha)$. For small t we use the a.s. expansion

$$M_n(u_0 e^{it}) = M_n(u_0) e^{itM_n'(u_0)/M_n(u_0) + O(t^2)}$$

and $W_n(u_0 e^{it}) = M_n(u_0 e^{it}) \mathbb{E}\, W_n(u_0 e^{it})$. For large t we use the estimate of Proposition 6.26 (resp. of its Corollary 6.27).

By Cauchy's formula we have

$$X_{n,k} = \frac{u_0^{-k}}{2\pi} \int_{-\pi}^{\pi} W_n(u_0 e^{it}) e^{-kit}\, dt,$$

where we will use $u_0 = k/\log n$. First, for any (sufficiently small) $\eta > 0$ we have by Corollary 6.27

$$|W_n(u_0 e^{it})| \le \frac{\mathbb{E}\, W_n(u_0)}{\log n} \quad a.s.$$

for $(3-\sqrt{2})/2 + \epsilon \le u_0 \le 1.58$ and $(\log n)^{-\frac{1-\eta}{2}} \le |t| \le \pi$. Hence

$$\left| \frac{u_0^{-k}}{2\pi} \int_{(\log n)^{-\frac{1-\eta}{2}} \le |t| \le \pi} W_n(u_0 e^{it}) e^{-kit}\, dt \right| = O\left(\frac{\mathbb{E}\, W_n(u_0)}{u_0^k \log n} \right).$$

Conversely, for real t with $|t| \le (\log n)^{-\frac{1-\eta}{2}}$ we have uniformly

$$\begin{aligned} W_n(u_0 e^{it}) e^{-kit} &= M_n(u_0 e^{it}) \mathbb{E}\, W_n(u_0 e^{it}) e^{-kit} \\ &= M_n(u_0) \mathbb{E}\, W_n(u_0) e^{2u_0 \log n (e^{it}-1) - kit} (1 + O(|t|) + O(n^{-1})) \\ &= M_n(u_0) \mathbb{E}\, W_n(u_0) e^{-u_0 \log n\, t^2} (1 + O((\log n)^{-\frac{1-3\eta}{2}})). \end{aligned}$$

This implies that a.s.

$$\begin{aligned} X_{n,k} &= \frac{u_0^{-k}}{2\pi} \int_{|t| \le (\log n)^{-\frac{1-\eta}{2}}} W_n(u_0 e^{it}) e^{-kit}\, dt + O\left(u_0^{-k} \frac{\mathbb{E}\, W_n(u_0)}{\log n} \right) \\ &= M_n(u_0) u_0^{-k} \sqrt{2\pi k}\, \mathbb{E}\, W_n(u_0) (1 + O((\log n)^{-\frac{1-3\eta}{2}})). \end{aligned}$$

Since we uniformly have

$$u_0^{-k} \sqrt{2\pi k}\, \mathbb{E}\, W_n(u_0) \sim \mathbb{E}\, X_{n,k}$$

and $M_n(u_0) \sim M(u_0)$, the result follows.

6.3.2 The Moment Method

Suppose that $0 \le \alpha \le 1$. We indicate a proof of (6.24) in this range by showing that for all integers $k \ge 1$

$$\lim_{n\to\infty} \mathbb{E} \left(\frac{X_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \right)^k = \mathbb{E} X(\alpha)^k.$$

This *moment method* only works for $0 \le \alpha \le 1$, since $X(\alpha)$ does not have moments of arbitrary order if $\alpha > 1$ (compare with [87]).

Let $P_{n,k}(u) = \mathbb{E}\, u^{X_{n,k}}$ denote the probability generating function of $X_{n,k}$. Then the recurrence (6.21) implies that

$$P_{n,k}(u) = \frac{1}{n-1} \sum_{j=1}^{n-1} P_{j,k-1}(u) P_{n-j,k}(u), \qquad (n \ge 2,\ k \ge 1),$$

with $P_{n,0}(u) = u$ for $n \ge 1$ and $P_{0,k} = 1$. Taking derivatives and setting $u = 1$ yields the recurrence

$$\mathbb{E}\, X_{n,k} = \frac{1}{n-1} \sum_{j=1}^{n-1} (\mathbb{E}\, X_{j,k-1} + \mathbb{E}\, X_{n-j,k})$$

with $\mathbb{E}\, X_{n,0} = 1$ for $n \ge 1$ and $\mathbb{E}\, X_{0,k} = 0$. This recurrence is different from (6.20) but can be extended to higher moments.

By taking m-th derivatives we are led to the recurrence

$$A_{n,k}^{(m)} = \frac{1}{n-1} \sum_{j=1}^{n-1} \left(A_{j,k-1}^{(m)} + A_{n-j,k}^{(m)} \right) + B_{n,k}^{(m)}$$

with $A_{n,0}^{(1)} = 1$ for $n \ge 1$ and $A_{n,0}(m) = 0$ for $m \ge 2$ and $n \ge 1$, where

$$A_{n,k}^{(m)} = \mathbb{E}\, (X_{n,k}(X_{n,k} - 1) \cdots (X_{n,k} - m + 1)) = P_{n,k}^{(m)}(1)$$

and where $B_{n,k}^{(m)}$ abbreviates

$$B_{n,k}^{(m)} = \sum_{h=1}^{m-1} \binom{m}{h} \frac{1}{n-1} \sum_{j=1}^{n-1} A_{j,k-1}^{(h)} A_{n-j,k}^{(m-h)}.$$

Hence, for every fixed $m \ge 1$ the sequence $(A_{n,k}^{(m)})$ satisfies a recurrence of the form

$$a_{n,k} = \sum_{j=1}^{n-1} (a_{j,k-1} + a_{n-j,k}) + b_{n,k}, \qquad (n \ge 2,\, k \ge 1), \tag{6.37}$$

with given initial values $a_{1,k}$ and a given sequence $b_{n,k}$. For the interest of consistency we set $b_{1,k} = a_{1,k}$. Then (6.37) holds also for $n = 1$. Remarkably the recurrence (6.37) has an explicit solution of the following form (see [87]).

Lemma 6.30. *Suppose that the sequence $a_{n,k}$ satisfies the recurrence (6.37). Then we have (for $n \ge 1$ and $k \ge 0$)*

$$a_{n,k} = b_{n,k} + \sum_{j=1}^{n-1} \sum_{r=0}^{k} \frac{b_{j,k-r}}{j} \, [u^r] \, (u+1) \prod_{j<\ell<k} \left(1 + \frac{u}{\ell}\right).$$

Proof. Setting $a_n(u) = \sum_k a_{n+1,k} u^k$ and $b_n(u) = \sum_k b_{n+1,k} u^k$ the recurrence (6.37) rewrites to

$$a_n(u) = \frac{1+u}{n} \sum_{j=0}^{n-1} a_j(u) + b_n(u) \qquad (n \ge 1)$$

with initial condition $a_0(u) = \sum_k a_{1,k}u^k$. By taking the difference $na_n(u) - (n-1)a_{n-1}(u)$ we thus obtain

$$a_n(u) = \left(1 + \frac{u}{n}\right) a_{n-1}(u) + b_{n-1}(u) - \frac{n-1}{n} b_{n-1}(u), \qquad (n \ge 2),$$

and consequently

$$a_n(u) = b_n(u) + (1+u) \sum_{j=0}^{n-1} \frac{b_j(u)}{j+1} \prod_{j+1 \le \ell \le n} \left(1 + \frac{u}{\ell}\right).$$

Taking the coefficient of u^k on both sides provides the result.

For example, in the case $m = 1$ we have $b_{n,k} = \delta_{n,1}\delta_{0,k}$ for $n \ge 1$ and $k \ge 0$. Thus, we recover the explicit (and also the asymptotic) formula for the expected value $A_{n,k}^{(1)} = \mathbb{E}\, X_{n,k}$:

$$\mathbb{E}\, X_{n,k} = [u^k] \prod_{1 \le \ell < n} \left(1 + \frac{u}{j}\right) = \frac{|s_{n,k+1}|}{(n-1)!} \sim \frac{(\log n)^k}{k!\Gamma\left(1 + \frac{k}{\log n}\right)}.$$

Starting from this solution one can iteratively apply Lemma 6.30 to obtain asymptotic expansions for $A_{n,k}^{(m)}$ for $m \ge 2$.

Proposition 6.31. *For* $0 \le \alpha \le m^{1/(m-1)}$ *set* $\nu_0(\alpha) = \nu_1(\alpha) = 1$ *and recursively (for* $m \ge 2$*)*

$$\nu_m(\alpha) = \frac{1}{m - \alpha^{m-1}} \sum_{h=1}^{m-1} \binom{m}{h} \nu_h(\alpha)\nu_{m-h}(\alpha) \frac{\Gamma(h\alpha+1)\Gamma((m-h)\alpha+1)}{\Gamma(m\alpha+1)}$$

Then for every fixed $0 \le \alpha \le m^{1/(m-1)}$ *we have, as* $n \to \infty$,

$$\mathbb{E}\, X^m_{n,\lfloor \alpha \log n \rfloor} \sim \nu_m(\alpha) \left(\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}\right)^m \tag{6.38}$$

It is easy to check that for every $0 \le \alpha \le 1$ the sequence $\nu_m(\alpha)$ is the moment sequence of a random variable $X(\alpha)$. From the contraction method (see Chapter 8) it follows that (6.41) has a unique solution $X(\alpha)$ (with $\mathbb{E}\, X(\alpha) = 1$) and by definition the sequence $\nu_m(\alpha)$ is the moment sequence of $X(\alpha)$. By induction it follows that $\nu_m(\alpha)\Gamma(m\alpha+1)/m! = O(K^m)$ (for some $K > 0$, see [107]). Hence, the sequence $\nu_m(\alpha)$ determines uniquely the distribution of $X(\alpha)$. Consequently, for $0 \le \alpha \le 1$, Proposition 6.31 implies the limit relation (6.24).

If $\alpha > 1$ then only few moments exist. More precisely if $\alpha \le m^{1/(m-1}$ then $X(\alpha)$ has moments up to order m (and we have convergence of moments up to this order) but the $(m+1)$-st moment is infinite.

The proof of Proposition 6.31 is very technical. Following [87] one starts with a uniform upper bound for the expected profile of the form

$$A_{n,k}^{(1)} = O\left((v\log n)^{-1/2} v^{-k} n^{v}\right), \qquad (0 \le k < n),$$

where $0 < v \le v_0$. If k is of logarithmic order then this bound is optimal, since we can set $v = k/\log n$. Then Lemma 6.30 implies inductively

$$A_{n,k}^{(m)} = O\left(\frac{1}{m - v^{m-1}}\left((v\log n)^{-1/2} v^{-k} n^{v}\right)^{m}\right),$$

uniformly for $0 \le k < n$, where $0 < v < m^{1/(m-1)}$.

By using this a priori bound Proposition 6.31 follows by induction on m. More precisely one shows that for every integer $m \ge 1$

$$A_{n,k}^{(m)} \sim \nu_m(\alpha)\left(\frac{(\log n)^k}{k!\Gamma(\alpha+1)}\right)^{m}, \tag{6.39}$$

where $k = \lfloor \alpha \log n \rfloor$. We just observed that this is true for $m = 1$. By assuming that (6.39) holds all integers $< m$ and by using the definition of $\nu_m(\alpha)$ one obtains

$$B_{n,k}^{(m)} \sim \nu_m(\alpha)\frac{m\alpha - \alpha^m}{m\alpha + 1}\left(\frac{(\log n)^k}{k!\Gamma(\alpha+1)}\right)^{m}.$$

Furthermore, from

$$[u^r]\prod_{j<\ell<n}\left(1 + \frac{u}{\ell}\right) = \frac{(\log(n/j))^r}{j}\left(1 + O(r^2/j)\right)$$

(for $\epsilon n \le j \le (1-\epsilon)n$ and $0 \le r \le k = o(\sqrt{j})$) it follows that

$$\begin{aligned}
&\sum_{j=1}^{n-1}\sum_{r=0}^{k}\frac{B_{j,k-r}^{(m)}}{j}[u^r](u+1)\prod_{j<\ell<k}\left(1+\frac{u}{\ell}\right)\\
&\qquad = \mu_m(\alpha)\frac{m\alpha - \alpha^m}{m\alpha+1}\left(\alpha^m + 1\right)\left(\frac{(\log n)^k}{k!\Gamma(\alpha+1)}\right)^{m}\int_0^1 x^{m\alpha - \alpha^m - 1}\,dx\\
&\qquad = \mu_m(\alpha)\frac{\alpha^m + 1}{m\alpha+1}\left(\frac{(\log n)^k}{k!\Gamma(\alpha+1)}\right)^{m}
\end{aligned}$$

Hence, (6.39) holds for m, too.

Finally we have

$$\mathbb{E}\, X_{n,\lfloor \alpha\log n\rfloor}^{m} \sim A_{n,\lfloor \alpha\log n\rfloor}^{(m)}$$

which implies (6.38) for all $m \ge 1$.

6.3.3 The Contraction Method

The contraction method makes use of the fact that there is a fixed point equation for the (expected) limiting distribution.

Assume for a moment that we already know a limiting relation for the profile $X_{n,k}$ of the form $X_{n,\lfloor \alpha \log n \rfloor}/\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor} \xrightarrow{d} X(\alpha)$. Motivated by that we rewrite the recurrence (6.21) in the form:

$$\frac{X_{n,k}}{\mathbb{E}\, X_{n,k}} \stackrel{\mathrm{d}}{=} \frac{\mathbb{E}\, X_{I_n,k-1}}{\mathbb{E}\, X_{n,k}} \frac{X_{I_n,k-1}}{\mathbb{E}\, X_{I_n,k-1}} + \frac{\mathbb{E}\, X^*_{n-I_n,k}}{\mathbb{E}\, X_{n,k}} \frac{X^*_{n-I_n,k}}{\mathbb{E}\, X^*_{n-I_n,k}}. \tag{6.40}$$

Recall that $X^*_{n,k}$ is an independent copy of $X_{n,k}$ and that $(X_{n,k})$, $(X^*_{n,k})$, and (I_n) are independent. By (6.23) we have for $k = \lfloor \alpha \log n \rfloor$

$$\frac{\mathbb{E}\, X_{I_n,k-1}}{\mathbb{E}\, X_{n,k}} \sim \frac{k}{\log n} \left(\frac{\log I_n}{\log n} \right)^{k-1} \sim \alpha \left(\frac{I_n}{n} \right)^{\alpha},$$

and

$$\frac{\mathbb{E}\, X^*_{n-I_n,k}}{\mathbb{E}\, X_{n,k}} \sim \left(1 - \frac{I_n}{n} \right)^{\alpha}.$$

Hence, by (formally) taking the limit $n \to \infty$ and by observing that $I_n/n \xrightarrow{d} U$, the (expected) limit relation for $X(\alpha)$ is

$$X(\alpha) \stackrel{\mathrm{d}}{=} \alpha U^{\alpha} X(\alpha) + (1-U)^{\alpha} X(\alpha)^*. \tag{6.41}$$

Interestingly we can do a similar calculation for the profile polynomial $W_n(u)$. For simplicity we assume that u is a positive real number. Then we already know that $M_n(u) = W_n(u)/\mathbb{E}\, W_n(u) \xrightarrow{d} M(u)$, where $M(u)$ is the limiting martingale. Again from (6.21) we obtain

$$W_n(u) \stackrel{\mathrm{d}}{=} u W_{I_n}(u) + W_{n-I_n}(u)^*,$$

and consequently

$$\begin{aligned} M_n(u) &\stackrel{\mathrm{d}}{=} u \frac{\mathbb{E}\, W_{I_n}(u)}{\mathbb{E}\, W_n(u)} M_{I_n}(u) + \frac{\mathbb{E}\, W_{n-I_n}(u)^*}{\mathbb{E}\, W_n(u)} M^*_{n-I_n}(u) \\ &\sim u \left(\frac{I_n}{n} \right)^u M_{I_n}(u) + \left(\frac{n-I_n}{n} \right)^u M_{n-I_n}(u)^*. \end{aligned}$$

By taking the limit $n \to \infty$ we get

$$M(u) \stackrel{\mathrm{d}}{=} u U^u M(u) + (1-U)^u M(u)^*. \tag{6.42}$$

Thus, $X(\alpha)$ and $M(u)$ satisfy the same distributional fixed point equation (6.41), resp. (6.42) when we set $\alpha = u$. Actually we have $X(\alpha) \stackrel{\mathrm{d}}{=} M(\alpha)$. By the martingale method we already know that $X_{n,\lfloor \alpha \log n \rfloor}/\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}$ has a limit $X(\alpha)$ which equals $M(\alpha)$.

Nevertheless, one can proceed directly without using profile polynomials. We will make this more precise in Section 8.2.3, however, the underlying idea is the following. In order to solve a distributional fixed point equation of the form (6.41) one has to introduce a proper metric on a space of measures so that the right-hand-side of the fixed point equation becomes a contraction. In particular, one gets a unique solution and since the normalised version (6.40) of the original underlying recurrence is close to the limit equation (6.41), one can expect that $X_{n,k}/\mathbb{E}\,X_{n,k}$ (for $k \sim \alpha \log n$) stabilises around the limit $X(\alpha)$.

This concept works quite well if the random variable of interest is not double indexed. For example, the so-called contraction method (introduced in Chapter 8) applies to the profile polynomial $W_n(u)$. In the case of the profile $X_{n,k}$ one has to take care of the second index k, too, which causes some extra work (compare with [87]). In fact, it is possible to prove Theorem 6.18 in the full range $0 < \alpha < e$. We will give more details in Section 8.2.3.

6.4 The Height of Recursive Trees

We encode the distribution of the height H_n of recursive trees by the generating function

$$y_k(x) = \sum_{n\ge 1} \mathbb{P}\{H_n < k\} \frac{x^n}{n}.$$

This is consistent with the generating function $y(x) = \log(1/(1-x))$ of recursive trees. More precisely, let $y_{n,k}$ denote the number of recursive trees of size n and height $< k$. Then

$$\mathbb{P}\{H_n < k\} = \frac{y_{n,k}}{(n-1)!},$$

and $y_k(x)$ also is the exponential generating function of these numbers:

$$y_k(x) = \sum_{n\ge 1} y_{n,k} \frac{x^n}{n!}.$$

Thus, we can use the usual counting procedure and obtain recursively

$$y_{k+1}'(x) = e^{y_k(x)}, \qquad (y_{k+1}(0) = 0),$$

with the initial condition $y_0(x) = 0$. In parallel we will work with

$$Y_k(x) = y_k'(x) = \sum_{n\ge 0} \mathbb{P}\{H_{n+1} < k\}\, x^n.$$

Here we have $Y_1(x) = 1$ and

$$Y_{k+1}'(x) = Y_{k+1}(x) Y_k(x), \qquad (Y_{k+1}(0) = 1).$$

The analysis of these generating functions is the basis of the following result (that is due to Drmota [57]).

Theorem 6.32. *The height H_n of random recursive trees has expected value*

$$\mathbb{E}\, H_n \sim e \log n. \tag{6.43}$$

Furthermore there are exponential tail estimates of the form

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge \eta\} = O\left(e^{-c\eta}\right)$$

with some $c > 0$.

This result is in accordance with Pittel's result [175] saying that $H_n/\log n \to e$ a.s.

Note that the exponential tail estimate shows that the distribution of the height is heavily concentrated around its expected value. In particular it follows that, as $n \to \infty$,

$$\mathbb{V}\mathrm{ar}\, H_n = O(1).$$

One can also be much more precise about the distribution of the height H_n. In particular the values $Y_k(1)$ play an essential role in the analysis. For example, we have

$$\mathbb{E}\, H_n = \max\{k : Y_k(1) \le n\} + O(1)$$

(compare with (6.48)), and

$$\mathbb{P}\{H_n < k\} = F(n/Y_k(1)) + o(1),$$

uniformly for all $k \ge 0$ as $n \to \infty$, where $F(y)$ satisfies the integral equation

$$y\, F(y/e^{1/e}) = \int_0^y F(z/e^{1/e}) F(y-z)\, dz. \tag{6.44}$$

We will comment on this problem at the end of the section. Anyway, this already shows that although the results on the height are quite precise, the behaviour of the distribution is in some sense implicit, since it uses $y_k'(1) = Y_k(1) = \sum_{n\ge 1} \mathbb{P}\{H_n < k\}$, a value that is unknown. We will just prove $\log Y_k(1) \sim k/e$ which provides (6.43). It is conjectured that

$$\log Y_k(1) = \frac{k}{e} + c_1 \log k + c_2(\log k) + o(1)$$

for some constant $c_1 > 0$ and a periodic function $c_2(x)$. Of course, if this conjectural relation can be verified then we actually get an explicit asymptotic representation for the height distribution.

The proof of Theorem 6.32 is divided into several steps. We start with some technical lemmas.

Lemma 6.33. *Suppose that $Y_1(x), Y_2(x), \overline{Y}_1(x), \overline{Y}_2(x)$ are non-negative continuous functions that are defined for $x \ge 0$ such that $Y_1(0) < \overline{Y}_1(0)$, $Y_2(0) < \overline{Y}_2(0)$, $Y_2'(x) = Y_2(x)Y_1(x)$, $\overline{Y}_2'(x) = \overline{Y}_2(x)\overline{Y}_1(x)$, and that the difference $\overline{Y}_1(x) - Y_1(x)$ has exactly one positive zero. Then the difference $\overline{Y}_2(x) - Y_2(x)$ has at most one positive zero.*

Proof. For $j = 1, 2$ set

$$y_j(x) = \int_0^x Y_j(t)\,dt \quad \text{and} \quad \overline{y}_j(x) = \int_0^x \overline{Y}_j(t)\,dt.$$

Then we have $y_1(x) < \overline{y}_1(x)$, $y_2(x) < \overline{y}_2(x)$ (at least) for a small interval $0 < x < \zeta$ and also $y_2'(x) = e^{y_1(x)}$ and $\overline{y}_2'(x) = e^{\overline{y}_1(x)}$. Since $\overline{Y}_1(x) - Y_1(x)$ is positive (for small positive x) and has at most one positive zero, the same follows for

$$\overline{y}_1(x) - y_1(x) = \int_0^x \left(\overline{Y}_1(t) - Y_1(t)\right)\,dt.$$

Namely, if $\overline{Y}_1(t) - Y_1(t) \ge 0$ for all $t \ge 0$ then $\overline{y}_1(x) - y_1(x)$ is an increasing function with $\overline{y}_1(0) - y_1(0) = 0$. If, however, $\overline{Y}_1(t) - Y_1(t) \ge 0$ for $0 \le t \le t_0$ and $\overline{Y}_1(t) - Y_1(t) \le 0$ for $t \ge t_0$ then $\overline{y}_1(x) - y_1(x)$ is increasing for $0 \le x \le t_0$ and decreasing for $x \ge t_0$. In the first case the function $\overline{y}_1(x) - y_1(x)$ has no positive zero and in the second case at most one.

Now observe that $(e^y - e^z)/(y - z) > 0$ for real $y \ne z$. Hence,

$$\overline{Y}_2(x) - Y_2(x) = \overline{y}_2'(x) - y_2'(x) = e^{\overline{y}_1(x)} - e^{y_1(x)} = \frac{e^{\overline{y}_1(x)} - e^{y_1(x)}}{\overline{y}_1(x) - y_1(x)} \left(\overline{y}_1(x) - y_1(x)\right)$$

has at most one positive zero, too.

Lemma 6.34. *For all $k \ge 0$ we have*

$$\frac{Y_{k+2}(1)}{Y_{k+1}(1)} \le \frac{Y_{k+1}(1)}{Y_k(1)}.$$

Proof. For $0 \le \gamma < 1$ set

$$V_k(x,\gamma) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \le x \le 1-\gamma, \\ \gamma^{-1} Y_k\left(\frac{x-(1-\gamma)}{\gamma}\right) & \text{for } 1-\gamma \le x \le 1. \end{cases}$$

These functions satisfy

$$V_{k+1}'(x,\gamma) = V_{k+1}(x,\gamma) V_k(x,\gamma),$$

$V_k(0) = 1$ and $V_k(1,\gamma) = \gamma^{-1} Y_k(1)$. In particular, for $\gamma_k = Y_{k+1}(1)/Y_k(1)$ we have $V_k(1,\gamma_k) = Y_{k+1}(1)$. Now the inductive application of Lemma 6.33 shows that $Y_{k+1}(x) - V_k(x,1)$ have (at most) one positive zero. Hence we get

$$Y_{k+1}(x) \le V_k(x,\gamma_k) \quad \text{for} \quad 0 \le x \le 1,$$

and after integration

$$Y_{k+2}(1) \le V_{k+1}(1,\gamma_k) = \gamma_k^{-1} Y_{k+1}(1) = \frac{Y_{k+1}(1)^2}{Y_{k+1}(1)}.$$

Corollary 6.35 *There exists* $\alpha_0 > 1$ *with*

$$\alpha_0 = \lim_{k\to\infty} \frac{Y_{k+1}(1)}{Y_k(1)}. \tag{6.45}$$

Proof. Lemma 6.34 implies that there exists $\alpha_0 \ge 1$ with

$$\alpha_0 = \lim_{k\to\infty} \frac{Y_{k+1}(1)}{Y_k(1)}.$$

We show that $Y_k(1)$ grows at least exponentially which implies $\alpha_0 > 1$.

Set $\delta_k = y(x) - y_k(x)$. Then it follows by induction that

$$\delta_k(x) \le \frac{L(x)^{k+1}}{(k+1)!} \qquad (0 \le x \le 1), \tag{6.46}$$

where $L(x)$ abbreviates $\log(1/(1-x))$. By definition (6.46) is true for $k = 0$. Now observe that $y_k(x) \le y(x)$ implies

$$\delta'_{k+1}(x) = e^{y(x)} - e^{y_k(x)} \le e^{y(x)}\delta_k(x) \le \frac{1}{1-x}\frac{L(x)^{k+1}}{(k+1)!}.$$

Since $\delta_{k+1}(0) = 0$, integration gives

$$\delta_{k+1}(x) \le \int_0^x \frac{1}{1-t}\frac{L(t)^{k+1}}{(k+1)!}\,dt = \frac{L(x)^{k+2}}{(k+2)!}.$$

For $x_k = 1 - e^{-(k+1)/e}$ we have $y(x_k) = (k+1)/e$ and

$$\frac{L(x_k)^{k+1}}{(k+1)!} \sim \frac{1}{\sqrt{2\pi k}}.$$

Hence it follows that (for sufficiently large k)

$$y_k(1) \ge y_k(x_k) \ge y(x_k) - \frac{L(x_k)^{k+1}}{(k+1)!} = \frac{k+1}{e} - \frac{1}{\sqrt{2\pi k}}(1+o(1)) \ge \frac{k}{e}.$$

This completes the proof of the lemma since

$$Y_k(1) = y'_k(1) = e^{y_{k-1}(1)} \ge e^{(k-1)/e}.$$

Remark 6.36 *If we replace* x_k *in the proof of Corollary (6.35) by*

$$x_k = 1 - e^{-(k+\log\sqrt{k})/e}$$

we get the slightly better lower bound

$$Y_k(1) \ge e^{(k+\log\sqrt{k})/e+O(1)}.$$

However, it is conjectured that the correct order of magnitude is

$$Y_k(1) = e^{(k+\frac{3}{2}\log k)/e+O(1)},$$

which cannot be reached by the method mentioned above.

The property that $Y_k(1)$ has an approximately exponentially growing behaviour is sufficient to prove exponential tail estimates for the distribution of H_n.

Lemma 6.37. *We have the estimates*

$$\mathbb{P}\{H_n < k\} \le \frac{Y_k(1)}{n}$$

and

$$\mathbb{P}\{H_n \ge k\} \le e\,\frac{n}{Y_k(1)}.$$

Proof. By considering the evolution model of recursive trees it immediately follows that the height is an increasing sequence in n. Hence we have for all $n, k \ge 0$

$$\mathbb{P}\{H_{n+1} < k\} \le \mathbb{P}\{H_n < k\}.$$

It follows directly that

$$\begin{aligned} Y_k(1) &= \sum_{\ell \ge 0} \mathbb{P}\{H_{\ell+1} < k\} \\ &\ge \sum_{0 \le \ell < n} \mathbb{P}\{H_{\ell+1} < k\} \\ &\ge n\,\mathbb{P}\{H_n < k\} \end{aligned}$$

or $\mathbb{P}\{H_n < k\} \le Y_k(1)/n$.

Next consider the function

$$\overline{Y}(x) = \frac{1}{1 + \frac{1}{Y_k(1)} - x},$$

which satisfies the differential equation $\overline{Y}'(x) = \overline{Y}(x)^2$. Since $0 < \overline{Y}(0) < 1$ there is exactly one positive zero of the difference $Y_0(x) - \overline{Y}(x)$. Hence, an inductive application of Lemma 6.33 implies that this is also true for the difference $Y_k(x) - \overline{Y}(x)$. However, by construction $\overline{Y}(1) = Y_k(1)$. Since $\overline{Y}(0) < Y_k(0)$, it follows that $Y_k(x) \le \overline{Y}(x)$ for $0 \le x \le 1$. In particular, if we set $x_n = 1 - \frac{1}{n}$ we obtain

$$\begin{aligned} Y(x_n) - \overline{Y}(x_n) &\ge Y(x_n) - Y_k(x_n) \\ &= \sum_{\ell \ge 0} \mathbb{P}\{H_{\ell+1} \ge k\} x_n^\ell \\ &\ge \sum_{\ell \ge n-1} \mathbb{P}\{H_{\ell+1} \ge k\} x_n^\ell \\ &\ge \mathbb{P}\{H_n \ge k\} \frac{x_n^{n-1}}{1 - x_n} \\ &\ge \mathbb{P}\{H_n \ge k\} \frac{n}{e}. \end{aligned}$$

Since

$$Y(x_n) - \overline{Y}(x_n) = n - \frac{1}{\frac{1}{n} + \frac{1}{Y_k(1)}} \le \frac{n^2}{Y_k(1)},$$

we finally get

$$\mathbb{P}\{H_n \ge k\} \le e\,\frac{n}{Y_k(1)}.$$

Set

$$h_n = \max\{k : Y_k(1) \le n\}.$$

Then Lemma 6.37 combined with the property that $Y_{k+1}(1) \ge (\alpha_0 - \epsilon)Y_k(1)$ (for sufficiently large k) provides exponential tail estimates of the kind

$$\mathbb{P}\{|H_n - h_n| \ge \eta\} = O\left(e^{-\eta c}\right), \tag{6.47}$$

for a properly chosen constant c. Of course, (6.47) implies

$$\mathbb{E}\, H_n = h_n + O(1) \tag{6.48}$$

and consequently

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge \eta\} = O\left(e^{-\eta c}\right),$$

which also implies that $\mathbb{V}\mathrm{ar}\, H_n = O(1)$ as $n \to \infty$. Thus, in order to complete the proof of Theorem 6.32 we need more precise information on $Y_k(1)$ (resp. on h_n). Namely, the limit relation of Corollary 6.35 implies

$$h_n \sim (\log n)/(\log \alpha_0)$$

and therefore

$$\mathbb{E}\, H_n \sim \frac{\log n}{\log \alpha_0}.$$

It remains to prove $\alpha_0 = e^{1/e}$. For this purpose we will define auxiliary functions $\overline{y}_k(\alpha, x)$ and $\overline{Y}_k(\alpha, x)$. But before we can do that we have to solve an integral equation.

Lemma 6.38. *Suppose that $1 < \alpha < e^{1/e}$ and that $\beta < e$ denotes the smallest positive solution of*

$$\alpha^\beta = \beta.$$

Furthermore, let $\mathcal{F}$ denote the set of monotonically decreasing and continuous functions $F(y)$ $(y \ge 0)$ that satisfy

$$F(y) = 1 - y^\beta + O(y^e)) \qquad (y \to 0+)$$

and $F(y) \to 0$ as $y \to \infty$. Then there exists a unique solution $F_\alpha \in \mathcal{F}$ of the integral equation

$$yF(y/\alpha) = \int_0^y F(z/\alpha)F(y-z)\,dz. \tag{6.49}$$

Moreover, there exists $C > 0$ *such that*

$$F_\alpha(y) = O\left(e^{-Cy}\right) \tag{6.50}$$

as $y \to \infty$.

Proof. It is easy to show that d defined by

$$d(F_1, F_2) = \sup_{y \ge 0} \left(|F_1(y) - F_2(y)| y^{-e}\right)$$

is a complete metric on $\mathcal{F}$ and that (6.49) is a fixed point equation on $\mathcal{F}$ that can be rewritten as $F = A(F)$, where

$$A(F) = \frac{1}{\alpha y} \int_0^{\alpha y} F(z/\alpha) F(\alpha y - z)\, dz.$$

Moreover, A is a contraction with Lipschitz constant

$$K = \frac{1 + \alpha^e}{1 + e} < 1.$$

Hence, by Banach's fixed point theorem there is a unique solution.

Finally, (6.50) can be proved in an inductive way. Let $F_0(y) = \max\{1 - y^e, 0\}$ and $F_{n+1} = A(F_n)$. By keeping track of the contraction A it follows that there exists $C_0 > 0$ and $y_0'' > 0$ such that

$$\sup_{n \ge 0} F_n(y) < 1 - y^\beta + C_0 y^e < 1$$

for $0 \le y \le y_0''$. Hence, for every $0 < y_0' < y_0''$ there exists $C > 0$ such that $F_n(y) \le e^{-Cy}$ for all $n \ge 0$ and for $y_0'\alpha \le y \le y_0''$ and that $F_0(y) \le e^{-Cy}$ for $y \ge y_0'$. More precisely we can choose $C = C(y_0')$ with $C \sim (y_0')^{\beta - 1}$ as $y_0' \to 0$. In particular we can assure that

$$\frac{1 - e^{-C(\alpha-1)y}}{C} + y_0' e^{C y_0'/\alpha} \le (\alpha - 1)y + y_0' \tag{6.51}$$

for $y \ge y_0''$.

It remains to verify the upper bound $F_n(y) \le e^{-Cy}$ for $y \ge y_0''$ which we will do by induction. Recall that

$$F_{n+1}(y) = \frac{1}{\alpha y} \int_0^{\alpha y} F_n(z/\alpha) F_n(\alpha y - z)\, dz.$$

By using the a priori bound we can assume that $F_n(y) \le e^{-Cy}$ even for $y \ge y_0'$. We split the integral into tree parts: $0 \le z \le (\alpha - 1)y$, $(\alpha - 1)y \le z \le \alpha y - y_0'$ and $\alpha y - y_0' \le z \le \alpha y$.

For the first part of the integral we have $\alpha y - z \ge y \ge y_0''$ and consequently

$$\int_0^{(\alpha-1)y} F_n(z/\alpha)F_n(\alpha y - z)\,dz \le \int_0^{(\alpha-1)y} e^{-C(\alpha y - z)}\,dz$$
$$= e^{-C\alpha y}\frac{e^{C(\alpha-1)y}-1}{C}.$$

In the second range we have $z/\alpha \ge y_0'$ and $\alpha y - z \ge y_0'$. Hence we get

$$\int_{(\alpha-1)y}^{\alpha y - z} F_n(z/\alpha)F_n(\alpha y - z)\,dz \le \int_{(\alpha-1)y}^{\alpha y - z} e^{-C(z/\alpha+\alpha y - z)}\,dz$$
$$\le (\alpha y - z - (\alpha-1)y)\; e^{-Cy}.$$

Finally, for the third integral we have

$$\int_{\alpha y - z}^{\alpha y} F_n(z/\alpha)F_n(\alpha y - z)\,dz \le y_0' e^{-Cy + Cy_0'/\alpha}.$$

For proving $F_{n+1}(y) \le e^{-Cy}$, it is thus sufficient to check that the sum of these three upper bounds is smaller or equal $\alpha y e^{-Cy}$ for $y \ge y_0''$:

$$e^{-C\alpha y}\frac{e^{C(\alpha-1)y}-1}{C} + (\alpha y - z - (\alpha-1)y)\; e^{-Cy} + y_0' e^{-Cy+Cy_0'/\alpha} \le \alpha y e^{-Cy}.$$

However, this inequality is equivalent to (6.51). Hence, we have shown $F_{n+1}(y) \le e^{-Cy}$ (for $y \ge y_0''$).

By taking the limit $n \to \infty$, we observe the same upper bound for $F_\alpha(y)$.

Note that for every solution $F(y)$ of (6.49) the function $F(cy)$ (where $c > 0$) is also a solution of (6.49). Thus, we can assume without loss of generality that (after a proper scaling) $F_\alpha(y)$ satisfies (6.49) and

$$\int_0^\infty F_\alpha(y)\,dy = 1.$$

Note that $\alpha = e^{1/e}$ is a critical value for the fixed point equation (6.49) because the Lipschitz constant K would equal 1 and, thus, we could not apply Banach's fixed point theorem.[5]

Next, consider the Laplace transforms

$$\Phi_\alpha(u) = \int_0^\infty F_\alpha(y)e^{-yu}\,dy.$$

They satisfy $\Phi_\alpha(0) = 1$ and

$$\Phi_\alpha'(u) = -\frac{1}{\alpha}\Phi_\alpha(u)\Phi_\alpha(u/\alpha).$$

[5] It is, however, possible to solve (6.49) even for $\alpha = e^{1/e}$, see [57].

The crucial step in this analysis it to introduce (for $1 < \alpha < e^{1/e}$) the following auxiliary functions

$$\overline{Y}_k(\alpha, x) = \alpha^k \Phi_\alpha \left(\alpha^k (1-x)\right), \tag{6.52}$$

where k can be considered a real (not necessarily integral) parameter and

$$\overline{y}_k(\alpha, x) = \int_0^x \overline{Y}_k(\alpha, t)\, dt = \log \overline{Y}_{k+1}(\alpha, x).$$

The next lemma collects some useful facts on $\overline{Y}_k(\alpha, x)$ if $\alpha < e^{1/e}$. In fact, these auxiliary functions have almost the same properties as the function $Y_k(x)$. The proof is immediate by translating the corresponding properties of F_α and Φ_α respectively.

Lemma 6.39. *Suppose that $1 < \alpha < e^{1/e}$ and let $\beta < e$ be given by $\alpha^\beta = \beta$. Let $\overline{Y}_k(\alpha, x)$ be defined by (6.52). Then the following assertions hold:*

1. *For all $k > 0$ the function $\overline{Y}_k(\alpha, x)$ is monotone for $x \ge 0$. It satisfies $0 < \overline{Y}_k(\alpha, 0) < 1$, more precisely, $1 - \overline{Y}_k(\alpha, 0) \sim C\alpha^{-\beta k}$ for some constant C depending on α. Furthermore we have $\overline{Y}_k(\alpha, 1) = \alpha^k$.*
2. *They satisfy the recurrence relation*

$$\overline{Y}'_{k+1}(x) = \overline{Y}_{k+1}(x)\overline{Y}_k(x).$$

3. *For all integers $\ell \ge 0$ and for all real numbers $k > 0$ the difference $Y_\ell(x) - \overline{Y}_k(\alpha, x)$ has exactly one positive zero $x_{\ell,k}$. In particular we have $\overline{Y}_k(x) \le Y_\ell(x)$ for $0 \le x \le x_{\ell,k}$ and $\overline{Y}_k(x) \ge Y_\ell(x)$ for $x \ge x_{\ell,k}$.*

Note that $y_k(x) = \log Y_{k+1}(x)$ and $\overline{y}_k(\alpha, x) = \log \overline{Y}_{k+1}(\alpha, x)$. Hence, the above properties can be translated to $\overline{y}_k(\alpha, x)$ and $y_k(x)$.

The next lemma completes the proof of Theorem 6.32.

Lemma 6.40. *We have* $\lim_{k\to\infty} Y_{k+1}(1)/Y_k(1) = e^{1/e}$ *and consequently*

$$\mathbb{E}\, H_n \sim e \log n. \tag{6.53}$$

Proof. Suppose that $1 < \alpha < e^{1/e}$ and set $e_k := y_k(1)/(\log \alpha) - 1$. Then the function $\overline{y}_{e_k}(\alpha, x)$ satisfies $\overline{y}_{e_k}(\alpha, 0) < y_k(0)$ and

$$\overline{y}_{e_k}(\alpha, 1) = y_k(1).$$

Hence, by Lemma 6.39 (reformulated for $\overline{y}_k(\alpha, x)$) it follows that $\overline{y}_{e_k}(\alpha, x) \le y_k(x)$ for $0 \le x \le 1$. Hence, by integration it also follows that $\overline{y}_{e_k+1}(\alpha, x) \le y_{k+1}(x)$ for $0 \le x \le 1$. In particular,

$$\overline{y}_{e_k+1}(\alpha, 1) = y_k(1) + \log \alpha \le y_{k+1}(1).$$

Thus, we have $y_{k+1}(1) - y_k(1) \ge \log \alpha$ for all $\alpha < e^{1/e}$ and consequently $y_{k+1}(1) - y_k(1) \ge 1/e$. This also shows that $Y_k(1)/Y_{k-1}(1) \ge e^{1/e}$.

In a second step we will show that for every $\epsilon > 0$

$$y_k(1) \le \frac{k}{e}(1+\epsilon)$$

for sufficiently $k \ge k_0(\epsilon)$. This is sufficient to complete the proof of Lemma 6.40.

We again fix $\alpha < e^{1/e}$ and define $t(\alpha) > 0$ by

$$(1+t(\alpha))\alpha^\beta \log \alpha = 1.$$

Note that $\lim_{\alpha \to e^{1/e}} t(\alpha) = 0$.

Set $\delta_k(x) = y_k(x) - \overline{y}_{k+r}(\alpha, x)$, where $r \ge 0$ is a parameter that will be chosen appropriately. Note that $y_k(x) \le y(x)$ and $\overline{y}_{k+r}(\alpha, x) \le y(x)$ for $0 \le x < 1$. By induction it follows that

$$\delta_k(x) \le \sum_{\ell=0}^{k} \delta_\ell(0) \frac{L(x)^{k-\ell}}{(k-\ell)!},$$

where $L = \log 1/(1-x)$. We now suppose that $r = 2kt(\alpha) - 1$, set $x' = 1 - \alpha^{-k(1+t(\alpha))}$ and estimate $y_k(x') = \overline{y}_{k+2kt(\alpha)-1}(\alpha, x') + \delta_k(x')$ from above. We have

$$\overline{y}_{k+2kt(\alpha)-1}(\alpha, x') = k(1+t(\alpha)) - \frac{C}{\alpha^{\beta k t(\alpha)}}(1+o(1)),$$

and

$$\begin{aligned}
\delta_k(x') &\le \sum_{\ell=0}^{k} \frac{C}{\alpha^{\beta\ell + 2\beta k t(\alpha)}} \frac{(k(1+t(\alpha))\log\alpha)^{k-\ell}}{(k-\ell)!} \\
&= \frac{C}{\alpha^{\beta k(1+2t(\alpha))}} \sum_{\ell=0}^{k} \frac{\left(k(1+t(\alpha))\alpha^\beta \log\alpha\right)^{k-\ell}}{(k-\ell)!} \\
&= \frac{C}{\alpha^{\beta k(1+2t(\alpha))}} \sum_{\ell=0}^{k} \frac{k^{k-\ell}}{(k-\ell)!} \\
&\sim \frac{C}{\alpha^{\beta k(1+2t(\alpha))}} \frac{e^k}{2} \\
&= \frac{1}{2} \frac{C}{\alpha^{\beta k t(\alpha)}}.
\end{aligned}$$

Consequently we obtain

$$y_k(x') \le k(1+t(\alpha))\log\alpha - \frac{1}{2}\frac{C}{\alpha^{\beta k t(\alpha)}}(1+o(1)).$$

If we compare that with

$$\overline{y}_{k+3kt(\alpha)-1}(\alpha, x') = k(1+t(\alpha))\log\alpha - \frac{C}{\alpha^{2\beta k t(\alpha)}}(1+o(1))$$

we observe that (for sufficiently large k)

$$y_k(x') \le \overline{y}_{k+3kt(\alpha)-1}(\alpha, x').$$

Since $y_k(0) > \overline{y}_{k+3kt(\alpha)}(\alpha, 0)$, it follows from Lemma 6.39 (resp. from its reformulation to $\overline{y}_k(\alpha, x)$), that $y_k(x) \le \overline{y}_{k+3kt(\alpha)}(\alpha, x)$ even for all $x \ge x'$. In particular we have (for sufficiently large k)

$$y_k(1) \le \overline{y}_{k+3kt(\alpha)-1}(\alpha, 1) = k(1+3t(\alpha))\log\alpha \le \frac{k}{e}(1+3t(\alpha)).$$

Since we can choose α such that $t(\alpha)$ is arbitrarily small this completes the proof of Lemma 6.40.

We close this section with some comments on the distribution of H_n. In Lemma 6.38 we have observed that for every $\alpha < e^{1/e}$ the integral equation (6.49) has a solution F_α, since it is a fixed point of a proper contraction. For $\alpha_0 = e^{1/e}$ the method fails. Nevertheless it is possible to find a non-trivial continuous solution for α_0 (see [57, 36]). As above we consider the Laplace transform Φ_{α_0} and the auxiliary functions

$$\begin{aligned}
\overline{Y}_k(x) &= \alpha_0^k \Phi_{\alpha_0}(\alpha_0^k(1-x)) \\
&= \alpha_0^k \int_0^\infty F_{\alpha_0}(y) e^{-y\alpha_0^k(1-x)}\, dy \\
&= \int_0^\infty F_{\alpha_0}(y\alpha_0^{-k}) e^{-y} e^{yx}\, dy \\
&= \sum_{n\ge 0} \left(\frac{1}{n!} \int_0^\infty y^n e^{-y} F_{\alpha_0}(y\alpha_0^{-k})\, dy \right) x^n.
\end{aligned}$$

Observe that

$$\frac{1}{n!}\int_0^\infty y^n e^{-y}\, dy = 1$$

and that the integrand $y^n e^{-y} = e^{n\log y - y}$ is highly concentrated around $y = n$. Hence, by a simple application of the Laplace method and by continuity of F_{α_0} we obtain

$$[x^n]\overline{Y}_k(x) = F_{\alpha_0}(n\alpha_0^{-k}) + o(1)$$

uniformly for n and k with $C_1 \le n\alpha^{-k} \le C_2$, where $0 < C_1 < C_2$ are arbitrary constants.

Recall that $Y_{k+1}(1)/Y_k(1) \sim \alpha_0 = e^{1/e}$. By definition we also have the relation $\overline{Y}_{k+1}(1)/\overline{Y}_k(1) = \alpha_0 = e^{1/c}$. Since $Y_k(x)$ and $\overline{Y}_k(x)$ satisfy the same recurrence $Y_{k+1}'(x) = Y_{k+1}(x)Y_k(x)$ resp. $\overline{Y}_{k+1}'(x) = \overline{Y}_{k+1}(x)\overline{Y}_k(x)$, one is led to the conjecture that $Y_k(x)$ and $\overline{Y}_k(x)$ are quite close. This is definitely true for small x but not for x close to 1. However, we can do the following trick. Define d_k by the relation

$$\alpha_0^{d_k} = Y_k(1),$$

that is, we have $\overline{Y}_{d_k}(1) = Y_k(1)$. Then

$$Y_k'(1) = Y_k(1)Y_{k-1}(1) \sim Y_k(1)^2\alpha_0^{-1} = \alpha_0^{2d_k-1} = \overline{Y}_k(1)\overline{Y}_{k-1}(1), = \overline{Y}'_{d_k}(1)$$

and similarly $Y_k^{(j)}(1) \sim \overline{Y}_{d_k}^{(j)}(1)$ for all $j \geq 1$. Thus, $Y_k(x)$ can be properly approximated by $\overline{Y}_{d_k}(x)$ in a complex neighbourhood of $x = 1$. Together with some further (technical but easy) estimates (compare with [62]) it follows via Cauchy's formula that

$$\begin{aligned}
\mathbb{P}\{H_{n+1} < k\} &= \frac{1}{2\pi i}\int_{|x|=1} \frac{Y_k(x)}{x^{n+1}}\,dx \\
&= \frac{1}{2\pi i}\int_{|x|=1} \frac{\tilde{Y}_{d_k}(x)}{x^{n+1}}\,dx + o(1) \\
&= F_{\alpha_0}(n\alpha_0^{-d_k}) + o(1) \\
&= F_{\alpha_0}(n/Y_k(1)) + o(1),
\end{aligned}$$

which leads to the proposed asymptotic representation of the height distribution.

6.5 Profile and Height of Binary Search Trees and Related Trees

Recursive trees have several generalisations including binary search trees, m-ary increasing trees and (generalised) plane oriented recursive trees. A second line of generalisations are m-ary search trees as introduced in Section 1.4.2.

All these classes of trees have similar asymptotic properties. In contrast to Galton-Watson trees their height is of order $\log n$. The corresponding limiting profile processes are similar to each other but not identical.

6.5.1 The Profile of Binary Search Trees and Related Trees

In the case of binary search trees we distinguish between internal and external nodes. Accordingly we consider the internal profile $I_{n,k}$ and the external profile $B_{n,k}$, that is, the number of internal resp. external nodes at level k in trees with a total number of n internal nodes. Both profiles are closely related to each other. Due to the structure of a binary tree

$$B_{n,k} = 2I_{n,k-1} - I_{n,k} \quad \text{and} \quad I_{n,k} = \sum_{j>k} 2^{k-j} B_{n,j}. \tag{6.54}$$

Actually, the external profile is easier to study. It has almost the same properties as the profile or recursive trees. First, it is related to the insertion depth D_n (of the internal nodes) by

$$\mathbb{P}\{D_n = k\} = \frac{\mathbb{E}\, B_{n-1,k}}{n-1}.$$

Second, we have

$$B_{n,k} \stackrel{\mathrm{d}}{=} B_{I_n,k-1} + B^*_{n-1-I_n,k-1},$$

where I_n is uniformly distributed over $\{0, 1, \ldots, n-1\}$, $B_{n,k} \stackrel{\mathrm{d}}{=} B^*_{n,k}$, and I_n, $(B_{n,k})$ and $(B^*_{n,k})$ are independent. We also have

$$\begin{aligned}\mathbb{E}(B_{n,k}|T_{n-1}) &= (B_{n-1,k}+2)\frac{B_{n-1,k-1}}{n} + (B_{n-1,k}-1)\frac{B_{n-1,k}}{n}\\ &\quad + B_{n-1,k}\left(1 - \frac{B_{n-1,k-1}+B_{n-1,k}}{n}\right)\\ &= \frac{2B_{n-1,k-1}}{n} + \frac{(n-1)B_{n-1,k}}{n}.\end{aligned}$$

For example, by using the last property it follows that

$$\mathbb{E}\, W_n(z) = \mathbb{E}\sum_{k\ge 0} B_{n,k}x^k = (-1)^n\binom{-2x}{n},$$

and consequently

$$\mathbb{E}\ B_{n,k} = \frac{2^k}{n!}|s_{n,k}|. \tag{6.55}$$

(It seems that this explicit formula was first observed by Lynch [145], compare also with [146]).

There are several other ways to derive (6.55). If we introduce the generating functions

$$Y_k(x,u) = \sum_{n\ge 0}\mathbb{E}\ u^{B_{n,k}}, x^n.$$

Then we have $Y_0(x,u) = u + x/(1-x)$ and recursively

$$\frac{\partial Y_{k+1}(x,u)}{\partial x} = Y_k(x,u)^2 \tag{6.56}$$

with $Y_{k+1}(0,u) = 1$ (for $k \ge 0$). By taking derivatives with respect to u we obtain

$$Z_k(x) = \left[\frac{\partial Y_k(x,u)}{\partial u}\right]_{u=1} = \sum_{n\ge 0}\mathbb{E}\ B_{n,k}\, x^n$$

which satisfies $Z_0(x) = 1$ and by (6.56)

$$Z'_{k+1}(x) = 2Y_k(x,1)Z_k(x) = \frac{2}{1-x}Z_k(x),$$

with $Z_{k+1}(0) = 0$ (for $k \ge 0$). Hence,

$$Z_k(x) = \frac{2^k}{k!}\left(\frac{1}{1-x}\right)^k,$$

which translates to (6.55).

By using the asymptotic expansion for Stirling numbers (from Lemma 6.1) the representation (6.55) implies

$$\mathbb{E}\, B_{n,k} = \frac{2^k (\log n)^k}{k!\, n\, \Gamma\left(\frac{k}{\log n}\right)} \sim \frac{n^{\alpha(1-\log(\alpha/2))-1}}{\sqrt{2\pi k}} \left(1 + O\left(\frac{1}{n}\right)\right), \tag{6.57}$$

where $\alpha = \frac{k}{\log n}$. In particular, if we just consider a local expansion for k close to $2\log n$ we obtain

$$\mathbb{E}\, B_{n,k} = \frac{n}{\sqrt{4\pi \log n}} \left(e^{-\frac{(k-2\log n)^2}{4\log n}} + O\left(\frac{1}{\sqrt{\log n}}\right)\right). \tag{6.58}$$

By (6.54) we get the same for $I_{n,k}$:

$$\mathbb{E}\, I_{n,k} = \frac{n}{\sqrt{4\pi \log n}} \left(e^{-\frac{(k-2\log n)^2}{4\log n}} + O\left(\frac{1}{\sqrt{\log n}}\right)\right).$$

This indicates that the majority of the nodes of a binary search tree T_n is concentrated around level $2\log n$. In particular, a central limit theorem for the insertion depth D_n follows.

From (6.58) it also follows that $\mathbb{E}\, B_{n,k} \to \infty$ (for $k \sim \alpha \log n$), if and only if $\alpha \in (\alpha_-, \alpha_+)$, where $0 < \alpha_- < 2 < \alpha_+$ are the solutions of the equation

$$\alpha \log\left(\frac{2e}{\alpha}\right) = 1, \quad \alpha_- = 0.373\ldots, \quad \alpha_+ = 4.311\ldots$$

This implies that the natural range of the external profile is $\alpha_- \log n \le k \le \alpha_+ \log n$. The corresponding limit theorem is the following one (see [37, 87]).

Theorem 6.41. *For $\alpha_- < \alpha < \alpha_+$ there exists a random variable $X(\alpha)$ that satisfies the equation*

$$X(\alpha) \stackrel{\mathrm{d}}{=} \frac{\alpha}{2} U^{\alpha-1} X(\alpha) + \frac{\alpha}{2}(1-U)^{\alpha-1} X(\alpha)^*$$

with $\mathbb{E}\, X(\alpha) = 1$, where $X(\alpha)$, $X(\alpha)^*$, U are independent, $X(\alpha) \stackrel{\mathrm{d}}{=} X(\alpha)^*$, and U is uniformly distributed on $[0,1]$, such that

$$\frac{B_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \xrightarrow{d} X(\alpha).$$

The internal profile behaves different. Here we have (see [87])

$$\mathbb{E}\, I_{n,k} \sim \begin{cases} 2^k - \frac{(2\log n)^k}{nk!\left(1-\frac{k}{\log n}\right)\Gamma\left(\frac{k}{\log n}\right)} & \text{for } k \le \log n - O(\sqrt{\log n}), \\ 2^k \Phi\left(-\frac{k-\log n}{\sqrt{\log n}}\right) & \text{for } k = \log n + o((\log n)^{2/3}), \\ \frac{(2\log n)^k}{nk!\left(1-\frac{k}{\log n}\right)\Gamma\left(\frac{k}{\log n}\right)} & \text{for } \log n + O(\sqrt{\log n}) \le k \le O(\log n), \end{cases}$$

where $\Phi(x)$ denotes the normal distribution function. This means that the binary search tree is almost full up to level $k = \log n$, that is, the expected number of internal nodes is approximately 2^k. For higher level the situation is different. There the number of internals and externals are (asymptotically) proportional. According to these two regimes we have

$$\frac{2^k - I_{n,\lfloor \alpha \log n \rfloor}}{2^k - \mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \xrightarrow{d} X(\alpha)$$

for $\alpha_- < \alpha < 1$ and

$$\frac{I_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \xrightarrow{d} X(\alpha)$$

for $1 < \alpha < \alpha_+$, where $X(\alpha)$ is the same limiting random variable as in Theorem 6.41.

There are again at least three different proofs for these properties: the martingale method (see [37, 38, 39]), the moment method and the contraction method (see [87]) which work exactly as in the case of recursive trees.

Binary search trees can be considered as special cases of d-ary recursive trees (see Section 1.3.3), that are determined by the generating series $\Phi(x) = (1+x)^d$. This implies that the differential equation of the generating function $y(x) = \sum_{n\ge 1} y_n x^n/n!$ of d-ary recursive trees is given by $y'(x) = (1+y(x))^d$. We have (see Section 6.2.2)

$$y(x) = (1-(d-1)x)^{-\frac{1}{d-1}} - 1$$

and

$$y_n = n!(-(d-1))^n \binom{-\frac{1}{d-1}}{n} \sim n!(d-1)^n \frac{n^{\frac{2-d}{d-1}}}{\Gamma\left(\frac{1}{d-1}\right)}.$$

We slightly change the generating function to $Y(x) = 1 + y(x)$. Here we have the differential equation $Y'(x) = Y(x)^d$ with initial condition $Y(0) = 1$.

In order to study the external profile we introduce the generating functions

$$Y_k(x,u) = \sum_{n\ge 0} \mathbb{E}\, u^{B_{n,k}}\, y_n \frac{x^n}{n!}. \tag{6.59}$$

As in the binary case we have $Y_0(x,u) = u + y(x)$ and recursively

$$\frac{\partial Y_{k+1}(x,u)}{\partial x} = Y_k(x,u)^d$$

with $Y_k(0,u)=1$. The corresponding generating function

$$Z_k(x) = \left[\frac{\partial Y_k(x,u)}{\partial u}\right]_{u=1} = \sum_{n\ge 0} \mathbb{E}\, B_{n,k}\, y_n \frac{x^n}{n!}$$

of the expected profile satisfies

$$Z'_{k+1}(x) = \frac{d}{1-(1-d)x} Z_k(x)$$

(with $Z_{k+1}(0)=0$ for $k\ge 0$). The solution

$$Z_k(x) = \frac{d^k}{k!}\left(\log\left(\frac{1}{1-(1-d)x}\right)\right)^k$$

rewrites to

$$\begin{aligned}
\mathbb{E}\, B_{n,k} &= \frac{d^k |s_{n,k}|}{k!(-1)^n \binom{-\frac{1}{d-1}}{n}} \\
&\sim \frac{(d\log n)^k \Gamma\left(\frac{1}{d-1}\right)}{n^{\frac{1}{d-1}} k!\, \Gamma\left(\frac{k}{\log n}\right)} \\
&\sim \frac{n^{\alpha(1-\log(\alpha/d))-\frac{1}{d-1}}\, \Gamma\left(\frac{1}{d-1}\right)}{\sqrt{2\pi k}\,\Gamma(\alpha)},
\end{aligned}$$

where $\alpha = k/\log n = O(1)$. The profile is thus concentrated around the level $k = d\log n$ and by $\mathbb{P}[D_n = k] = \mathbb{E}\, B_{n-1,k}/((d-1)(n-1)+1)$ we again get a central limit theorem for the insertion depth D_n with $\mathbb{E}\, D_n = d\log n + O(1)$ and $\mathbb{V}\mathrm{ar}\, D_n = d\log n + O(1)$.

Furthermore it follows that $\mathbb{E}\, B_{n,k}\to\infty$ (for $k\sim\alpha\log n$), if and only if $\alpha\in(\alpha_{d,-},\alpha_{d,+})$, where $0<\alpha_{d,-}<d<\alpha_{d,+}$ are the solutions of the equation

$$\alpha\log\left(\frac{de}{(d-1)\alpha}\right) = \frac{1}{d-1}.$$

Hence, the natural range for the external profile is $\alpha_{d,-}\log n \le k\le \alpha_{d,+}\log n$ for which we have the following property (which is a direct extension of Theorem 6.41).

Theorem 6.42. *Let $B_{n,k}$ denote the external profile of d-ary recursive trees.*

Then for $\alpha_{d,-} < \alpha < \alpha_{d,+}$ there exists a random variable $X(\alpha)$ that satisfies the equation

$$X(\alpha) \overset{\mathrm{d}}{=} \frac{\alpha}{d} V_1^{\alpha-\frac{1}{d-1}} X(\alpha)^{(1)} + \cdots \frac{\alpha}{d} V_d^{\alpha-\frac{1}{d-1}} X(\alpha)^{(d)}$$

with $\mathbb{E}\, X(\alpha) = 1$, *where* $X(\alpha)^{(1)}, \ldots, X(\alpha)^{(d)}$, $(V_1, \ldots, V_d)$ *are independent,* $X(\alpha)^{(j)} \stackrel{\mathrm{d}}{=} X(\alpha)$, *and* $(V_1, \ldots, V_d)$ *is a Dirichlet distribution on the simplex* $\Delta = \{(s_1, \ldots, s_d) : s_j \ge 0, s_1 + \cdots + s_d = 1\}$ *with density*

$$f(s_1, \ldots, s_d) = (d-1)!(s_1 \cdots s_d),$$

such that

$$\frac{B_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \xrightarrow{d} X(\alpha).$$

The properties of the internal profile of the binary case extend to the d-ary one. Again one has to distinguish between the range $\alpha_{d,-} < \alpha < 1$ and $1 < \alpha < \alpha_{d,+}$.

Next we consider the class of increasing trees that are determined by the generating series $\Phi(x) = (1-x)^{-r}$, where r is a positive real parameter (see Section 1.3.3). The case $r = 1$ corresponds to plane oriented recursive trees. The generating function $y(x) = \sum_{n \ge 1} y_n x^n/n! = 1 - \bigl(1 - (r+1)x\bigr)^{1/(r+1)}$ satisfies the differential equation $y' = 1/(1-y)^r$ and we have $y_n = n!(-1)^{n-1}(r+1)^n \binom{1/(r+1)}{n}$. Equivalently we can work with the function

$$Y(x) = y'(x) = \sum_{n \ge 0} y_{n+1} \frac{x^n}{n!} = \frac{1}{(1-(r+1)x)^{\frac{r}{r+1}}},$$

that satisfies the differential equation

$$Y'(x) = rY(x)^{\frac{1}{r}+2}.$$

Let $X_{n,k}$ denote the corresponding profile. Then the generating function

$$Y(x,u) = \sum_{n \ge 0} \mathbb{E}\, u^{X_{n+1,k}}\, y_{n+1} \frac{x^n}{n!}$$

satisfies the differential equation

$$\frac{\partial Y_{k+1}(x,u)}{\partial x} = rY_{k+1}^{\frac{1}{r}+1} Y_k(x,u).$$

The analysis of this class of increasing trees is more involved than in the previous cases (compare with Hwang [106], Sulzbach [194] and Schopp [189]). Nevertheless it is possible to obtain a corresponding results for the limit.

Theorem 6.43. *Let* $X_{n,k}$ *denote the profile of increasing trees defined by the generating series* $\Phi(x) = (1-x)^{-r}$, *where* $1/r$ *is a positive integer, and let* $\alpha = \alpha_0 > 0$ *be the unique solution of the equation*

$$\alpha \left(\log \alpha + \log \frac{r+1}{r} - 1 \right) = \frac{1}{r+1}.$$

Then for $0 < \alpha < \alpha_0$ *there exists a random variable* $X(\alpha)$ *such that*

$$\frac{X_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \xrightarrow{d} X(\alpha). \tag{6.60}$$

The distribution of $X(\alpha)$ *is determined by the stochastic fixed point equation*

$$X(\alpha) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{\frac{1}{r}+1} V_i^{\alpha+\frac{1}{r+1}} X(\alpha)^{(i)} + \frac{r+1}{r} \alpha V_{\frac{1}{r}+2} X(\alpha)^{(\frac{1}{r}+2)}$$

with $\mathbb{E}\, X(\alpha) = 1$*, where* $X(\alpha)^{(1)}, \ldots, X(\alpha)^{(\frac{1}{r}+2)}$*,* $(V_1, \ldots, V_{\frac{1}{r}+2})$ *are independent,* $X(\alpha)^{(j)} \stackrel{\mathrm{d}}{=} X(\alpha)$*, and* $(V_1, \ldots, V_{\frac{1}{r}+2})$ *are given by* $V_j = G_j / (\sum_{i=1}^{\frac{1}{r}+2} G_i)$ *with* G_i *independent and gamma distributed* $\Gamma(\frac{r}{r+1}, \frac{r}{r+1})$.[6]

The proof method of Schopp [189] is based on an martingale approach and leads to an optimal result (almost sure convergence), but only if $1/r$ is a positive integer. If $1/r$ is not an integer then one can use a moment method in order to obtain (6.60) at least in the range $0 < \alpha < r/(r+1)$ (compare with [106]). Note also that $\alpha = r/(r+1)$ is the level, where most of the nodes are concentrated. As in the previous cases we obtain a central limit theorem for the insertion depth D_n with $\mathbb{E}\, D_n = \frac{r}{r+1} \log n + O(1)$ and $\mathbb{V}\mathrm{ar}\, D_n = \frac{r}{r+1} \log n + O(1)$.

Finally we discuss the profile of fringe balanced m-ary search trees as described in Section 1.4.2. Recall that every (internal) node in an m-ary search tree can hold up to $m-1$ keys. Hence, there are m different kinds of nodes and, thus, m different possible kinds of profiles. For simplicity we only discuss the key profile $X_{n,k}$, that is the number of keys that are stored in nodes with depth k.

By construction (see Section 1.4.2) we have the recurrence

$$X_{n,k} \stackrel{\mathrm{d}}{=} X^{(1)}_{V_n^{(1)},k-1} + X^{(2)}_{V_n^{(2)},k-1} + \cdots + X^{(m)}_{V_n^{(m)},k-1}, \tag{6.61}$$

where $(X^{(1)}_{n,k}), \ldots, (X^{(1)}_{n,k}), (V_n^{(1)}, \ldots, V_n^{(m)})$ are independent, $(X^{(1)}_{n,k}) \stackrel{\mathrm{d}}{=} (X_{n,k})$, and where the splitter $(V_n^{(1)}, \ldots, V_n^{(m)})$ has distribution

$$\mathbb{P}\{(V_n^{(1)}, \ldots, V_n^{(m)}) = (n_1, \ldots, n_m)\} = \frac{\binom{n_1}{t} \cdots \binom{n_m}{t}}{\binom{n}{mt+m-1}}. \tag{6.62}$$

Let

$$W_n(z) = \sum_{k \ge 0} X_{n,k} z^k$$

[6] A non-negative random variable X is gamma distributed $\Gamma(\alpha, \beta)$ with positive parameters α, β if the density of the distribution is given by $f(x) = x^{\alpha-1} e^{-x/\beta} / (\beta^\alpha \Gamma(\alpha))$ for $x > 0$.

denote the corresponding profile polynomial. By (6.61) it is recursively given by $W_n(z) = n$ for $n \le m-1$ and

$$W_n(z) \stackrel{\mathrm{d}}{=} zW^{(1)}_{V_n^{(1)}}(z) + zW^{(2)}_{V_n^{(2)}}(z) + \cdots + zW^{(m)}_{V_n^{(m)}}(z) + m - 1, \qquad n \ge m, \quad (6.63)$$

where $W_\ell^{(j)}(z)$, $j = 1, \ldots, m$, are independent copies of $W_\ell(z)$ that are independent of $(V_n^{(1)}, \ldots, V_n^{(m)})$, $\ell \ge 0$. From this relation we obtain a recurrence for the expected profile polynomial $\mathbb{E}\, W_n(z)$. We have, for $n \ge mt + m - 1$ and $z \in \mathbb{C}$,

$$\mathbb{E}\, W_n(z) = mz \sum_{\ell=0}^{n-1} \frac{\binom{\ell}{t}\binom{n-\ell-1}{(m-1)t+m-2}}{\binom{n}{mt+m-1}} \mathbb{E}\, W_\ell(z) + m - 1. \qquad (6.64)$$

As in the previous cases we use the expected profile polynomial to get asymptotic information on the expected profile. Set

$$F(\theta) = \frac{t!}{m(mt+m-1)!}(\theta+t)(\theta+t+1)\cdots(\theta+mt+m-2), \qquad (6.65)$$

and let $\lambda_j(z)$, $j = 1, \ldots, (m-1)(t+1)$ denote the roots of $F(\theta) = z$ (counted with multiplicities), arranged in decreasing order of the real parts: $\Re\lambda_1(z) \ge \Re\lambda_2(z) \ge \ldots$. Further, let D_s, for real s, be the set of all complex z such that $\Re\lambda_1(z) > s$ and $\Re\lambda_1(z) > \Re\lambda_2(z)$ (in particular, $\lambda_1(z)$ is a simple root). It can be seen that the set D_s is open and that $\lambda_1(z)$ is an analytic function of $z \in D_s$. If $z \in D_s$ is real, then $\lambda_1(z)$ has to be real (and thus $> s$), because otherwise $\overline{\lambda_1(z)}$ would be another root with the same real part.

Lemma 6.44. *Let $W_n(z) = \sum_{k\ge0} X_{n,k} z^k$ denote the (random) profile polynomials.*

1. *If K is a compact subset of D_1 then there exists $\delta > 0$ and an analytic function $E(z)$ such that*

$$\mathbb{E}\, W_n(z) = n^{\lambda_1(z)-1}\left(E(z) + O(n^{-\delta})\right) \qquad (6.66)$$

uniformly for $z \in K$.

2. *K is a compact subset of $\mathbb{C}$. Then there exists $D \ge 0$ such that*

$$|\mathbb{E}\, W_n(z)| = O\left(n^{\max\{\Re(\lambda_1(z))-1,0\}}(\log n)^D\right) \qquad (6.67)$$

uniformly for $z \in K$.

We indicate a possible proof for the first part of the lemma. We introduce the generating function $\Psi(\zeta; z) = \sum_{n\ge0} \mathbb{E}\, W_n(z)\zeta^n$. Let $\Lambda(\theta; z)$ be the polynomial (in θ) of degree $r := mt + m - 1$ defined by

$$\begin{aligned}\Lambda(\theta;z) &= \theta(\theta+1)\cdots(\theta+mt+m-2)\\ &\quad - mz\frac{(mt+m-1)!}{t!}\theta(\theta+1)\cdots(\theta+t-1). \\ &= \frac{m(mt+m-1)!}{t!}\theta(\theta+1)\cdots(\theta+t-1)\left(F(\theta)-z\right)\end{aligned} \tag{6.68}$$

and let ϑ denote the differential operator $(1-\zeta)\frac{d}{d\zeta}$. Then (6.64) is equivalent to the differential equation (in ζ, with z fixed)

$$\Lambda(\vartheta;z)\Psi(\zeta;z) = (m-1)r!\,(1-\zeta)^{-1}. \tag{6.69}$$

The functions $(1-\zeta)^{-\lambda_j(z)}$ are then solutions of the corresponding homogeneous differential equation. With the help of this observation it follows that (for ζ in a Δ region and for $z \in D_1$)

$$\Psi(\zeta;z) = E(z)(1-\zeta)^{-\lambda_1(z)} + O\left(|1-\zeta|^{-\Re(\lambda_2(z))}\right).$$

This implies the first part of Lemma 6.44 by using the transfer lemma (Lemma 2.12). The second part can be similarly proved by estimating the appearing Cauchy integral instead of applying the transfer lemma (see [68]).

By using Lemma 6.44 we derive bivariate asymptotic expansions for $\mathbb{E}\,X_{n,k}$ in a large range. For convenience we use the abbreviation

$$\alpha_0 = \left(\frac{1}{t+1}+\frac{1}{t+2}+\cdots+\frac{1}{(t+1)m-1}\right)^{-1}.$$

Lemma 6.45. *Suppose that α_1,α_2 with $\alpha_0<\alpha_1<\alpha_2<\infty$ are given and let $\beta(\alpha)$ be defined by $\beta(\alpha)\lambda_1'(\beta(\alpha))=\alpha$. Then*

$$\mathbb{E}\,X_{n,k} = \frac{E(\beta(\alpha_{n,k}))n^{\lambda_1(\beta(\alpha_{n,k}))-\alpha_{n,k}\log(\beta(\alpha_{n,k}))-1}}{\sqrt{2\pi(\alpha_{n,k}+\beta(\alpha_{n,k})^2\lambda_1''(\beta(\alpha_{n,k})))\log n}}\left(1+O((\log n)^{-1/2})\right)$$

uniformly for $\alpha_{n,k}=k/\log n \in [\alpha_1,\alpha_2]$ as $n,k\to\infty$.

The proof is a simple application of Cauchy's formula

$$\mathbb{E}\,X_{n,k} = \frac{1}{2\pi i}\int_{|z|=\beta}\mathbb{E}\,W_n(z)z^{-k-1}\,dz,$$

where β is chosen to be $\beta(\alpha_{n,k})$, that is, the saddle point of the dominant part of the integrand:

$$n^{\lambda_1(z)}z^{-k} = e^{\lambda_1(z)\log n - k\log z}.$$

In order to formulate a limit theorem for the profile $X_{n,k}$ we need the solution of the distributional fixed point equation

$$Y(z) \stackrel{\mathrm{d}}{=} zV_1^{\lambda_1(z)-1}Y^{(1)}(z)+zV_2^{\lambda_1(z)-1}Y^{(2)}(z)+\cdots+zV_m^{\lambda_1(z)-1}Y^{(m)}(z), \quad (6.70)$$

where $Y^{(1)}(z), \ldots, Y^{(m)}(z), (V_1, \ldots, V_m)$ are independent, $Y^{(j)}(z) \stackrel{\mathrm{d}}{=} Y(z)$, and where $(V_1, \ldots, V_m)$ is supported on the simplex $\Delta = \{(s_1, \ldots, s_m) : s_j \geq 0, s_1 + \cdots + s_m = 1\}$ with density

$$f(s_1, \ldots, s_m) = \frac{((t+1)m-1)!}{(t!)^m} (s_1 \cdots s_m)^t.$$

In particular one first shows that for suitable $z \in \mathbb{C}$

$$\frac{W_n(z)}{\mathbb{E}\, W_n(z)} \xrightarrow{d} Y(z)$$

(compare with [68]). Finally, by applying Cauchy's formula one then derives a limit theorem for $X_{n,k}$.

Theorem 6.46. *Let $m \geq 2$ and $t \geq 0$ be given integers and let $(X_{n,k})_{k\geq 0}$ be the profile of the corresponding random search tree with n keys.*

Set $I = \{\beta > 0 : 1 < \lambda_1(\beta^2) < 2\lambda_1(\beta) - 1\}$, $I' = \{\beta\lambda_1'(\beta) : \beta \in I\}$, and let $\beta(\alpha) > 0$ be defined by $\beta(\alpha)\lambda_1'(\beta(\alpha)) = \alpha$. Then for every $\alpha \in I'$

$$\frac{X_{n,\lfloor \alpha \log n \rfloor}}{\mathbb{E}\, X_{n,\lfloor \alpha \log n \rfloor}} \xrightarrow{d} Y(\beta(\alpha)). \quad (6.71)$$

A corresponding result holds for several different kinds of profiles according to the m different kinds of nodes that appear in m-ary search trees. Theorem 6.46 generalises to a functional limit theorem, too (see [68]).

6.5.2 The Height of Binary Search Trees and Related Trees

In the previous section we have discussed the profile of d-ary recursive trees, generalised plane oriented recursive trees and m-ary search trees. For all these classes of random trees we have observed a structure that is similar to that of recursive trees. Since the height is closely related to the profile we expect a similar phenomenon for the height. All these trees are so-called $\log n$-trees, that is, the height is of order $\log n$. The second observation is that the height is highly concentrated around its mean.

We start with d-ary recursive trees that include the class of binary search trees. Besides the height H_n there is a second level of interest. The saturation level $\overline{H}_n$ is the maximum level with $I_{n,k} = d^k$, that is, up to this level the tree is a complete d-ary tree.

Theorem 6.47. *The height H_n and the saturation level $\overline{H}_n$ of random d-ary recursive trees have expected value*

$$\mathbb{E}\, H_n \sim \alpha_{d,+} \log n \quad \text{and} \quad \mathbb{E}\, \overline{H}_n \sim \alpha_{d,-} \log n, \quad (6.72)$$

where $0 < \alpha_{d,-} < d < \alpha_{d,+}$ *are the solutions of the equation*

$$\alpha \log\left(\frac{de}{(d-1)\alpha}\right) = \frac{1}{d-1}.$$

Furthermore there are exponential tail estimates of the form

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge \eta\} = O\left(e^{-c\eta}\right) \quad \text{and} \quad \mathbb{P}\{|\overline{H}_n - \mathbb{E}\,\overline{H}_n| \ge \eta\} = O\left(e^{-c\eta}\right)$$

with some $c > 0$.

Remark 6.48 *We want to note that in the binary case* $\Phi(x) = (1+x)^2$ *(that is equivalent to binary search trees) various results of that kind (and even much more precise ones) are well known (see [49, 61, 181]). For example, it is known that*

$$\mathbb{E}\, H_n = \alpha_{2,+} \log n - \frac{3\alpha_{2,+}}{2(\alpha_{2,+}-1)} \log\log n + O(1).$$

Of course, we expect similar relations for general d-ary recursive trees.

The first breakthrough was due to Devroye [49] who related the height of binary search trees to branching random walks and could prove that $H_n/\log n \to \alpha_{2,+} = 4.331...$ *a.s. Previously Pittel [172] had shown that* $H_n/\log n \to \gamma$ *for some* $\gamma \le \alpha_{2,+}$. *Based on numerical data Robson conjectured that the variance* $\mathbb{V}\mathrm{ar}\, H_n$ *is bounded. In fact, he could prove (see [185]) that there is an infinite subsequence for which* $\mathbb{E}|H_n - \mathbb{E}\, H_n|$ *stays bounded. Eventually Robson's conjecture was independently solved by Reed [181] and (few month later) by Drmota [62].*

Nowadays the height constant is known in very general setting, for example, in [27] it is shown that $H_n/\log n \to \alpha_{d,+}$ *for all polynomial classed of increasing trees of degree d.*

The analysis of the height H_n of d-ary recursive trees is very similar to that of recursive trees. The generating functions

$$y_k(x) = \sum_{n\ge 1} \mathbb{P}\{H_n \le k\} y_n \frac{x^n}{n!}$$

with $y_n = (-1)^n n!(d-1)^n \binom{-1/(d-1)}{n}$ satisfy $y_0(x) = 0$ and

$$y_{k+1}'(x) = (1 + y_k(x))^d, \quad y_{k+1}(x) = 0$$

(compare also with Section 6.2.2). For convenience we will work with $Y_k(x) = 1 + y_k(x)$ that satisfy $Y_{k+1}'(x) = Y_k(x)^d$ and $Y_{k+1}(0) = 1$. Note that by definition

$$\lim_{k\to\infty} Y_k(x) = Y(x) = \frac{1}{(1-(d-1)x)^{\frac{1}{d-1}}}.$$

Thus, one might expect that the behaviour of the sequence $Y_k(1/(d-1))$ will play an essential role in the analysis as it was with $Y_k(1)$ in the case of recursive trees. This is actually true. In particular Lemma 6.33 and Lemma 6.34 have direct counterparts. Hence the limit

$$\alpha_0 = \lim_{k\to\infty} \frac{Y_{k+1}(1/(d-1))}{Y_k(1/(d-1))} \tag{6.73}$$

exists. Furthermore, Lemma 6.37 generalises to

$$\mathbb{P}\{H_n \le k\} = O\left(\frac{Y_k(\frac{1}{d-1})}{n^{1/(d-1)}}\right) \quad \text{and} \quad \mathbb{P}\{H_n > k\} = O\left(\frac{n}{Y_k(\frac{1}{d-1})^{d-1}}\right). \tag{6.74}$$

The essential part of the proof is to show that $\alpha_0 > 1$. Here we use the estimate

$$\delta_k(x) = Y(x) - Y_k(x) \le d^k \sum_{\ell > k} \frac{1}{\ell!}\left(\frac{1}{d-1}\log\frac{1}{1-(d-1)x}\right)^{\ell},$$

which follows by induction with the help of the inequality $\delta_{k+1}'(x) \le dY(x)^{d-1}\delta_k(x)$. If we set $x_k = \frac{1}{d-1}\left(1 - A^{-k}\right)$ with $A = e^{1/\alpha_{d,+}}$ then this inequality implies

$$Y_k(1/(d-1)) \ge Y_k(x_k) \ge A^{k/(d-1)}\left(1 - O\left(k^{-1/2}\right),\right)$$

and consequently $\alpha_0 \ge A^{1/(d-1)} > 1$.

By combining (6.73) and (6.74) it follows (as in the case of recursive trees) that

$$\mathbb{E}\, H_n = \max\left\{k : Y_k(1/(d-1)) \le n^{1/(d-1)}\right\} + O(1) \sim \frac{\log n}{(d-1)\log\alpha_0}$$

and that we have exponential tail estimates of the form

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge \eta\} = O\left(e^{-c\eta}\right).$$

It remains to show that $\alpha_0 = A^{1/(d-1)}$. Since $\alpha_{d,+} = ((d-1)\log A)^{-1}$, this will imply $\mathbb{E}\, H_n \sim \alpha_{d,+}\log n$. For this purpose we consider solutions of the integral equation

$$y^{\frac{1}{d-1}}F(y/A) = \frac{\Gamma\left(\frac{d}{d-1}\right)}{\Gamma\left(\frac{1}{d-1}\right)^d}\int_{y_1+\cdots+y_d=y,\, y_j\ge 0}\prod_{j=1}^{d}\left(F(y_j)y_j^{\frac{1}{d-1}-1}\right)d\mathbf{y}, \tag{6.75}$$

for $1 < A < e^{1/\alpha_{d,+}}$ and auxiliary functions

$$\overline{Y}_k(x) = A^{k/(d-1)}\Phi\left(A^k\left(x - \frac{1}{d-1}\right)\right),$$

where

$$\Phi(u) = \frac{1}{(d-1)^{\frac{1}{d-1}}\Gamma\left(\frac{1}{d-1}\right)} \int_0^\infty F(y)\, y^{\frac{1}{d-1}-1} e^{-uy}\, dy.$$

These functions satisfy the recurrence $\overline{Y}'_{k+1}(x) = \overline{Y}^d_k(x)$ and can be used to provide upper bounds for $Y_k(1/(d-1))$ (compare with Lemmas 6.38–6.40, see also [57]).

Remark 6.49 *An even more refined analysis (see [57]) shows that*

$$\mathbb{P}\{H_n \le k\} = F\big((d-1)n/Y_k(1/(d-1))^{d-1}\big) + o(1)$$

uniformly for all $k \ge 0$ as $n \to \infty$, where $F(y)$ satisfies the integral equation (6.75) for $A = e^{1/\alpha_{d,+}}$.

The analysis of the saturation level $\overline{H}_n$ of d-ary recursive trees runs along similar lines. Here one has to study the generating functions

$$\overline{y}_k(x) = \sum_{n\ge 1} \mathbb{P}\{\overline{H}_n \ge k\}\, y_n \frac{x^n}{n!}$$

that satisfy the recurrence

$$\overline{y}'_{k+1}(x) = \overline{y}_k(x)^d, \quad \overline{y}_{k+1}(0) = 0$$

with initial condition $\overline{y}_0(x) = y(x)$. The major difference between the analysis of the height and the saturation level is that one cannot consider $\overline{y}_k(1/(d-1))$, since the function $\overline{y}_k(x)$ is singular at $x = 1/(d-1)$. Instead, one can study the behaviour of the sequence $\overline{y}_k(\overline{x}_k)$ that is determined by the equation

$$\overline{y}_k(\overline{x}_k) = \frac{y(\overline{x}_k)}{2}.$$

For details see [60] and [36].

Next we turn to generalised plane oriented recursive trees.

Theorem 6.50. *The height H_n of random increasing trees defined by the generating series $\Phi(x) = (1-x)^{-r}$ (for some real $r > 0$) has expected value*

$$\mathbb{E}\, H_n \sim \alpha_r \log n, \tag{6.76}$$

where $\alpha_r > 0$ is the solution of the equation

$$\alpha_r \left(\log \alpha_r + \log \frac{r+1}{r} - 1 \right) = \frac{1}{r+1}.$$

Furthermore there are exponential tail estimates of the form

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge \eta\} = O\left(e^{-c\eta}\right)$$

with some $c > 0$.

Here the generating function $y(z) = \sum_{n\ge1} y_n z^n/n!$ satisfies $y'(z) = (1-y(z))^{-r}$ and is explicitly given by $y(z) = 1 - \big(1-(r+1)z\big)^{1/(r+1)}$ with coefficients $y_n = n!(-1)^{n-1}(r+1)^n\binom{1/(r+1)}{n}$. The height distribution is then encoded in the generating functions

$$y_k(x) = \sum_{n\ge0} y_n \mathbb{P}\{H_n \le k\}\frac{x^n}{n!},$$

which are given by $y_0(z) = 0$ and recursively by

$$y'_{k+1}(z) = \frac{1}{(1-y_k(z))^r}, \qquad y_{k+1}(0) = 0.$$

By taking derivatives it follows that

$$y''_{k+1}(z) = r\big(y'_{k+1}(z)\big)^{1+\frac{1}{r}} y'_k(z).$$

If we set

$$Y_k(z) = y'_k(z) = \sum_{n\ge0} y_{n+1}\mathbb{P}\{H_{n+1}\le k\}\frac{z^n}{n!}$$

then we have $Y_1(z) = 1$ and the recurrence relation

$$Y'_{k+1}(z) = r\,Y_{k+1}(z)^{1+\frac{1}{r}} Y_k(z), \qquad Y_{k+1}(0) = 1. \tag{6.77}$$

This equation looks like a mixture of the corresponding generating functions for recursive trees and d-ary recursive trees. Therefore a similar analysis applies (see also [60, 57]) and we obtain the result.

Remark 6.51 *If $r = \frac{A}{B} > 0$ is a rational number (with positive coprime integers A, B) then we have (uniformly for all $k \ge 0$ as $n \to \infty$)*

$$\mathbb{P}\{H_n \le k\} = G\big((r+1)n/(y'_k(r/(r+1)))^{1+\frac{1}{r}}\big) + o(1)$$

with

$$G(y) = \frac{\Gamma\left(\frac{A}{A+B}\right)}{\Gamma\left(\frac{1}{A+B}\right)^A}\int_{z_1+\cdots+z_d=1,z_j\ge0}\prod_{j=1}^{d}\left(F(yz_j)z_j^{\frac{1}{A+B}-1}\right)d\mathbf{z}$$

and where $F(y)$ satisfies the integral equation

$$y^{\frac{1}{d-1}}F(ye^{-1/\alpha_r}) = \frac{\Gamma\left(1+\frac{1}{A+B}\right)}{\Gamma\left(\frac{1}{A+B}\right)^{A+B+1}} \times \int_{y_1+\cdots+y_{A+B+1}=y,y_j\ge0}\prod_{j=1}^{B+1}\left(F(y_je^{-1/\alpha_r})y_j^{\frac{1}{A+B}-1}\right) \times \prod_{\ell=B+2}^{A+B+1}\left(F(y_\ell)y_\ell^{\frac{1}{A+B}-1}\right)d\mathbf{y}.$$

Finally we consider fringe balanced m-ary search trees.

Theorem 6.52. *The height H_n and the saturation level $\overline{H}_n$ of random fringe balanced m-ary search trees (with parameter t) have expected value*

$$\mathbb{E}\, H_n \sim \alpha_{m,t,+} \log n \quad \textit{and} \quad \mathbb{E}\, \overline{H}_n \sim \alpha_{m,t,-} \log n, \tag{6.78}$$

where $0 < \alpha_{m,t,-} < \alpha_{m,t,+}$ are given by

$$\alpha_{m,t,-} = \sum_{j=0}^{(m-1)(t+1)-1} \frac{1}{\beta_1 + t + 1 + j},$$
$$\alpha_{m,t,+} = \sum_{j=0}^{(m-1)(t+1)-1} \frac{1}{\beta_2 + t + 1 + j},$$

where $\beta_1 > 0$ and $\beta_2 < 0$ are the solutions of the equation

$$\sum_{j=0}^{(m-1)(t+1)-1} \log(\beta+t+1+j) - \log\left(\frac{(m(t+1))!}{(t+1)!}\right) = \sum_{j=0}^{(m-1)(t+1)-1} \frac{\beta}{\beta+t+1+j}. \tag{6.79}$$

Furthermore there are exponential tail estimates of the form

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge \eta\} = O\left(e^{-c\eta}\right) \quad \textit{and} \quad \mathbb{P}\{|\overline{H}_n - \mathbb{E}\, \overline{H}_n| \ge \eta\} = O\left(e^{-c\eta}\right)$$

with some $c > 0$.

In this case the generating functions

$$y_k(x) = \sum_{n\ge 0} \mathbb{P}\{H_n^{(m,t)} \le k\}\, x^n \quad \text{and} \quad \overline{y}_k(x) = \sum_{n\ge 0} \mathbb{P}\{\overline{H}_n^{(m,t)} \ge k\}\, x^n$$

satisfy the recurrence relation

$$y_{k+1}^{(m(t+1)-1)}(x) = \frac{(m(t+1)-1)!}{(t!)^m} \left(y_k^{(t)}(x)\right)^m, \tag{6.80}$$
$$\overline{y}_{k+1}^{(m(t+1)-1)}(x) = \frac{(m(t+1)-1)!}{(t!)^m} \left(\overline{y}_k^{(t)}(x)\right)^m, \tag{6.81}$$

with initial conditions

$$y_0(x) = 1,\; y_k(0) = y_k'(0) = \cdots = y_k^{(m-1)}(0) = 1,$$

resp.

$$\overline{y}_0(x) = \frac{x}{1-x},\; y_k(0) = y_k'(0) = \cdots = y_k^{(m-1)}(0) = 0.$$

Thus, we are again in a similar situation as in the above cases (for details we refer to [36]).

7

Tries and Digital Search Trees

Digital trees like tries or digital search trees are important in many computer science applications like data compression, pattern matching or hashing. For example, the popular Lempel-Ziv compression scheme [208] is strongly related to digital search trees.

The basic idea of the Lempel-Ziv algorithm is to partition a sequence over a finite alphabet into phrases (blocks) of variable sizes such that a new block is the shortest substring not seen in the past as a phrase. For example, the string 110010100010001000 is parsed into (1)(10)(0)(101)(00)(01)(000)(100). This algorithm can be implemented efficiently by using the digital tree structure (compare with Figure 1.13). Assume that the first phrase of the Lempel-Ziv scheme is an empty phrase that is stored in the root. When a new phrase is created, the search starts at the root and proceeds down the tree as directed by the input symbols exactly in the same manner as in the digital search tree construction. The search is completed when a branch is taken from an existing tree node to a new node that has not been visited before. Then, an edge and a new node are added to the tree. Phrases created in such a way are stored directly in the nodes of the tree. The example string produces exactly the digital search tree depicted in Figure 1.13.

Digital trees have been widely studied in the literature (see [146, 197] and the references therein). Here we will concentrate on the profile and comment shortly on the height. The motivation of studying the profiles of such trees is multifold. Of course, digital trees are used in various applications (for example, the profile $X_{n,k}$ represents the number of phrases of length k in the Lempel-Ziv'78 built over n phrases). Second, the profile is a fine shape measure closely connected to many other cost measures (height, saturation level, depth, path length, etc.). And finally, the analytic problems are mathematically challenging and lead to interesting distributional results. It is remarkable that the profile process is almost deterministic. In contrast to the previously studied classes of trees (Galton-Watson trees in Chapter 4 and recursive trees in Chapter 6) the normalised profile $B_{n,k}/\mathbb{E}\, B_{n,k}$ converges to 1 and there is a central limit theorem.

The following treatment on tries is based on the work by Park, Hwang, Nicodème, and Szpankowski [169]. The corresponding result for digital search trees follows some recent work by Drmota and Szpankowski [70].

7.1 The Profile of Tries

Tries are prototype data structures useful for many indexing and retrieval purposes. They were first proposed by de la Briandais [44] in the late 1950s for information processing; Fredkin [86] suggested the current name is part of retrieval. Due to their simplicity and efficiency, tries found widespread use in diverse applications (see [87, 130, 146, 197]).

Tries are a natural choices of data structures when the input records involve a notion of alphabets or digits. They are often used to store such data so that future retrieval can be made efficient. Recall that (m-ary) tries can be constructed in the following way. Suppose that a sequence of n strings over an m-ary alphabet, $m \geq 2$, is given. If $n = 0$, then the trie is empty. If $n = 1$ then a single (external) node holding this string is allocated. If $n \geq 1$ then the trie consists of a (internal) root node directing strings to the m subtrees according to the first letter of each string, and strings directed to the same subtree are themselves tries (see [130, 146, 197] for more details). For simplicity, we only deal with binary tries here. Unlike other search trees, such as digital search trees and binary search trees, where records or keys are stored in the internal nodes, the internal nodes in tries are branching nodes used merely to direct records to each sub trie. The keys are all stored in external nodes that are leaves of such tries. A trie has more internal nodes than external nodes (fixed to be n), differing from almost all other search trees.

7.1.1 Generating Functions for the Profile

Our first and main goal is to study the profile of tries, where we have to distinguish, as in the case of binary search trees, between the internal and the external profile. We write $B_{n,k}$ to denote the number of external nodes (leaves) at distance k from the root (external profile) and the number of internal nodes at distance k from the root is denoted by $I_{n,k}$ (internal profile).

In the example (depicted in Figure 1.14) we have $B_{8,0} = B_{8,1} = 0$, $B_{8,2} = B_{8,3} = 1$, $B_{8,4} = 2$, $B_{8,5} = 6$ and $I_{8,0} = 1$, $I_{8,1} = 2$, $I_{8,2} = I_{8,3} = 3$, $I_{8,4} = 2$.

We also assume the simplest probabilistic model, namely the standard Bernoulli model. More precisely, we assume that the input is a sequence of n independent and identically distributed random variables, each being composed of an infinite sequence of Bernoulli random variables with mean p, where $0 < p < 1$ is the probability of a 1 and $q = 1 - p$ is the probability of a 0. The corresponding trie constructed from these n bit-strings is called a random trie. This simple model may seem to be too idealised for practical purposes, however, the typical behaviours of such a model often hold for more general models such as Markovian or dynamical sources, although the technicalities are usually more involved (see for example [41, 50, 51, 111]).

We start with the external profile $B_{n,k}$. Let

$$P_{n,k}(u) = \mathbb{E}\, u^{B_{n,k}}$$

be the corresponding probability generating function. Then by using the recursive definition of tries we have

$$P_{n,k}(u) = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} P_{j,k-1}(u) P_{n-j,k-1}(u), \qquad (n \ge 2, k \ge 1) \quad (7.1)$$

with initial conditions $P_{0,k}(u) = 1$ for $k \ge 0$, $P_{1,0}(u) = u$, $P_{1,k}(u) = 1$ for $k \ge 1$, and $P_{n,0}(u) = 1$ for $n \ge 1$. Recall that the first digit determines whether the corresponding string is put to the left or to the right subtree. Due to the independence assumption the number of strings where the first digit is 0 follows a binomial distribution. The splitting probabilities are thus given by $\binom{n}{j} p^j q^{n-j}$.

From (7.1) one gets directly a recurrence relation for the exponential generating functions

$$G_k(x,u) = \sum_{n\ge 0} P_{n,k}(u)\, \frac{x^n}{n!}$$

of the form

$$G_k(x,u) = G_{k-1}(px,u)G_{k-1}(qx,u) + (P_{1,k}(u) - P_{1,k-1}(u))x, \qquad (k \ge 1), \quad (7.2)$$

with initial condition $G_0(x,u) = e^x + x(u-1)$. With the help of the initial conditions of $P_{1,k}(u)$ it follows that

$$G_1(x,u) = e^x + \left(e^{px}qx + e^{qx}px - 1\right)(u-1) + pqx^2(u-1)^2$$

and

$$G_k(x,u) = G_{k-1}(px,u)G_{k-1}(qx,u), \qquad (k \ge 2). \quad (7.3)$$

The corresponding probability generating functions for the internal profile,

$$P^{[I]}_{n,k}(u) = \mathbb{E}\, u^{I_{n,k}},$$

satisfy the same recurrence

$$P^{[I]}_{n,k}(u) = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} P^{[I]}_{j,k-1}(u) P^{[I]}_{n-j,k-1}(u), \qquad (n \ge 2, k \ge 1),$$

but with initial conditions $P^{[I]}_{n,0}(u) = u$ for $n \ge 2$ and $P^{[I]}_{n,k}(u) = 1$ for $n \le 1$ and $k \ge 0$. Similarly to the external profile we introduce the exponential generating functions

$$G_k^{[I]}(x,u) = \sum_{n\ge 0} P_{n,k}^{[I]}(u)\,\frac{x^n}{n!},$$

that satisfy

$$G_k^{[I]}(x,u) = G_{k-1}^{[I]}(px,u)G_{k-1}^{[I]}(qx,u), \qquad (k \ge 1), \tag{7.4}$$

with initial condition $G_0^{[I]}(x,u) = ue^x - (1+x)(u-1)$.

We will come back to the relations (7.3) and (7.4) in Section 7.1.3. In order to get a first impression of the behaviour of such profiles we will study the expected profile (see Section 7.1.2). Set

$$E_k(x) = \sum_{n\ge 0} \mathbb{E}\, B_{n,k}\,\frac{x^n}{n!} = \left[\frac{\partial G_k(x,u)}{\partial u}\right]_{u=1}.$$

Then (7.3) implies

$$E_k(x) = e^{qx}E_{k-1}(px) + e^{px}E_{k-1}(qx) \qquad (k \ge 2).$$

Setting

$$\Delta_k(x) = e^{-x}E_k(x) = \sum_{n\ge 0} \mathbb{E}\, B_{n,k}\,\frac{x^n}{n!}e^{-x}$$

this rewrites to

$$\Delta_k(x) = \Delta_{k-1}(px) + \Delta_{k-1}(qx) \qquad (k \ge 2) \tag{7.5}$$

with initial conditions $\Delta_0(x) = xe^{-x}$ and $\Delta_1(x) = pxe^{-px} + qxe^{-qx} - xe^{-x}$.

Note that if x is a positive real number then $\Delta_k(x)$ can be considered as the expected value of the external profile (at level k) if the number n of input strings is Poisson distributed with parameter x. The mapping

$$(a_n)_{n\ge 0} \mapsto \tilde{A}(x) = \sum_{n\ge 0} a_n\,\frac{x^n}{n!}e^{-x}$$

is also called *Poisson transform*. In the context of tries the Poisson transform has turned out to be natural; the relation (7.5) is remarkably easy. With the help of the Mellin transform we will be able to provide an asymptotic solution of $\Delta_k(x)$ (see Section 7.1.2). The final step is then to derive asymptotics for $\mathbb{E}\, B_{n,k}$ from that of $\Delta_k(x)$, that is, we have to invert the Poisson transform. This is of independent interest (compare also with Jacquet and Szpankowski [113]). For example, the Poisson transform of the sequence $a_n = n$ equals $\tilde{A}(x) = x$, and the the Poisson transform of the sequence $b_n = n^2$ is $\tilde{B}(x) = x + x^2$. In these two examples we have $\tilde{A}(n) = a_n$ and $\tilde{B}(n) \sim b_n$. Such a property is surely not generally true (for example, for $a_n = 2^n$) but the crucial observation is that (under suitable assumptions) $\tilde{A}(n)$ and a_n are very close (see [113]). Thus, inversion of the Poisson transform is mostly relatively easy. We can expect that $\mathbb{E}\, B_{n,k}$ is approximated by $\Delta_k(n)$, but one has to keep track of the second parameter k.

From a computational point of view depoissonisation is done with Cauchy integration

$$a_n = \frac{1}{2\pi i} \int_{|x|=r} e^x \tilde{A}(x)\, \frac{dx}{x^{n+1}},$$

where the radius r is usually chosen to be n, since $x = n$ is the saddle point of the function $e^x x^{-n}$. Thus, if $\tilde{A}(x)$ has a regular behaviour, for example polynomial growth, then the integral is concentrated at $x = n$ and we (usually) get

$$\begin{aligned} a_n &\sim \frac{1}{2\pi i} \int\limits_{|x|=n,\, |\arg(x)|\le\epsilon} e^x \tilde{A}(x)\, \frac{dx}{x^{n+1}} \\ &\sim \tilde{A}(n) \cdot \frac{1}{2\pi i} \int\limits_{|x|=n,\, |\arg(x)|\le\epsilon} e^x\, \frac{dx}{x^{n+1}} \\ &\sim \tilde{A}(n), \end{aligned}$$

as expected.

7.1.2 The Expected Profile of Tries

In order to state our main result we need the following notations. For a real number α with $(\log \frac{1}{p})^{-1} < \alpha < (\log \frac{1}{q})^{-1}$, let

$$\rho = \rho(\alpha) = \frac{1}{\log(p/q)} \log \frac{1 - \alpha \log(1/p)}{\alpha \log(1/q) - 1}. \tag{7.6}$$

Equivalently, α and ρ satisfy the equation

$$\alpha = \frac{p^{-\rho} + q^{-\rho}}{p^{-\rho} \log \frac{1}{p} + q^{-\rho} \log \frac{1}{q}}.$$

Furthermore, we set

$$\beta(\rho) = \frac{p^{-\rho} q^{-\rho} \log(p/q)^2}{(p^{-\rho} + q^{-\rho})^2}, \tag{7.7}$$

and

$$\alpha_1 = \frac{1}{\log(1/p)}, \qquad \alpha_0 = \frac{2}{\log \frac{1}{p} + \log \frac{1}{q}},$$

$$\alpha_2 = \frac{p^2 + q^2}{p^2 \log \frac{1}{p} + q^2 \log \frac{1}{q}}, \qquad \alpha_3 = \frac{2}{\log(1/(p^2 + q^2))}.$$

Then we have the following relations.

Theorem 7.1. *Let $\mathbb{E}\,B_{n,k}$ denote the expected external profile in binary random tries with underlying (non-zero) probabilities $p > q = 1-p$. Then:*

1. *If $\alpha_1 + \epsilon \le \frac{k}{\log n} \le \alpha_2 - \epsilon$ (for some $\epsilon > 0$) then we have uniformly*

$$\mathbb{E}\,B_{n,k} = G\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(k^{-1/2}\right)\right),$$

where $G(\rho, x)$ is a non-zero periodic function with period 1 and $\rho_{n,k} = \rho(k/\log n)$.

2. *If $k = \alpha_2\left(\log n + \xi\sqrt{\alpha_2\beta(-2)\log n}\right)$, where $\xi = o((\log n)^{\frac{1}{6}})$ then*

$$\mathbb{E}\,B_{n,k} = 2pqn^2(p^2+q^2)^{k-1}\Phi(\xi)\left(1 + O\left(\frac{1+|\xi|^3}{\sqrt{\log n}}\right)\right),$$

where $\Phi(x)$ denotes the normal distribution function.

3. *If $\frac{k}{\log n} \ge \alpha_2 + \epsilon$ (for some $\epsilon > 0$) then uniformly*

$$\mathbb{E}\,B_{n,k} = 2pqn^2(p^2+q^2)^{k-1}\left(1 + O(n^{-\eta})\right)$$

for some $\eta > 0$.

Theorem 7.2. *Let $\mathbb{E}\,I_{n,k}$ denote the expected internal profile in binary random tries with underlying (non-zero) probabilities $p > q = 1-p$. Then:*

1. *If $\alpha_1 + \epsilon \le \frac{k}{\log n} \le \alpha_0 - \epsilon$ (for some $\epsilon > 0$) then we have uniformly*

$$\mathbb{E}\,I_{n,k} = 2^k - \overline{G}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(k^{-1/2}\right)\right),$$

where $\overline{G}(\rho, x)$ is a non-zero periodic function with period 1 and $\rho_{n,k} = \rho(k/\log n)$.

2. *If $k = \alpha_0\left(\log n + \xi\sqrt{\alpha_0\beta(0)\log n}\right)$, where $\xi = o((\log n)^{\frac{1}{6}})$ then*

$$\mathbb{E}\,I_{n,k} = 2^k\Phi(-\xi)\left(1 + O\left(\frac{1+|\xi|^3}{\sqrt{\log n}}\right)\right).$$

3. *If $\alpha_0 + \epsilon \le \frac{k}{\log n} \le \alpha_2 - \epsilon$ (for some $\epsilon > 0$) then uniformly*

$$\mathbb{E}\,I_{n,k} = \overline{G}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(k^{-1/2}\right)\right).$$

4. *If $k = \alpha_2\left(\log n + \xi\sqrt{\alpha_2\beta(-2)\log n}\right)$, where $\xi = o((\log n)^{\frac{1}{6}})$ then*

$$\mathbb{E}\,I_{n,k} = \frac{1}{2}\Phi(\xi)n^2(p^2+q^2)^k\left(1 + O\left(\frac{1+|\xi|^3}{\sqrt{\log n}}\right)\right).$$

5. *If* $\frac{k}{\log n} \ge \alpha_2 + \epsilon$ *(for some* $\epsilon > 0$*) then uniformly*

$$\mathbb{E}\, I_{n,k} = \frac{1}{2} n^2 (p^2+q^2)^{k-1} \left(1 + O(n^{-\eta})\right)$$

for some $\eta > 0$.

Remark 7.3 *The results stated above are not optimal. For example, the restriction* $\alpha_1 + \epsilon \le \frac{k}{\log n} \le \alpha_2 - \epsilon$ *(in the first part of Theorem 7.1) can be weakened to*

$$\alpha_1 \left(\log n - \log\log\log n - K_n\right) \le k \le \alpha_2 \left(\log n - K_n \sqrt{\log n}\right),$$

where K_n *is any sequence tending to infinity (compare with [169]).*

If we set $\alpha = k/\log n$ then we can rewrite

$$(p^{-\rho} + q^{-\rho})^k n^{-\rho} = n^{\alpha \log(p^{-\rho}+q^{-\rho}) - \rho}.$$

Thus, for $\alpha_0 < \alpha < \alpha_2$ the behaviour of $\mathbb{E}\, B_{n,k}$ and $\mathbb{E}\, I_{n,k}$ is governed by a power of n depending on the ratio $\alpha = k/\log n$. The maximum exponent is obtained for

$$\alpha = \frac{1}{h} = \frac{1}{p \log \frac{1}{p} + q \log \frac{1}{q}},$$

where $h = p \log \frac{1}{p} + q \log \frac{1}{q}$ denotes the entropy of the Bernoulli source. Actually, the expected number of nodes at level $k = \frac{1}{h} \log n$ is of order $n/\sqrt{\log n}$. Thus, as in the case of recursive trees (or binary search trees), almost all nodes are concentrated around this *typical level*.

Let D_n denote the depth of a random node in a random trie with n keys. Then the distribution of D_n is related to the external profile by

$$\mathbb{P}\{D_n = k\} = \frac{\mathbb{E}\, B_{n,k}}{n}.$$

Hence, a direct application of Theorem 7.1 provides an unusual local limit theorem. (Here we also use the notation $h_2 = p(\log \frac{1}{p})^2 + q(\log \frac{1}{q})^2$.)

Theorem 7.4. *Let* D_n *denote the depth of a random node in a binary random trie with underlying (non-zero) probabilities* $p > q = 1 - p$*. Then we have*

$$\mathbb{P}\{D_n = k\} = \frac{G\left(-1, \log_{p/q} p^k n\right)}{\sqrt{2\pi (h_2 - h^2)/h^3 \, \log n}} \exp\left(- \frac{\left(k - \frac{1}{h}\log n\right)^2}{2(h_2-h^2)/h^3 \, \log n} \right) \times \left(1 + O\left(\frac{1}{\sqrt{\log n}} + \frac{\left|k - \frac{1}{h}\log n\right|^3}{(\log n)^2} \right)\right)$$

uniformly for k *and* n *with* $\left|k - \frac{1}{h}\log n\right| = o\left((\log n)^{2/3}\right)$.

The unusualness in this result is the periodic factor. Thus, although the depth D_n follows a central limit theorem (see [110, 111]) it does not follow a corresponding local central limit theorem (compare also with [169]).

The proof of Theorem 7.1 relies on a precise analysis of the recurrence (7.5) for $\Delta_k(x)$ that is based on the Mellin transform and analytic depoissonisation. Let $\Delta_k^*(s)$ denote the Mellin transform

$$\Delta_k^*(s) = \int_0^\infty \Delta_k(x) x^{s-1}\, dx$$

that exists for $s \in \mathbb{C}$ with $\Re(s) > -2$ (if $k \ge 1$). Then (7.5) rewrites to

$$\Delta_k^*(s) = (p^{-s} + q^{-s})\Delta_{k-1}^*(s), \qquad (k \ge 2),$$

with initial condition

$$\Delta_1^*(s) = \Gamma(s+1)(p^{-s} + q^{-s} - 1).$$

Thus, for $k \ge 2$ we explicitly have

$$\Delta_k^*(s) = \Gamma(s+1)(p^{-s} + q^{-s} - 1)(p^{-s} + q^{-s})^{k-1}.$$

Hence, by the inverse Mellin transform (see [81])

$$\Delta_k(x) = \frac{1}{2\pi i} \int_{\rho - i\infty}^{\rho + i\infty} \Gamma(s+1) g(s) (p^{-s} + q^{-s})^k x^{-s}\, ds \tag{7.8}$$

with $\rho > -2$ and where

$$g(s) = 1 - \frac{1}{p^{-s} + q^{-s}}.$$

We are mainly interested in the behaviour of $\Delta_k(x)$ for $x = n$, since by analytic depoissonisation we expect that $\mathbb{E}\, X_{n,k} \sim \Delta_k(n)$.

For convenience we set

$$T(s) = p^{-s} + q^{-s}.$$

For the asymptotic analysis of the integral (7.8) it is natural to choose $\rho = \rho_{n,k}$ as the saddle point of the function

$$T(s)^k n^{-s} = e^{k \log T(s) - s \log n}$$

that is given by the relation

$$\frac{\partial}{\partial s} \left(k \log T(s) - s \log n\right) = 0.$$

Equivalently we have

$$\frac{k}{\log n} = \frac{p^{-\rho} + q^{-\rho}}{p^{-\rho} \log \frac{1}{p} + q^{-\rho} \log \frac{1}{q}},$$

that is, $\rho = \rho_{n,k} = \rho(k/\log n)$.

However, we have to distinguish between several cases. We start with the range $\alpha_1 < k/\log n < \alpha_2$. Here we have $-2 < \rho_{n,k} < \infty$ and we can choose $\rho = \rho_{n,k}$. Another observation is that on the line $\Re(s) = \rho$ there will be infinitely many saddle points

$$s_j = \rho + \frac{2\pi i j}{\log \frac{p}{q}}.$$

This is due to the fact that $T(s_j) = e^{-2\pi i j(\log p)/(\log p/q)} T(\rho)$ and consequently the behaviour of $T(s)^k z^{-s}$ around $s = s_j$ is almost the same as that of $T(s)^k z^{-s}$ around $s = \rho$. This phenomenon gives a periodic leading factor in the asymptotics of $\Delta_k(n)$ and also of $\mu_{n,k} = \mathbb{E}\, B_{n,k}$.

Lemma 7.5. *Suppose that* $\alpha_1 + \epsilon \le \frac{k}{\log n} \le \alpha_2 - \epsilon$ *(for some* $\epsilon > 0$*). Then we have*

$$\Delta_k(n) = G\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(\frac{1}{\sqrt{k}}\right)\right),$$

where

$$G(\rho, x) = \sum_{j\in\mathbb{Z}} g(\rho + it_j)\Gamma(\rho + it_j + 1)e^{-2j\pi i x}$$

is a non-zero periodic function with period 1 *and* $t_j = 2j\pi/\log(p/q)$.

Proof. For convenience we set $J_k(n, s) = n^{-s}\Gamma(s+1)g(s)T(s)^k$. We split the integral (7.8) into two parts where we use the substitution $s = \rho + it$. Let us start with the range $|t| \ge \sqrt{\log n}$ and recall that by Stirling's formula $\Gamma(\rho + 1 + it) = O\left(|t|^{\rho+1/2} e^{-\pi|t|/2}\right)$:

$$\begin{aligned}
\frac{1}{2\pi} \int\limits_{|t|\ge\sqrt{\log n}} J_k(n, \rho + it)\, dt &= O\left(n^{-\rho} T(\rho)^k \int_{\sqrt{\log n}}^{\infty} |\Gamma(\rho + it)\, dt\right) \\
&= O\left(n^{-\rho} T(\rho)^k \int_{\sqrt{\log n}}^{\infty} t^{\rho+1/2} e^{-\pi t/2}\right) \\
&= O\left(n^{-\rho} T(\rho)^k (\log n)^{\rho/2+1/4} e^{-\pi\sqrt{\log n}/2}\right) \\
&= O\left(n^{-\rho} T(\rho)^k e^{-\sqrt{\log n}}\right).
\end{aligned}$$

Next set

$$T_j = \frac{1}{2\pi} \int_{|t-t_j|\le\pi/\log(p/q)} J_k(n, \rho + it)\, dt,$$

where $t_j = \frac{2\pi j}{\log \frac{p}{q}}$. We have to study these integrals for all $|j| \le j_0 = \lfloor \sqrt{\log n}\log(p/q)/(2\pi) \rfloor$.

Since there exists $c_0 > 0$ with

$$p^{-\rho-it} + q^{-\rho-it} \le T(\rho)e^{-c_0(t-t_j)^2}$$

for $|t-t_j| \le \pi/\log(p/q)$, we obtain an upper bound of the integral (for $j \ne 0$)

$$\begin{aligned} T_j' &= \frac{1}{2\pi} \int\limits_{k^{-2/5} \le |t-t_j| \le \pi/\log(p/q)} J_k(n, \rho+it)\,dt \\ &= O\left(|\Gamma(\rho+it_j)| n^{-\rho} T(\rho)^k \int_{k^{-2/5}}^{\infty} e^{-c_0 t^2}\,dt \right) \\ &= O\left(|\Gamma(\rho+it_j)| n^{-\rho} T(\rho)^k k^{-3/5} e^{-c_0 k^{1/5}} \right). \end{aligned}$$

For $j = 0$ we can replace the factor $|\Gamma(\rho+it_j)|$ by 1.

Finally, for $|t-t_j| \le k^{-2/5}$ we use the approximation

$$\begin{aligned} J_k(n,\rho+it) &= \Gamma(\rho+it)g(\rho+it)n^{\rho+it}T(\rho+it)^k \\ &= \Gamma(\rho+it)g(\rho+it)e^{-it_j \log(p^k n)} n^{-\rho+i(t-t_j)} T(\rho+i(t-t_j))^k \\ &= \Gamma(\rho+it_j)g(\rho+it_j)e^{-it_j \log(p^k n)} n^{-\rho} T(\rho)^k e^{-\frac{1}{2}\beta(\rho)(t-t_j)^2} \\ &\qquad \times \left(1 + O(|t-t_j|) + O(k|t-t_j|^3)\right) \end{aligned}$$

and standard saddle point techniques to derive the approximation

$$\begin{aligned} T_j'' &= \frac{1}{2\pi} \int\limits_{|t-t_j| \le k^{-2/5}} J_k(n,\rho+it)\,dt \\ &= \Gamma(\rho+it_j)g(\rho+it_j) \frac{n^{-\rho}T(\rho)^k}{\sqrt{2\pi\beta(\rho)k}} e^{-it_j\log(p^k n)} \left(1 + O(k^{-1/2})\right). \end{aligned}$$

Hence we finally get

$$\begin{aligned} \Delta_k(n) &= \sum_{|j| \le j_0} T_j + O\left(n^{-\rho}T(\rho)^k e^{-\sqrt{\log n}}\right) \\ &= \sum_{|j|\le j_0} \Gamma(\rho+it_j)g(\rho+it_j)e^{-it_j\log(p^k n)} \frac{n^{-\rho}T(\rho)^k}{\sqrt{2\pi\beta(\rho)k}} \left(1+O(k^{-1/2})\right) \\ &\quad + O\left(n^{-\rho}T(\rho)^k e^{-\sqrt{\log n}}\right) \\ &= G\left(\rho, \log_{p/q} p^k n\right) \frac{n^{-\rho}T(\rho)^k}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1+O(k^{-1/2})\right), \end{aligned}$$

as proposed.

For the depoissonisation procedure we also need some information about the asymptotic behaviour of $\Delta_k(ne^{i\theta})$. The above proof extends to the range $|\theta| \le \pi/2 - \epsilon$ (for some $\epsilon > 0$). In this range we have uniformly

$$\Delta_k(ne^{i\theta}) = \frac{T(\rho)^k}{\sqrt{2\pi\beta(\rho)k}} \sum_{|j|\le j_0} g(\rho+it_j)\Gamma(\rho+it_j)(ne^{i\theta})^{-\rho-it_j} p^{-ikt_j} \times \left(1 + O\left(k^{-1/2}\right)\right). \tag{7.9}$$

Furthermore, we also ovserve the following upper bound.

Lemma 7.6. *There exits a constant $c > 0$ such that*

$$|e^x \Delta_k(x)| \le e^r \Delta_k(r) e^{-cr\theta^2}$$

uniformly for $r \ge 0$ and $|\theta| \le \pi$, where $x = re^{i\theta}$.

Proof. Set $x = re^{i\theta}$. By (7.5) we can represent $\Delta_k(x)$ by

$$\Delta_k(x) = \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} \Delta_1\left(p^\ell q^{k-1-\ell} x\right).$$

Since $e^x\Delta_1(x)$ is a power series with non-negative coefficients, we have $|e^x\Delta_1(x)| \le e^r\Delta_1(r)$. Consequently, by using the inequality $1-\cos\theta \ge 2\vartheta^2/\pi^2$ (for $|\theta| \le \pi$) it follows for $k \ge 2$

$$\begin{aligned}
|e^x\Delta_k(x)| &\le \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} |e^{x(1-p^\ell q^{k-1-\ell})}| e^{rp^\ell q^{k-1-\ell}} \Delta_1\left(p^\ell q^{k-1-\ell} r\right) \\
&= \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} |e^{(1-p^\ell q^{k-1-\ell})r\cos\theta}| e^{rp^\ell q^{k-1-\ell}} \Delta_1\left(p^\ell q^{k-1-\ell} r\right) \\
&\le \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} |e^{(1-p^\ell q^{k-1-\ell})r(1-2\vartheta^2/\pi^2)}| e^{rp^\ell q^{k-1-\ell}} \Delta_1\left(p^\ell q^{k-1-\ell} r\right) \\
&\le e^{-2r\theta^2(1-p^{k-1})/\pi^2} e^r \Delta_k(r) \\
&\le e^{-2r\theta^2(1-p)/\pi^2} e^r \Delta_k(r).
\end{aligned}$$

This proves the lemma with $c = 2(1-p)/\pi^2$.

By using these preliminaries the proof of the first part of Theorem 7.1 is completed by inverting the Poisson transform. As explained above we use Cauchy integration with the contour $|x| = n$:

$$\begin{aligned}
\mathbb{E}\, B_{n,k} &= \frac{n!}{2\pi i} \int_{|x|=n} e^x \Delta_k(x) \frac{dx}{x^{n+1}} \\
&= \frac{n!\, n^{-n}}{2\pi} \int_{|\theta|\le\pi} e^{ne^{i\theta}} \Delta_k(ne^{i\theta}) e^{-in\theta}\, d\theta.
\end{aligned}$$

Fix $0 < \theta_0 < \pi/2$. Then Lemma 7.6 implies

$$\left| \frac{n!\,n^{-n}}{2\pi} \int\limits_{\theta_0 \le |\theta| \le \pi} e^{ne^{i\theta}} \Delta_k(ne^{i\theta}) e^{-in\theta}\, d\theta \right| \le \Delta_k(n) \frac{n!\,n^{-n}e^n}{2\pi} \int\limits_{\theta_0 \le |\theta| \le \pi} e^{-cn\theta^2}\, d\theta$$

$$= O\left(\Delta_k(n) e^{-c\theta_0^2 n} \right).$$

For the remaining part of the integral we use (7.9) and obtain

$$\begin{aligned}
&\frac{n!\,n^{-n}}{2\pi} \int_{|\theta| \le \theta_0} e^{ne^{i\theta}} \Delta_k(ne^{i\theta}) e^{-in\theta}\, d\theta \\
&\quad = \frac{n^{-\rho} T(\rho)^k}{\sqrt{2\pi\beta(\rho)k}} \sum_{|j| \le j_0} \Gamma(\rho + it_j) g(\rho + it_j) \\
&\quad \times \frac{n!\,n^{-n}}{2\pi} \int_{|\theta| \le \theta_0} e^{ne^{i\theta} - in\theta} e^{i\theta(\rho + it_j)}\, d\theta \cdot \left(1 + O(k^{-1/2})\right) \\
&\quad = \frac{n^{-\rho} T(\rho)^k}{\sqrt{2\pi\beta(\rho)k}} \sum_{|j| \le j_0} \Gamma(\rho + it_j) g(\rho + it_j) \\
&\quad \times \frac{n!\,n^{-n}e^n}{2\pi} \int_{|\theta| \le \theta_0} e^{-\frac{1}{2}n\theta^2} \left(1 + O(n|\theta|^3) + O(|t_j\theta|)\right)\, d\theta \cdot \left(1 + O(k^{-1/2})\right) \\
&\quad = \frac{n^{-\rho} T(\rho)^k}{\sqrt{2\pi\beta(\rho)k}} \sum_{|j| \le j_0} \Gamma(\rho + it_j) g(\rho + it_j) \left(1 + O(|t_j| n^{-1/2}) + O(k^{-1/2})\right) \\
&\quad = \Delta_k(n) \left(1 + O(k^{-1/2})\right).
\end{aligned}$$

This completes the proof of Theorem 7.1 for the range $\alpha_1 + \epsilon \le \frac{k}{\log n} \le \alpha_2 - \epsilon$.

The range $\frac{k}{\log n} \ge \alpha_2 + \epsilon$ can be handled in a similar way. First of all one has to shift the path of integration to the line $\Re(s) = \rho < -2$. Since the integrand has a polar singularity at $s = -2$, we thus have

$$\Delta_k(x) = -g(-2)x^2(p^2 + q^2)^k + \frac{1}{2\pi} \int_{-\infty}^{\infty} J_k(x, \rho + it)\, dt.$$

By definition

$$g(-2) = 1 - \frac{1}{p^2 + q^2} = -\frac{2pq}{p^2 + q^2}.$$

This suggests that the leading term of $\Delta_k(x)$ is of the form

$$2pqx^2(p^2 + q^2)^{k-1},$$

which corresponds (by depoissonisation) to the proposed leading term

$$2pqn^2(p^2+q^2)^{k-1}$$

of $\mathbb{E}\, B_{n,k}$. Therefore we only have to concentrate on the remaining integral.

We set $x = ne^{i\theta}$ and obtain (by splitting up the integral into integrals T_j, etc.)

$$\int_{-\infty}^{\infty} J_k(x;\rho+it)\,dt = O\left(n^{-\rho}T(\rho)^k k^{-1/2}\right)$$

uniformly for $|\theta| \le \theta_0$ (where $\theta_0 < \pi/2$) and $-K' \le \rho \le -2-\epsilon'$ (where $\epsilon' > 0$ and $K' > 2$ are arbitrary constants). In fact, if we choose ρ sufficiently close to -2 (depending of ϵ) we obtain

$$\Delta_k(ne^{i\theta}) = 2pqn^2e^{2i\theta}(p^2+q^2)^{k-1}\left(1+O(n^{-\eta})\right)$$

uniformly for $|\theta| \le \theta_0$ for some $\eta > 0$. Hence, together with Lemma 7.6 and analytic depoissonisation we get

$$\mathbb{E}\, B_{n,k} = 2pqn^2(p^2+q^2)^{k-1}\left(1+O(n^{-\eta})\right)$$

uniformly for $k/\log n \ge \alpha_2 + \epsilon$.

Finally suppose that $k/\log n$ is close to α_2:

$$k = \alpha_2\left(\log n + \xi\sqrt{\alpha_2\beta(-2)\log n}\right),$$

where $\xi = o((\log n)^{1/6})$. We move the line of integration to the saddle point

$$\Re(s) = \rho = \frac{1}{\log(p/q)}\log\frac{1-\alpha\log(1/p)}{\alpha\log(1/q)-1} = -2 - \frac{\xi}{\sqrt{\alpha_2\beta(-2)\log n}} + O\left(\xi^2/\log n\right).$$

First assume that $k < \alpha_2 \log n$, so that $\xi < 0$ and $\rho > -2$. This means that we do not pass the polar singularity, which is located at $s = -2$. Hence, as above we obtain

$$\begin{aligned}\Delta_k(ne^{i\theta}) &= \frac{1}{2\pi}\int_{|t|\le(\log n)^{-2/5}} J_k(ne^{i\theta},\rho+it)\,dt \\ &\quad + O\left(|\Gamma(\rho+1+i(\log n)^{-2/5})|\, n^{-\rho}T(\rho)^k e^{-c_0(\log n)^{1/5}}\right) \\ &\quad + O\left(k^{-1/2}n^{-\rho}T(\rho)^k\right).\end{aligned}$$

Since

$$|\Gamma(\rho+1+i(\log n)^{-2/5})| = O\left(\frac{1}{|\xi|(\log n)^{-1/2}+(\log n)^{-2/5}}\right) = O((\log n)^{2/5}),$$

we can neglect the first error term.

Next we replace the factor $\Gamma(s+1)g(s)$ in $J_k(x,s)$ by

$$-\frac{g(2)}{s+2}.$$

Since the sum $\Gamma(s+1)g(s) + g(2)/(s+2)$ is analytic, we have

$$\int_{|t|\le(\log n)^{-2/5}} \left(\Gamma(s+1)g(s) + \frac{g(2)}{s+2}\right)(ne^{i\theta})^{-\rho-it}T(\rho+it)^k\,dt$$
$$= O\left(\frac{n^{-\rho}T(\rho)^k}{\sqrt{k}}\right)$$

and consequently

$$\begin{aligned}
\Delta_k(ne^{i\theta}) &= \frac{-g(-2)}{2\pi}\int_{|t|\le(\log n)^{-2/5}} \frac{(ne^{i\theta})^{-\rho-it}T(\rho+it)^k}{\rho+2+it}\,dt \\
&= \frac{-g(-2)}{2\pi}n^{-\rho}e^{-i\theta\rho}T(\rho)^k\int_{|t|\le(\log n)^{-2/5}} \frac{e^{\theta t-\beta(\rho)kt^2/2+O(k|t|^3)}}{\rho+2+it}\,dt \\
&= \frac{-g(-2)}{2\pi}n^{-\rho}e^{-i\theta\rho}T(\rho)^k\int_{-\infty}^{\infty} \frac{e^{-t^2/2}}{\xi_0+it}\left(1+O\left(\frac{|t|+|t|^3}{\sqrt{\log n}}\right)\right)dt,
\end{aligned}$$

where

$$\xi_0 = (\rho+2)\sqrt{\beta(\rho)k} = -\xi + O(\xi^2(\log n)^{-1/2}).$$

Since $\xi_0 > 0$, we obtain

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-t^2/2}}{\xi_0+it}\,dt &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-t^2/2}\int_0^{\infty} e^{-v(\xi_0+it)}\,dv\,dt \\
&= \frac{1}{2\pi}\int_0^{\infty} e^{-v\xi_0}\int_{-\infty}^{\infty} e^{-t^2/2-itv}\,dt\,dv \\
&= \frac{1}{\sqrt{2\pi}}e^{-v^2/2-v\xi_0}\,dv \\
&= e^{\xi_0^2/2}\Phi(-\xi_0).
\end{aligned}$$

The error term is estimated similarly:

$$\begin{aligned}
&\frac{1}{\sqrt{\log n}}\int_{-\infty}^{\infty}\frac{(|t|+|t|^3)e^{-t^2/2}}{|(\rho+2)\sqrt{\beta(\rho)k}+it|}\,dt \\
&\quad = O\left(\frac{1}{\sqrt{\log n}}\int_0^{\infty}(v+v^3)e^{-v^2/2-v\xi_0}\,dv\right) \\
&\quad = O\left(\frac{1}{\sqrt{\log n}}e^{\xi_0^2/2}\Phi(-\xi_0)(1+|\xi_0|^3)\right).
\end{aligned}$$

Thus we get

$$\Delta_k(ne^{i\theta}) = -g(-2)(ne^{i\theta})^{-\rho}T(\rho)^k e^{\xi_0^2/2}\Phi(-\xi_0)\left(1+O\left(\frac{1+|\xi_0|^3}{\sqrt{\log n}}\right)\right)$$
$$+ O\left(k^{-1/2}n^{-\rho}T(\rho)^k\right).$$

By using the local expansions

$$n^{-\rho}T(\rho)^k = n^2T(2)^k e^{-\xi^2/2+O(|\xi^3|(\log n)^{-1/2})},$$
$$e^{\xi_0^2/2}\Phi(-\xi_0) = e^{\xi^2/2}\Phi(\xi)\left(1+O\left(|\xi|^3(\log n)^{-1/2}\right)\right)$$

we end up with the expansion

$$\Delta_k(ne^{i\theta}) = -g(-2)(ne^{i\theta})^2T(2)^k\Phi(\xi)\left(1+O\left(\frac{1+|\xi_0|^3}{\sqrt{\log n}}\right)\right)$$
$$+O\left(k^{-1/2}n^2T(2)^k e^{-\xi^2/2}\right),$$

that holds uniformly for $|\theta| \le \theta_0$. By combining this expansion with Lemma 7.6 (together with analytic depoissonisation) we finally obtain the proposed asymptotic expansion for $\mathbb{E}\, B_{n,k}$, if $k < \alpha_2 \log n$. The missing case $k \ge \alpha_2 \log n$ can be treated in a similar way. Therefore we have also completed the proof in the second case of Theorem 7.1.

The proof of Theorem 7.2 is very similar to the proof of Theorem 7.1. The corresponding Poisson transform $\overline{\Delta}_k(x)$ is given by

$$\overline{\Delta}_k(x) = 2^k - \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty}(s+1)\Gamma(s)T(s)^k x^{-s}\,ds,$$

where $\rho > 0$. Due to the (additional) polar singularity at $\rho = 0$ there is a second phase transition. However, the analysis uses exactly the same methods.

7.1.3 The Limiting Distribution of the Profile of Tries

We next consider the limiting distribution of the profile that is Gaussian in the range of interest. We first observe that in the range $\alpha_1+\epsilon \le \frac{k}{\log n} \le \alpha_2-\epsilon$ the variances $\mathbb{V}\mathrm{ar}\, B_{n,k}$ and $\mathbb{V}\mathrm{ar}\, I_{n,k}$ have asymptotic representations of the form

$$\mathbb{V}\mathrm{ar}\, B_{n,k} = H\left(\rho_{n,k}, \log_{p/q} p^k n\right)\frac{(p^{-\rho_{n,k}}+q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}}\left(1+O\left(k^{-1/2}\right)\right)$$

and

$$\mathbb{V}\mathrm{ar}\, I_{n,k} = \overline{H}\left(\rho_{n,k}, \log_{p/q} p^k n\right)\frac{(p^{-\rho_{n,k}}+q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}}\left(1+O\left(k^{-1/2}\right)\right)$$

where $H(\rho,x)$ and $\overline{H}(\rho,x)$ are non-zero periodic functions with period 1 (we do not give proofs here, for details see [169]). In particular, $\mathbb{V}\mathrm{ar}\, B_{n,k} \to \infty$ and $\mathbb{V}\mathrm{ar}\, I_{n,k} \to \infty$ in this range. Remarkably, there is no phase transition for $\mathbb{V}\mathrm{ar}\, B_{n,k} \to \infty$ for $k \sim \alpha_0 \log n$ which means that for $\alpha_1 < \frac{k}{\log n} < \alpha_0$ the expected value and the variance of the internal profile do not have the same order of magnitude. Nevertheless, there is a central limit theorem even in this range.

Theorem 7.7. *Suppose that* $\alpha_1 + \epsilon \le \frac{k}{\log n} \le \alpha_2 - \epsilon$ *(for some* $\epsilon > 0$*). Then*

$$\frac{B_{n,k} - \mathbb{E}\, B_{n,k}}{\sqrt{\mathbb{V}\mathrm{ar}\, B_{n,k}}} \xrightarrow{d} N(0,1) \quad \text{and} \quad \frac{I_{n,k} - \mathbb{E}\, I_{n,k}}{\sqrt{\mathbb{V}\mathrm{ar}\, I_{n,k}}} \xrightarrow{d} N(0,1).$$

We present an outline of the proof of Theorem 7.7 (for details see [169]). It is based on an analysis of the generating function $G_k(x,u)$ resp. $G_k^{[I]}(x,u)$. Recall that $G_k(x,u)$ satisfies the recurrence $G_k(x,u) = G_{k-1}(px,u)G_{k-1}(qx,u)$ (for $k \ge 2$). The principle idea of the proof is to introduce the functions

$$\Delta_k(x,u) = \log(e^{-x} G_k(x,u))$$

and the Mellin transforms

$$\Delta_k^*(s,u) = \int_0^\infty \Delta_k(x,u) x^{s-1}\, dx.$$

Since $G_1(x,u) = e^x + (e^{px}qx + e^{qx}px - 1)(u-1) + pqx^2(u-1)^2$, it follows that the Mellin transform $\Delta_1^*(s,u)$ exists for $\Re(s) > -2$. Thus,

$$\Delta_k^*(s,u) = \Delta_1^*(s,u) T(s)^{k-1}$$

also exists for s in this range. Consequently, we have

$$\log(e^{-x} G_k(x,u)) = \Delta_k(x,u) = \frac{1}{2\pi i} \int_{\rho - i\infty}^{\rho + i\infty} \Delta_1^*(s,u) T(s)^{k-1} x^{-s}\, ds, \quad (7.10)$$

where we can choose ρ as the saddle point $\rho_{n,k}$ for x with modulus $|x| = n$.

Thus, the plan of the proof is to use a saddle point method to obtain asymptotics for $\Delta_k(x,u)$ and consequently for $G_k(x,u)$. Second, one has to depoissonise $G_k(x,u)$ to obtain asymptotics for $\mathbb{E}\, u^{B_{n,k}}$ (for $|u| = 1$) which will lead to the central limit theorem.

We collect some properties that are used for the proof (see [169]).

Lemma 7.8. *There exists a constant* $c > 0$ *such that*

$$|G_k(re^{i\theta}, u)| \le e^{r - cr\theta^2}$$

uniformly for $r \ge 0$, $|\theta| \le \pi$, *and* $|u| = 1$.

The proof is similar to the proof of Lemma 7.6.

Lemma 7.9. *Let* $V_k(x) = \Delta_k^{(2)}(x) - \Delta_k(x)^2$, *where*

$$\Delta_k^{(2)}(x) = e^{-x} \sum_{n \ge 0} \mathbb{E}(B_{n,k}^2) \frac{x^n}{n!}.$$

Suppose that $\alpha_1 + \epsilon \le \frac{k}{\log n} \le \alpha_2 - \epsilon$ *(for some* $\epsilon > 0$*) and set* $\sigma_{n,k} = \sqrt{\mathbb{V}\mathrm{ar}\, B_{n,k}}$ *and* $\rho_0 = \max\{1, \rho_{n,k}\}$. *Then we have*

$$G_k(ne^{i\theta}, e^{i\varphi}) = \exp\Big(n - \frac{n}{2}\theta^2 + \Delta_k(n)i\varphi - \Delta_k'(n)\varphi\theta - \frac{V_k(n)}{2}\varphi^2 + O\left(n|\theta|^3 + \rho_0^2\sigma_{n,k}^2|\varphi|\theta^2 + \rho_0\sigma_{n,k}^2\varphi^2|\theta| + \sigma_{n,k}^2|\varphi|^3\right)\Big)$$

uniformly for $|\theta| \le n^{2/5}$ *and* $|\varphi| = o(\sigma_{n,k}^{-2/3})$.

The proof is based on the explicit representation (7.10), local expansions, proper estimates for derivatives of $\Delta_k(x)$ and $\Delta_k^{(2)}(x)$, and a saddle point analysis (see [169]).

By using these two properties one obtains (after working out the asymptotics of the integrals)

$$\begin{aligned}\mathbb{E}\, e^{iB_{n,k}\varphi} &= \frac{n!}{2\pi i}\int_{|x|=n} G_k(x, e^{i\varphi})\frac{dx}{x^{n+1}} \\ &= e^{i\Delta_k(n)\varphi - \frac{1}{2}\varphi^2(V_k(n) - n\Delta_k'(n)^2)}\left(1 + O(\sigma_{n,k}^{-1} + \sigma_{n,k}^2|\varphi|^3)\right)\end{aligned}$$

and, thus

$$\mathbb{E}\exp\left(\frac{B_{n,k} - \Delta_k(n)}{\sqrt{V_k(n) - n\Delta_k'(n)^2}}i\varphi\right) = e^{-\varphi^2/2}\left(1 + O\left(\frac{1+|\varphi|^3}{\sigma_{n,k}}\right)\right),$$

uniformly for $\varphi = o(\sigma_{n,k}^{1/5})$. Since $\mathbb{E}\, B_{n,k} \sim \Delta_k(n)$ and $\mathbb{V}\mathrm{ar}\, B_{n,k} \sim V_k(n) - n\Delta_k'(n)^2$ (see [169]), the result for the external profile follows. The proof of the central limit theorem for the internal profile is very similar to that of the external profile. For details we again refer to [169].

7.1.4 The Height of Tries

The height H_n of tries was already studied by Flajolet and Steyaert [85], Devroye [48], and Pittel [173], and Szpankowski [196]. The limiting behaviour was finally determined by Pittel [174] (even in a slightly more general setting).

Theorem 7.10. *The distribution of the height H_n of random tries is asymptotically given by*

$$\mathbb{P}\{H_n \le k\} = \exp\left(-\frac{1}{2}e^{-2(\alpha_3 k - \log n)}\right) + o(1)$$

uniformly for all $k \ge 0$ as $n \to \infty$ (where $\alpha_3 = 2/\log(1/(p^2+q^2))$).

The analytic methods that have been presented to derive asymptotic properties of the profile can also be used to handle the height. Let $G_k(x)$ denote the generating function

$$G_k(x) = \sum_{n\ge 0}\mathbb{P}\{H_n \le k\}\frac{x^n}{n!}.$$

Then we have (as for the profile)

$$G_k(x) = G_{k-1}(px)G_{k-1}(qx), \qquad (k \ge 2),$$

and can we proceed as in Section 7.1.3.

7.1.5 Symmetric Tries

The case of symmetric binary tries with underlying probabilities $p = q = \frac{1}{2}$ is very easy to analyse. In this case the generating functions $G_k(x,u)$ and $G_k^{[I]}(x,u)$ are explictly given by

$$G_k(x,u) = \left(e^{x/2^{k-1}} + (u-1)\frac{x}{2^{k-1}}\left(e^{x/2^k} - 1\right) + (u-1)^2\frac{x^2}{4}\right)^{2^{k-1}},$$

and by

$$G_k^{[I]}(x,u) = \left(ue^{x/2^k} + (1-u)\left(1 + \frac{x}{2^k}\right)\right)^{2^k}.$$

Hence, we also get explicit results.

Theorem 7.11. *The expected values $\mathbb{E}\, B_{n,k}$ and $\mathbb{E}\, I_{n,k}$ of the external and internal profile of symmetric binary tries are explicitly given by*

$$\mathbb{E}\, B_{n,k} = n\left(1 - 2^{-k}\right)^{n-1} - n\left(1 - 2^{1-k}\right)^{n-1},$$
$$\mathbb{E}\, I_{n,k} = 2^k\left(1 - \left(1 - 2^{-k}\right)^n\right) - n\left(1 - 2^{-k}\right)^{n-1},$$

and asymptotically by

$$\mathbb{E}\, B_{n,k} \sim \begin{cases} n\left(1 - 2^{-k}\right)^{n-1} & \text{if } 2^{-k}n \to \infty, \\ n\left(e^{-n/2^k} - e^{-n/2^{k-1}}\right) & \text{if } 4^{-k}n \to 0, \end{cases}$$

$$\mathbb{E}\, I_{n,k} \sim \begin{cases} 2^k - n\left(1 - 2^{-k}\right)^{n-1} & \text{if } 2^{-k}n \to \infty, \\ 2^k - (2^k + n)e^{-n/2^k} & \text{if } 4^{-k}n \to 0. \end{cases}$$

Furthermore, if

$$\frac{\log n - \log\log n}{\log 2} \le k \le \frac{2\log n}{\log 2},$$

then we have

$$\frac{B_{n,k} - \mathbb{E}\, B_{n,k}}{\sqrt{\mathbb{V}\mathrm{ar}\, B_{n,k}}} \xrightarrow{d} N(0,1),$$
$$\frac{I_{n,k} - \mathbb{E}\, I_{n,k}}{\sqrt{\mathbb{V}\mathrm{ar}\, I_{n,k}}} \xrightarrow{d} N(0,1),$$

where the variances are asymptotically given by

$$\mathbb{V}\mathrm{ar}\, B_{n,k} \sim \mathbb{V}\mathrm{ar}\, I_{n,k} \sim n\left(1-2^{-k}\right)^{n-1}, \qquad \text{if } 2^{-k}n \to \infty,$$

and by

$$\mathbb{V}\mathrm{ar}\, B_{n,k} \sim n\left(e^{-n/2^k} - e^{-n/2^{k-1}}\right) + 2^{-k}n^2 e^{-n/2^{k-1}} - 2^{1-k}n^2\left(e^{-n/2^k} - e^{-n/2^{k-1}}\right)^2$$

and

$$\mathbb{V}\mathrm{ar}\, I_{n,k} \sim (2^k+n)e^{-n/2^k} - 2^k\left(1+2^{-k}\right)^2 e^{-n/2^{k-1}},$$

if $4^{-k}n \to 0$.

Accordingly, the generating function for the height distribution is given by

$$G_k(x) = \left(1+\frac{x}{2^k}\right)^{2^k},$$

which leads to the explicit height distribution

$$\mathbb{P}\{H_n \le k\} = \frac{n!}{2^{nk}}\binom{2^k}{n}.$$

7.2 The Profile of Digital Search Trees

Although digital search trees are very well studied there are only partial results on the profile (in particular in the symmetric case $p = q = \frac{1}{2}$ [126], [130, Section 6.3]; we will come back to this special case in Section 7.2.3). The closest related quantity is the typical depth D_n that measures the path length from the root to a randomly selected node; it is equal to the ratio of the average profile to the number of nodes. Unfortunately, all estimations of the depth [130, 142, 144, 177, 195] only deal with the typical depth around the most likely value, namely $k = (1/h)\log n + O(1)$, where $h = -p\log p - q\log q$ is the entropy rate.

7.2.1 Generating Functions for the Profile

Let $B_{n,k}$ resp. $I_{n,k}$ denote the (random) number of external resp. internal nodes at level k in a digital search tree, when the total number of internal nodes equals n, that is, we have the empty string and $n-1$ strings that are generated by a memoryless source with parameters $p > q = 1-p$. By the definition of the probability generating function, $P_{n,k}(u) = \mathbb{E}\, u^{B_{n,k}}$, satisfies the following recurrence relation

$$P_{n+1,k}(u) = \sum_{\ell=0}^{n} \binom{n}{\ell} p^\ell q^{n-\ell} P_{n,k-1}(u) P_{n,k-1}(u), \qquad (7.11)$$

while the corresponding exponential generating function

$$G_k(x,u) = \sum_{n\ge 0} P_{n,k}(u) \frac{x^n}{n!}$$

satisfies the functional recurrence

$$\frac{\partial}{\partial x} G_k(x,u) = G_{k-1}(px,u) G_{k-1}(qx,u), \qquad (k \ge 1), \qquad (7.12)$$

with initial conditions $G_0(x,u) = u + e^x - 1$ and $G_k(0,u) = 1$ $(k \ge 1)$.

Accordingly, the corresponding generating functions for the internal profile

$$G_k^{[I]}(x,u) = \sum_{n\ge 0} \mathbb{E}\, u^{I_{n,k}} \frac{x^n}{n!}$$

satisfy the same recurrence relation

$$\frac{\partial}{\partial x} G_k^{[I]}(x,u) = G_{k-1}^{[I]}(px,u) G_{k-1}^{[I]}(qx,u), \qquad (k \ge 1), \qquad (7.13)$$

with initial conditions $G_0^{[I]}(x,u) = 1 + u(e^x - 1)$ and $G_k^{[I]}(0,u) = 1$ $(k \ge 1)$.

We are interested in the expected profiles $\mathbb{E}\, B_{n,k}$ and $\mathbb{E}\, I_{n,k}$. By taking derivatives with respect to u and setting $u = 1$ we obtain for the exponential generating function

$$E_k(x) = \sum_{n\ge 0} \mathbb{E}\, B_{n,k} \frac{x^n}{n!},$$

that satisfies the functional recurrence

$$E_k'(x) = e^{qx} E_{k-1}(px) + e^{px} E_{k-1}(qx), \qquad (7.14)$$

with initial condition $E_0(x) = 1$ and $E_k(0) = 0$ $(k \ge 1)$.

The corresponding generating functions for the internal profile

$$E_k^{[I]}(x) = \sum_{n\ge 0} \mathbb{E}\, I_{n,k} \frac{x^n}{n!}$$

satisfy the recurrence (7.14), too, however with initial conditions $E_0^{[I]}(x) = e^x - 1$ and $E_k^{[I]}(0) = 0$ $(k \ge 1)$.

7.2.2 The Expected Profile of Digital Search Trees

In order to state the results for the expected profile of digital search trees we use the same notation as for tries. For a real number α with $(\log \frac{1}{p})^{-1} < \alpha < (\log \frac{1}{q})^{-1}$ we define $\rho = \rho(\alpha)$ by (7.6) and $\beta(\rho)$ by (7.7). We also use the abbreviation

$$\alpha_0 = \frac{2}{\log \frac{1}{p} + \log \frac{1}{q}}.$$

The results for the expected profile of digital search trees are formally similar to those of tries (compare with Theorems 7.1 and 7.2).

Theorem 7.12. *Let $\mathbb{E}\, B_{n,k}$ denote the expected external profile in (asymmetric) digital search trees with underlying (non-zero) probabilities $p > q = 1-p$. If n and k are positive integers with $\frac{1}{\log \frac{1}{p}} + \epsilon \le \frac{k}{\log n} \le \frac{1}{\log \frac{1}{q}} - \epsilon$ (for some $\epsilon > 0$) then we uniformly have*

$$\mathbb{E}\, B_{n,k} = G\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(k^{-1/2}\right)\right),$$

where $G(\rho, x)$ is a non-zero periodic function with period 1 and $\rho_{n,k} = \rho(k/\log n)$.

Theorem 7.13. *Let $\mathbb{E}\, I_{n,k}$ denote the expected internal profile in (asymmetric) digital search trees with underlying (non-zero) probabilities $p > q = 1-p$. Let k and n be positive integers such that $k/\log n$ satisfies $(\log \frac{1}{p})^{-1} < k/\log n < (\log \frac{1}{q})^{-1}$. Then the following assertions hold:*

1. *If $\frac{1}{\log \frac{1}{p}} + \epsilon \le \frac{k}{\log n} \le \alpha_0 - \epsilon$ (for some $\epsilon > 0$) then we have uniformly*

$$\mathbb{E}\, I_{n,k} = 2^k - \overline{G}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(k^{-1/2}\right)\right),$$

where $\overline{G}(\rho, x)$ is a non-zero periodic function with period 1.

2. *If $k = \alpha_0 \left(\log n + \xi\sqrt{\alpha_0\beta(0)\log n}\right)$, where $\xi = o((\log n)^{\frac{1}{6}})$, then*

$$\mathbb{E}\, I_{n,k} = 2^k \Phi(-\xi) \left(1 + O\left(\frac{1 + |\xi|^3}{\sqrt{\log n}}\right)\right).$$

3. *If $\alpha_0 + \epsilon \le \frac{k}{\log n} \le \frac{1}{\log \frac{1}{q}} - \epsilon$ (for some $\epsilon > 0$) then we have uniformly*

$$\mathbb{E}\, I_{n,k} = \overline{G}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + O\left(k^{-1/2}\right)\right)$$

with the same function $\overline{G}(\rho, x)$ as in 1.

A direct consequence of these properties is a local limit theorem for the depth D_n of a random internal node. The result is precisely the same as Theorem 7.4 (only the periodic function is different). As in the case of tries the depth satisfies a central limit theorem (see [112]) but not a local central limit theorem.

The proofs of Theorems 7.12 and 7.13 are also similar to the proofs of the corresponding theorems (Theorems 7.1 and 7.2) for the expected profile of tries. However, we have to add one further step, since the Mellin transforms $\Delta_k^*(s)$ cannot be written in a direct explicit form.

As for tries the first step is to consider the Poisson transform

$$\Delta_k(x) = e^{-x} \sum_{n\ge 0} \mathbb{E}\, B_{n,k} \frac{x^n}{n!} = E_k(x)e^{-x}, \qquad (k \ge 0).$$

It is clear that the above recurrence translates to

$$\Delta_k(x) + \Delta_k'(x) = \Delta_{k-1}(px) + \Delta_{k-1}(qx), \qquad (k \ge 1), \tag{7.15}$$

with initial conditions $\Delta_0(x) = e^{-x}$ and $\Delta_k(0) = 0$ $(k \ge 1)$.

The second step is to use the Mellin transform

$$\Delta_k^*(s) = \int_0^\infty \Delta_k(x) x^{s-1}\, dx.$$

By induction it is easy to prove that $\Delta_k(x)$ can be represented as a finite linear combination of functions of the form $e^{-p^{\ell_1}q^{\ell_2}x}$ with $\ell_1, \ell_2 \ge 0$ and $0 \le \ell_1 + \ell_2 \le k$. Hence, $\Delta_k^*(s)$ surely exists for all s with $\Re(s) > 0$. Furthermore, $B_{n,k} = 0$ for $k > n$. Thus, $E_k(x) = O(x^k)$ for $x \to 0$ which ensures that $\Delta_k^*(s)$ actually exists for s with $\Re(s) > -k$. We can also write

$$\Delta_k^*(s) = \Gamma(s) F_k(s),$$

where $F_k(s)$ is now a finite linear combination of functions of the form $p^{-\ell_1 s} q^{-\ell_2 s}$ with $\ell_1, \ell_2 \ge 0$ and $0 \le \ell_1 + \ell_2 \le k$. Thus, $F_k(s)$ can be considered as an entire function. The relation (7.15) translates to

$$F_k(s) - F_k(s-1) = (p^{-s} + q^{-s}) F_{k-1}(s) = T(s) F_{k-1}(s), \qquad (k \ge 1), \tag{7.16}$$

with initial condition $F_0(s) = 1$. Note that the relation (7.16) not only holds for $\Re(s) > k$ where the Mellin transform exists. Since $F_k(s)$ analytically continues to an entire function, (7.16) holds for all s.

At this formal level the essential difference between tries and digital search trees is that the recurrence (7.16) cannot be solved in a direct way. Nevertheless the asymptotic behaviour is of the same type. We have to add a third step to our analysis. We consider the power series

$$f(w,s) = \sum_{k\ge 0} F_k(s) w^s.$$

It turns out that $f(w,s)$ can we represented (see Lemma 7.14) as

$$f(w,s) = \frac{g(w,s)}{g(w,-1)},$$

where $g(w,s)$ satisfies the relation

$$g(w,s) = 1 + w \sum_{j\ge 0} g(w,s-j)(p^{-s+j} + q^{-s+j}). \tag{7.17}$$

Granted the above, an asymptotic analysis follows. We start with a singularity analysis of $g(w,s)$, in particular we will show (see Lemma 7.17) that $g(w,s)$ (usually) has a polar singularity at $w = 1/(p^{-s}+q^{-s})$. Thus it will be possible to get proper asymptotics for $F_k(s)$. In fact we get $F_k(s) \sim f(s)(p^{-s} + q^{-s})^k$ (for a proper function $f(x)$ and for s in the interesting range). Actually this resembles the exact expression for tries. This is also the reason why the overall behaviour of the profile of biased tries and biased digital search trees is almost of the same form. Only the periodic functions are slightly different. The final two steps (inverting the Mellin transform and depoissonisation) are almost identical to the methods presented in Section 7.1.2.

Before we study the generating function $f(w,s) = \sum_{k\ge 0} F_k(s)w^k$, we will collect some basic properties of $F_k(s)$. We recall that $F_k(s)$ can be considered as an entire function.

Let $\mathbf{A}$ be an functional operator that is defined by

$$\mathbf{A}[f](s) = \sum_{j\ge 0} f(s-j)T(s-j), \tag{7.18}$$

where

$$T(s) = p^{-s} + q^{-s}. \tag{7.19}$$

In the next lemma we find an explicit representation of $F_k(x)$ through the operator $\mathbf{A}$.

Lemma 7.14. *The functions $F_k(s)$ are recursively given by*

$$F_k(s) = \mathbf{A}[F_{k-1}](s) - \mathbf{A}[F_{k-1}](0) \qquad (k \ge 1) \tag{7.20}$$

with initial function $F_0(s) = 1$. Furthermore, if we set $R_k(s) = \mathbf{A}^k[1](s)$ then we have the formal identity

$$\sum_{k\ge 0} F_k(s)w^k = \frac{\sum_{\ell\ge 0} R_\ell(s)w^\ell}{\sum_{\ell\ge 0} R_\ell(0)w^\ell}. \tag{7.21}$$

Finally for $k \ge 1$ we have $F_k(-\ell) = 0$ for $\ell = 0,1,2,\ldots,k-1$.

Proof. Set $\tilde{F}_k(s) = 1$ and recursively

$$\tilde{F}_k(s) = \mathbf{A}[\tilde{F}_{k-1}](s) - \mathbf{A}[\tilde{F}_{k-1}](0) \qquad (k \ge 1).$$

It is easy to see that $\tilde{F}_k(s)$ is a well defined entire function. In particular it follows that $\tilde{F}_k(s)$ is (as it is for $F_k(s)$) a finite linear combination of a function of the form $p^{-\ell_1 s}q^{-\ell_2 s}$ with $\ell_1, \ell_2 \ge 0$ and $\ell_1 + \ell_2 \le k$. Further (by definition) these functions satisfy $\tilde{F}_k(0) = 0$ (for $k \ge 1$) and fulfil the relation

$$\tilde{F}_k(s) - \tilde{F}_k(s-1) = T(s)\tilde{F}_{k-1}(s)$$

for $k \ge 0$ and all s.

Now we can proceed by induction to show that $F_k(s) = \tilde{F}_k(s)$. By definition we have $F_0(s) = \tilde{F}_0(s)$. Now suppose that $F_k(s) = \tilde{F}_k(s)$ holds for some $k \ge 0$. Then with the help of the above considerations it follows that $F_{k+1}(s) = \tilde{F}_{k+1}(s) + G(s)$, where $G(s)$ satisfies

$$G(0) = 0 \quad \text{and} \quad G(s) - G(s-1) = 0, \qquad (\Re(s) > -k). \tag{7.22}$$

By the above observations $G(s)$ has to be a finite linear combination of functions of the form $p^{-\ell_1 s}q^{-\ell_2 s}$. However, the only periodic function of this form that meets conditions (7.22) is the zero function. Hence, $F_{k+1}(s) = \tilde{F}_{k+1}(s)$.

Now we prove (7.21). First, (7.21) is equivalent to

$$\sum_{\ell=0}^{k} F_\ell(s) R_{k-\ell}(0) = R_k(s), \qquad (k \ge 0),$$

or to

$$F_k(s) = R_k(s) - \sum_{\ell=0}^{k-1} F_\ell(s) R_{k-\ell}(0), \qquad (k \ge 0).$$

We will prove this relation by induction. Of course, it is satisfied for $k = 0$. Now suppose that is holds for some $k \ge 0$. Then from (7.20) we find

$$\begin{aligned}
F_{k+1}(s) &= \mathbf{A}[F_k](s) - \mathbf{A}[F_k](0) \\
&= \mathbf{A}[R_k](s) - \mathbf{A}[R_k](0) \\
&\quad - \sum_{\ell=0}^{k-1} (\mathbf{A}[F_\ell](s) - \mathbf{A}[F_\ell](0)) R_{k-\ell}(0) \\
&= R_{k+1}(s) - R_{k+1}(0) - \sum_{\ell=0}^{k-1} F_{\ell+1}(s) R_{k-\ell}(0) \\
&= R_{k+1}(s) - \sum_{\ell=0}^{k} F_\ell(s) R_{k+1-\ell}(0).
\end{aligned}$$

This completes the induction proof.

Finally, since $F_k(s) = -\Delta_k^*(s)/\Gamma(s)$ is analytic for s with $\Re(s) > -k$ and $1/\Gamma(-\ell) = 0$, it also follows that $F_k(-\ell) = 0$ for $\ell = 0, 1, \ldots, k-1$.

Remark 7.15 *The proof of (7.20) (and consequently that of (7.21)) makes use of the fact that $F_k(0) = 0$ for $k \ge 1$. However, we also have $F_k(-r) = 0$ for $k > r$. In particular, if we set $s = -r$ in (7.21) we get*

$$\sum_{k=0}^{r} F_k(-r)w^k = \frac{\sum_{\ell \ge 0} R_k(-r)w^k}{\sum_{\ell \ge 0} R_k(0)w^k},$$

and consequently

$$\sum_{k \ge 0} F_k(s)w^k = \frac{\sum_{\ell \ge 0} R_k(s)w^k}{\sum_{\ell \ge 0} R_k(-r)w^k} \sum_{k=0}^{r} F_k(-r)w^r. \tag{7.23}$$

Our next goal is to study the function $g(w,s) = \sum_{\ell \ge 0} R_\ell(s)w^\ell$, where we now consider w as a complex variable, too. Note that $g(w,s)$ satisfies the (at the moment formal) identity

$$g(w,s) = 1 + w\mathbf{A}[g(w,\cdot)](s) = 1 + \sum_{j \ge 0} g(w, s-j)T(s-j). \tag{7.24}$$

In the next lemma we establish a crucial property of $g(w,s)$.

Lemma 7.16. *There exists a function $h(w,s)$ that is analytic for all w and s with*

$$wT(s-m) \ne 1 \quad \textit{for all } m \ge 1$$

such that

$$g(w,s) = \frac{h(w,s)}{1 - wT(s)}. \tag{7.25}$$

Thus, $g(w,s)$ has a meromorphic continuation where $w_0 = 1/T(s)$ is a polar singularity.

Proof. We recall that $R_k(s) = \mathbf{A}^k[1](s)$. In particular, the first few functions $R_k(s)$ are given by

$$\begin{aligned}
R_0(s) &= 1, \\
R_1(s) &= \frac{p^{-s}}{1-p} + \frac{q^{-s}}{1-q}, \\
R_2(s) &= \frac{p^{-2s}}{(1-p)(1-p^2)} + \frac{p^{-s}q^{-s}}{(1-p)(1-pq)} \\
&\quad + \frac{p^{-s}q^{-s}}{(1-q)(1-pq)} + \frac{q^{-2s}}{(1-q)(1-q^2)}.
\end{aligned}$$

With the help of (7.18) we can derive corresponding representations for general k. Hence, by using the assumption $p > q$ it follows that

$$|R_k(s)| \le \frac{1}{\prod_{j \ge 1}(1-p^j)} (p^{-\Re(s)} + q^{-\Re(s)})^k.$$

Thus, if $|w| < T(\Re(s))^{-1}$ then the series

$$g(w,s) = \sum_{\ell \ge 0} R_\ell(s) w^\ell = \left(\sum_{\ell \ge 0} w^\ell \mathbf{A}^\ell \right) [1](s) \tag{7.26}$$

converges absolutely and represents an analytic function. We can rewrite (7.26) to

$$g(w,s) = (\mathbf{I} - w\mathbf{A})^{-1}[1](s),$$

or to

$$(\mathbf{I} - w\mathbf{A})[g(w,\cdot)](s) = g(w,s) - w \sum_{j \ge 0} g(w, s-j) T(s-j) = 1, \tag{7.27}$$

which is the same as (7.24).

If we substitute $g(w,s)$ in (7.27) by

$$g(w,s) = \frac{h(w,s)}{1 - wT(s)},$$

we get a relation for $h(w,s)$ of the form

$$h(w,s) = 1 + \sum_{j \ge 1} h(w, s-j) \frac{wT(s-j)}{1 - wT(s-j)}. \tag{7.28}$$

Recall that we already know that $h(w,s)$ exists for $|w| < T(\Re(s))^{-1}$. We will now use (7.28) to show that $h(w,s)$ can be analytically continued to the range $|w| < T(\Re(s) - 1)^{-1}$ (and even to the range where $wT(s-m) \ne 1$) so that we also get a meromorphic continuation as proposed.

For this purpose we introduce another operator $\mathbf{B}$ by

$$\mathbf{B}[f](s) = \sum_{j \ge 1} f(w, s-j) \frac{wT(s-j)}{1 - wT(s-j)}. \tag{7.29}$$

For convenience set $U(w,s) = wT(s)/(1 - wT(s))$. By induction it follows that

$$\begin{aligned}
\mathbf{B}^k[1](s) &= \sum_{i_1 \ge 1} \sum_{i_2 \ge 1} \cdots \sum_{i_k \ge 1} U(w, s-i_1) U(w, s-i_1-i_2) \\
&\qquad \cdots U(w, s-i_1-i_2-\cdots-i_k) \\
&= \sum_{m_k \ge k} \sum_{m_{k-1}=k-1}^{m_k - 1} \sum_{m_{k-2}=k-2}^{m_{k-1}-1} \\
&\qquad \cdots \sum_{m_1=1}^{m_2-1} U(w, s-m_1) U(w, s-m_2) \cdots U(w, s-m_k).
\end{aligned}$$

Hence, we get the upper bound

$$|\mathbf{B}^k[1](s)| \le \sum_{m_k \ge k} \sum_{m_{k-1} \ge k-1} \cdots \sum_{m_1 \ge 1} |U(w, s-m_1)U(w, s-m_2)\cdots U(w, s-m_k)|$$
$$= \sum_{m_1 \ge 1} |U(w, s-m_1)| \cdot \sum_{m_2 \ge 2} |U(w, s-m_2)| \cdots \sum_{m_k \ge k} |U(w, s-m_k)|.$$

By using the fact that $T(s-m) = O(q^m)$ it follows directly that the series

$$S := \sum_{m \ge 1} |U(w, s-m)| = \sum_{m \ge 1} \frac{|wT(s-m)|}{|1 - wT(s-m)|}$$

converges if $wT(s-m) \ne 1$ for all $m \ge 1$. Thus for any choice of w and s there are only finitely many exceptional points where $wT(s-m) = 1$. Let k_0 be any value with

$$\sum_{m \ge k_0} |U(w, s-m)| \le \frac{1}{2}.$$

Then we have for all $k \ge k_0$

$$|\mathbf{B}^k[1](s)| \le S^{k_0} 2^{-(k-k_0)} = (2S)^{k_0} 2^{-k}.$$

Hence, we can set

$$h(w, s) = \sum_{k \ge 0} \mathbf{B}^k[1](s), \tag{7.30}$$

which obviously satisfies (7.28). Furthermore, we have the upper bound $|h(w,s)| \le 2(2S)^{k_0}$.

Finally, we are in the position to derive an asymptotic representation for $F_k(s)$.

Lemma 7.17. *For every real interval $[a, b]$ there exist k_0, $\eta > 0$ and $\epsilon > 0$ such that*

$$F_k(s) = f(s)T(s)^k \left(1 + O\left(e^{-\eta k}\right)\right) \tag{7.31}$$

uniformly for all s with $\Re(s) \in [a, b]$, $|\Im(s) - 2\ell\pi \log(q/p)| \le \epsilon$ for some integer ℓ and $k \ge k_0$, where $f(s)$ is an analytic function that satisfies $f(-r) = 0$ for $r = 0, 1, 2, \ldots$.

Furthermore, if $|\Im(s) - 2\ell\pi \log(q/p)| > \epsilon$ for for all integers ℓ then we have

$$F_k(s) = O\left(T(\sigma)^k \, e^{-\eta k}\right) \tag{7.32}$$

uniformly for $\Re(s) \in [a, b]$.

Proof. Suppose first that s is a real number with $-r-1 < s < -r$ for some integer $r \ge 0$. Here we use the representation

$$f(w,s) = \sum_{k=0}^{r+1} F_k(-r-1) w^k \frac{g(w,s)}{g(-r-1,w)}$$
$$= \sum_{k=0}^{r+1} F_k(-r-1) w^k \frac{h(w,s)}{h(-r-1,w)} \frac{1-wT(-r-1)}{1-wT(s)}.$$

By Lemma 7.16 there exists $\eta > 0$ such that $h(w,s)$ is analytic for $|w| \le e^{\eta}/T(s)$. Since $T(-r-1) < T(s)$, it also follows that $h(w,-r-1)$ is analytic in that region. Furthermore, since $h(w,-r-1)$ is non-zero for positive real $w < 1/T(-r-2)$ (compare with (7.30)), we obtain that the radius of convergence of the series $\sum_{k\ge 0} F_k(s) w^k$ equals $w_0 = 1/T(s)$.

With the help of this observation we can also deduce that the function $f(w,s)$ has no other singularities on the circle $|w| = 1/T(s)$. Suppose that $h(w,-r-1)$ has a zero w_1 with $|w_1| < 1/T(s)$. If $\sum_{k=0}^{r+1} F_k(-r-1) w_1^k \ne 0$ then w_1 has to be a zero of $h(w,s)$, too: $h(w_1,s) = 0$. However, if we slightly decrease s then certainly $h(w_1, s-\eta) \ne 0$. In this case the function $f(w,s)$ would be singular for $w = w_1$ although its radius of convergence is $1/T(s-\eta) > 1/T(s) > |w_1|$. This is, of course, a contradiction and, thus, $\sum_{k=0}^{r+1} F_k(-r-1) w_1^k = 0$ too. Actually, it also follows that the order of the zeros are the same. Furthermore, by a slight variation of the above argument, we also deduce that $f(w,s)$ has no singularities on the circle $|w| = 1/T(s)$ other than $w_0 = 1/T(s)$, as proposed.

Hence, by using a contour integration on the circle $|w| = e^{\eta}/T(s)$ and the residue theorem it follows that

$$F_k(s) = f(s) T(s)^k + O\left(|T(s)e^{-\eta}|^k\right),$$

where

$$f(s) = \sum_{k=0}^{r+1} F_k(-r-1) T(s)^{-k} \frac{h(1/T(s)), s}{h(1/T(s), -r-1)} \left(1 - \frac{T(-r-1)}{T(s)}\right).$$

These estimates are uniform for $s \in [a,b]$, where $-r-1 < a < b < r$. Furthermore, we get the same result if s is sufficiently close to the real axis. Thus, if $a \le \Re(s) \le b$ and $|\Im(s)| \le \epsilon$ for some sufficiently small $\epsilon > 0$ then we obtain (7.31), too.

Next, suppose that s is real (or sufficiently close to the real axis) and close to an integer $-r \le 0$, say $-r-\eta \le s \le -r+\eta$ (for some $\eta > 0$). Here we use the representation

$$\sum_{k\ge 0} F_k(s) w^k = \sum_{k=0}^{r} F_k(-r) w^k \frac{g(w,s)}{g(w,-r)}$$

$$= \sum_{k=0}^{r} F_k(-r) w^k \frac{h(w,s)}{h(w,-r)} \frac{1 - wT(-r)}{1 - wT(s)}$$

$$= \sum_{k=0}^{r} F_k(-r) w^k \frac{h(w,s) - h(w,-r)}{h(w,-r)} \frac{1 - wT(-r)}{1 - wT(s)}$$

$$+ \sum_{k=0}^{r} F_k(-r) w^k + \sum_{k=0}^{r-1} F_k(-r) w^{k+1} \frac{T(s) - T(-r)}{1 - wT(s)}.$$

Now if we subtract the finite sum $\sum_{k=0}^{r} F_k(-r)$ then we can safely multiply by $\Gamma(s)$ (that is singular at $s = -r$) and obtain a function of the form

$$\sum_{k=0}^{r} F_k(-r) w^k \frac{\Gamma(s)(h(w,s) - h(w,-r))}{h(w,-r)} \frac{1 - wT(-r)}{1 - wT(s)}$$

$$+ \sum_{k=0}^{r-1} F_k(-r) w^{k+1} \frac{\Gamma(s)(T(s) - T(-r))}{1 - wT(s)},$$

which we can handle in the same way as above. Actually, we proved (7.31) for $k \ge r$ with $f(-r) = 0$.

Finally, if $\Re(s)$ is positive (and $\Im(s)$ sufficiently close to $2\ell\pi/\log(q/p)$ for some integer ℓ) then we can also use

$$\sum_{k \ge 0} F_k(s) w^k = \frac{h(w,s)}{h(0,w)} \frac{1 - wT(0)}{1 - wT(s)}$$

and obtain the proposed result. (Note that $h(w,0)$ is analytic for $|w| < 1/T(0)$.)

Next suppose that $s = \sigma + it$, where t is not necessarily small. Then

$$T(s) = e^{it\log p} \left(p^{-\sigma} + q^{-\sigma} e^{it\log(q/p)} \right).$$

Consequently $|T(s)| = T(\rho)$, if and only if $t = 2j\pi/\log(q/p) = t_j$ for some integer j. Hence, if $|t - 2j\pi/\log(q/p)| \le \epsilon$ for some integer j we can do the same contour integration as above and again get (7.31).

Finally, if $|t - 2j\pi/\log(q/p)| > \epsilon$ for all integers j then we trivially estimate $F_k(s)$ by

$$|F_k(s)| \le \rho^{-k} \cdot \max_{|w|=\rho} |g(w,s)|,$$

where ρ is chosen in a way that $g(w,s)$ is analytic for $|w| \le \rho$. Since there is $\eta > 0$ with

$$|T(s-m)| = |p^{-\sigma+m} + e^{it\log(q/p)} q^{-\sigma+m}| \le e^{-2\eta} T(\sigma - m),$$

it follows that $h(w,s)$ exists for $|w| \le e^{\eta}/T(\sigma)$. Hence, we can actually set $\rho = e^{\eta}/T(\sigma)$ and obtain (7.32). In order to complete the proof, note that $\Delta_k^*(s) = -\Gamma(s) F_k(s)$ exists for $\Re(s) > -k$ and that $F_k(-r) = 0$ for $r = 0, 1, 2, \ldots$ and $k \ge r$. Thus, $f(-r) = 0$, too.

Putting these observations together it follows that we can represent $\Delta_k^*(s)$ as

$$\Delta_k^*(s) = \Gamma(s)f(s)T(s)^k \left(1 + O(e^{-\eta k})\right)$$

if $|\Im(s) - t_j| \le \epsilon$ for some integer j, where $\Gamma(s)f(s)$ is an entire function. Furthermore, we have proper upper bounds in the remaining range. Thus, we are precisely in the same situation as in the proof of Theorem 7.1. With the help of the Mellin inversion formula (with the line of integration $\Re(s) = \rho_{n,k}$) and by approximating the integral by infinitely many saddle point integrals around t_j we obtain

$$\Delta_k(ne^{i\theta}) = \frac{T(\rho)^k}{\sqrt{2\pi\beta(\rho)k}} \sum_{|j|\le j_0} f(\rho + it_j)\Gamma(\rho + it_j)(ne^{i\theta})^{-\rho - it_j} p^{-ikt_j} \times \left(1 + O\left(k^{-1/2}\right)\right) \tag{7.33}$$

uniformly for $|\theta| \le \theta_0$ (for some $0 < \theta_0 < \pi/2$).

It remains to have an a priori upper bound for $\Delta_k(ne^{i\theta})$ that is valid for all $|\theta| \le \pi$. Here we cannot use Lemma 7.6, since the recurrence relation is different. However, it can be shown that for every real ρ there exists a constant $C = C(\rho)$ and an integer $k_0 = k_0(\rho)$ such that

$$|e^x \Delta_k(x)| \le C(1 - c\theta^2)^{-k} r^{\max\{-\rho,0\}} T(\rho)^k e^{r(1-c\theta^2)} \tag{7.34}$$

for $k \ge k_0$ and uniformly for all $r \ge 0$ and $|\theta| \le \pi$, where $x = re^{i\theta}$. We indicate a proof for $\rho \le 0$. Obviously there exists k_0 and C such that (7.34) holds for $k = k_0$. Then by definition we have recursively (for $k \ge k_0$)

$$\begin{aligned}
|e^x \Delta_{k+1}(x)| &= \left| \int_0^x e^{\xi} \left(\Delta_k(p\xi) + \Delta_k(q\xi)\right) d\xi \right| \\
&= \left| \int_0^r e^{te^{i\theta}} \left(\Delta_k(pte^{i\theta}) + \Delta_k(qte^{i\theta})\right) dt \right| \\
&\le C(1 - c\theta^2)^{-k}\, T(\rho)^k \\
&\quad \times \int_0^r \left(e^{qt\cos\theta}(pt)^{-\rho} e^{pt(1-c\theta^2)} + e^{pt\cos\theta}(qt)^{-\rho} e^{qt(1-c\theta^2)}\right) dt \\
&\le C(1 - c\theta^2)^{-k} T(\rho)^{k+1} \int_0^r t^{-\rho} e^{t(1-c\theta^2)}\, dt \\
&\le C(1 - c\theta^2)^{-k-1} T(\rho)^{k+1}\, r^{-\rho} e^{r(1-c\theta^2)}.
\end{aligned}$$

A similar proof works for $\rho \ge 0$.

Finally by combining (7.33) and (7.34) with another Cauchy integration one obtains

$$\mathbb{E}\, B_{n,k} = \frac{n!}{2\pi i} \int_{|x|=n} e^x \Delta_k(x) \frac{dx}{x^{n+1}} \sim \Delta_k(n)$$

in the range of interest which also completes the proof of Theorem 7.12.

For the internal profile $I_{n,k}$ we use the Poisson transform

$$\overline{\Delta}_k(x) = e^{-x} \sum_{n\ge 0} \mathbb{E}\, I_{n,k}\, \frac{x^n}{n!}$$

that satisfies

$$\overline{\Delta}_k(x) + \overline{\Delta}_k'(x) = \overline{\Delta}_{k-1}(px) + \overline{\Delta}_{k-1}(qx), \qquad (k \ge 1), \tag{7.35}$$

with initial conditions $\overline{\Delta}_0(x) = 1 - e^{-x}$. The Mellin transforms $\overline{\Delta}_k^*(s)$ exist for $-k-1 < \Re(s) < 0$ and can be represented as

$$\overline{\Delta}_k^*(s) = -\Gamma(s)\overline{F}_k(s),$$

where $\overline{F}_0(s) = 1$ and

$$\overline{F}_k(s) - \overline{F}_k(s-1) = T(s)\overline{F}_{k-1}(s), \qquad (k \ge 1).$$

The corresponding explicit recurrence is

$$\overline{F}_k(s) = \mathbf{A}[\overline{F}_{k-1}](s) - \mathbf{A}[\overline{F}_{k-1}](-1), \qquad (k \ge 1),$$

and we have

$$\sum_{k\ge 0} \overline{F}_k(s) w^k = \frac{\sum_{\ell\ge 0} R_\ell(s) w^\ell}{\sum_{\ell \ge 0} R_\ell(-1) w^\ell}$$

and $\overline{F}_k(-\ell) = 0$ for $\ell = 1, 2, \ldots, k$. As above it follows that (for $|\Im(s) - t_j| \le \epsilon$ for some integer j)

$$\overline{F}_k(s) = \overline{f}(s) T(s)^k \left(1 + O(e^{-\eta k})\right),$$

where $\overline{f}(-r) = 0$ for $r = 1, 2, \ldots$. Thus,

$$\overline{\Delta}_k^*(s) = -\Gamma(s)\overline{f}(s)T(s)^k \left(1 + O(e^{-\eta k})\right)$$

has a polar singularity at $s = 0$ which induces the phase transition at the level $k = \alpha_0 \log n$. More precisely, if $k < \alpha_0 \log n$ one has to shift the line of integration $\Re(s) = \rho$ $(-k-1 < \rho < 0)$, where the Mellin transform exists, to $\Re(s) = \rho_{n,k} > 0$ so that the residue 2^k (at $\rho = 0$) has to be taken into account. In principle one can proceed as in the case of tries.

7.2.3 Symmetric Digital Search Trees

The symmetric case $p = q = \frac{1}{2}$ is different, since $\log \frac{1}{p}$ and $\log \frac{1}{q}$ coincide. Of course, the corresponding generating functions have as simpler structure. We have

$$\frac{\partial}{\partial x} G_k(x, u) = G_{k-1}\left(\frac{x}{2}, u\right)^2, \qquad (k \ge 1),$$

with initial conditions $G_0(x,u) = u + e^x - 1$ and $G_k(0,u) = 1$ (for $k > 0$) and

$$\frac{\partial}{\partial x} G_k^{[I]}(x,u) = G_{k-1}^{[I]}\left(\frac{x}{2},u\right)^2, \qquad (k \ge 1),$$

with initial conditions $G_0^{[I]}(x,u) = 1 + u(e^x - 1)$ and $G_k^{[I]}(0,u) = 1$. Thus, we obtain

$$E_k'(x) = 2e^{x/2} E_{k-1}\left(\frac{x}{2}\right), \tag{7.36}$$

with $E_0(x) = 1$ and $E_k(0) = 0$ for $k \ge 1$ or

$$\overline{E}_k'(x) = 2e^{x/2} \overline{E}_{k-1}\left(\frac{x}{2}\right), \tag{7.37}$$

with $\overline{E}_{k-1}(0) = e^x - 1$ and $\overline{E}_k(0) = 0$.

In this special case there is a simple explicit solution:

Lemma 7.18. *Set* $\gamma_0 = 1$ *and*

$$\gamma_\ell = \prod_{j=1}^{\ell} \left(1 - \frac{1}{2^j}\right) \qquad (\ell > 0).$$

Then we explicitly have

$$E_k(x) = 2^k e^x \sum_{m=0}^{k} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma_m \gamma_{k-m}} e^{-x2^{m-k}} \tag{7.38}$$

and

$$\overline{E}_k(x) = 2^k e^x \left(1 - \sum_{m=0}^{k} \frac{(-1)^m 2^{-\binom{m+1}{2}}}{\gamma_m \gamma_{k-m}} e^{-x2^{m-k}}\right), \tag{7.39}$$

or equivalently

$$\mathbb{E}\, B_{n,k} = 2^k \sum_{m=0}^{k} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma_m \gamma_{k-m}} \left(1 - \frac{1}{2^{k-m}}\right)^n \tag{7.40}$$

and

$$\mathbb{E}\, I_{n,k} = 2^k - 2^k \sum_{m=0}^{k} \frac{(-1)^m 2^{-\binom{m+1}{2}}}{\gamma_m \gamma_{k-m}} \left(1 - \frac{1}{2^{k-m}}\right)^n. \tag{7.41}$$

There are several ways to prove these relations. The simplest way is to use induction. It should be noted that the explicit formula (7.40) for $\mathbb{E}\, B_{n,k}$ has appeared several times in the literature [142, 143, 146, 177].

For the asymptotic analysis of $\mathbb{E}\, B_{n,k}$ and $\mathbb{E}\, I_{n,k}$ we introduce the function

$$F(z) = 1 - \sum_{m \ge 0} \frac{(-1)^m 2^{-\binom{m+1}{2}}}{\gamma \gamma_m} e^{-z2^m}, \tag{7.42}$$

where

$$\gamma = \prod_{j=1}^{\infty} \left(1 - \frac{1}{2^j}\right).$$

By definition we obviously have

$$F(z) = 1 - \frac{1}{\gamma} e^{-z} + O\left(e^{-2z}\right) \qquad (z \to \infty).$$

The asymptotic behaviour for $z \to 0+$ is much more involved and is covered in the following lemma.

Lemma 7.19. *The Laplace transform* $L(s) = \int_0^\infty F(z) e^{-sz}\, dz$ *is given by*

$$\begin{aligned} L(s) &= \frac{1}{s} - \sum_{m \ge 0} \frac{(-1)^m 2^{-\binom{m+1}{2}}}{\gamma\, \gamma_m \,(s + 2^m)} \\ &= \frac{1}{s} \prod_{j=0}^{\infty} \frac{1}{1 + s 2^{-j}}. \end{aligned}$$

Furthermore, for any fixed $r \ge 0$ *the* r*-th derivative* $F^{(r)}(z)$ *is asymptotically given by*

$$F^{(r)}(z) = C_r(z) 2^{-\frac{1}{2}\left(\log_2 \frac{1}{z}\right)^2 - \log_2 \frac{1}{z} \log_2 \log_2 \frac{1}{z}} \tag{7.43}$$

for $z \to 0+$*, where* $C_r(z)$ *is a function of at most polynomial growth for* $z \to 0+$*. In particular* $\lim_{z \to 0+} F(z) = 0$.

Remark 7.20 *Set* $\overline{L}(s) = sL(s)$ *which equals the Laplace transform of* $F'(z)$*. Then* $\overline{L}(s)$ *satisfies the functional equation*

$$\overline{L}(s) = s \sum_{j \ge 1} 2^j \overline{L}(s 2^j) \tag{7.44}$$

which follows from the fact that

$$\sum_{j \ge 1} \frac{2^j}{(1+2s)(1+2^2 s) \cdots (1+2^j s)} = \frac{1}{s}.$$

This relation (7.44) translates directly to the functional equation

$$F(z) = \sum_{j \ge 1} F'(z 2^{-j}).$$

Proof. Since $F(z)$ is bounded for $z \ge 0$, the Laplace transform $L(s)$ exists for s with $\Re(s) > 0$ and is given by

$$\int_0^\infty F(z) e^{-sz}\, dz = \frac{1}{s} - \sum_{m \ge 0} \frac{(-1)^m 2^{-\binom{m+1}{2}}}{\gamma\, \gamma_m \,(s + 2^m)}.$$

Since

$$\frac{1}{s}\prod_{j=0}^{\infty}\frac{1}{1+s2^{-j}} = \frac{1}{s} - \sum_{m\ge 0}\frac{(-1)^m 2^{-\binom{m+1}{2}}}{\gamma\,\gamma_m\,(s+2^m)},$$

we obtain the above representation of $L(s)$. With the help of the standard notation

$$Q(x) = \prod_{j=1}^{\infty}\left(1-\frac{x}{2^j}\right)$$

we also have

$$L(s) = \frac{1}{s\,Q(-2s)}.$$

Note that the Laplace transforms of the derivatives $F^{(r)}(z)$ are given by $s^r L(s)$.

In order to obtain the asymptotic expansion (7.43) we use the integral representation for the inverse Laplace transform

$$F^{(r)}(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} s^r L(s) e^{sz}\, ds,$$

where c is an arbitrary positive number (which will be properly chosen in the sequel). The idea is to use a kind of saddle point approximation of the integrand. For this purpose we need an asymptotic formula for $Q(-2x)$ for $x \to \infty$:

$$Q(-2x) = \exp\left(\frac{\log^2(x)}{2\log(2)} + \frac{\log x}{2} + f(\log_2(x)) + O\left(1/x\right)\right), \tag{7.45}$$

where $f(x)$ is a differentiable periodic function with period 1. This follows with the help of the Mellin transform applied to the logarithms. For $-1 < \Re(s) < 0$ we have

$$M(u) = \int_0^{\infty} \log Q(-2x) x^{u-1}\, dx = \frac{1}{1-2^u} g(u)$$

with

$$g(u) = \int_0^{\infty} \log(1+x) x^{u-1}\, dx = \frac{1}{u^2} + h(u),$$

where $h(u)$ has a meromorphic continuation to the whole complex plane with single poles at the positive integers $k \ge 1$ and residues $\mathrm{res}(h; u = k) = (-1)^k/k$. The inverse Mellin transform gives

$$Q(-2x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} M(u) x^{-u}\, du$$

with $-1 < c < 0$. By shifting the line of integration to the right and collecting the *contributions* from the triple pole at $u = 0$ and the single poles at $u = ki/\log 2$ ($k \in \mathbb{Z}\setminus\{0\}$), which constitute the periodic function, we obtain (7.45) (compare with [81]). It is also easy to extend the asymptotic relation (7.45) to a complex one of the form $|\arg(x)| \le \delta$ and $|x| \to \infty$, where δ is a small positive number.

Next we evaluate the r-th derivative $F^{(r)}(z)$ asymptotically. Let $0 < z < 1$ be given. We will compute the integral

$$F^{(r)}(z) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{s^{r-1}}{Q(-2s)} e^{sz}\, ds,$$

for

$$c = c(z) = \frac{\log(1/z)}{z \log 2} = \frac{\log_2(1/z)}{z}.$$

With the help of the usual Laplace method we obtain (after some algebra)

$$F^{(r)}(z) \sim c' e^{-f(\log_2 c(z))} c(z)^{r+\frac{1}{2}} (\log c(z))^{-\frac{3}{2}} e^{\log_2 \frac{1}{z} - \frac{\log 2}{2}(\log_2 c(z))^2},$$

which completes the proof of Lemma 7.19.

The main result on the expected profile of digital search trees is now formulated in terms of $F(z)$ resp. $F'(z)$.

Theorem 7.21. *We have*

$$\mathbb{E}\, B_{n,k} = 2^k F'(n2^{-k}) + F''(n2^{-k}) + O\left(n2^{-k}\right) \tag{7.46}$$

and

$$\mathbb{E}\, I_{n,k} = 2^k F(n2^{-k}) + F'(n2^{-k}) + O\left(n2^{-k}\right) \tag{7.47}$$

uniformly for all $n, k \ge 1$.

Remark 7.22 *These expansions are only of use if* $n2^{-k} \le 2^k$, *that is, for* $k \ge \frac{1}{2}\log_2 n$. *However, for small* k *there are no interesting phenomena. The range of interest is*

$$\log_2 n - \log\log n \le k \le \log_2 n + \sqrt{2\log_2 n},$$

and this range is covered by Theorem 7.21. Nevertheless, with slightly more care it is easy to obtain more precise expansions, e.g.

$$\mathbb{E}\, B_{n,k} = 2^k F'(n2^{-k}) + F''(n2^{-k}) - \frac{n}{2^{k+1}} F'''(n2^{-k}) + O\left(n4^{-k}\right) + O\left(2^{-k}\right).$$

Proof. We just have to use the explicit representations of Lemma 7.18 and approximate the leading terms by $2^k F'(n2^{-k})$ resp. by $2^k F(n2^{-k})$.

In particular, we first show that the terms for $m \ge k/3$ do not count:

$$\begin{aligned}\left| \sum_{m>k/3} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma_m \gamma_{k-m}} \left(1 - 2^{m-k}\right)^n \right| &\ll \sum_{m>k/3} 2^{-\binom{m}{2}} \\ &\ll 2^{-\binom{\lfloor k/3 \rfloor}{2}}.\end{aligned}$$

In a next step we use the approximation (for $m \le k/3$)

$$\left(1-2^{m-k}\right)^n = e^{-n2^{m-k}}\left(1+O\left(\frac{n}{4^{k-m}}\right)\right) \\ = e^{-n2^{m-k}} + O\left(\frac{n}{4^{k-m}}\right),$$

and obtain an error term of the form

$$\sum_{m\le k/3} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma_m \gamma_{k-m}} \cdot O\left(\frac{n}{4^{k-m}}\right) = O\left(\frac{n}{4^k}\right).$$

Finally, we approximate the ratio

$$\frac{\gamma}{\gamma_m} = 1 - \frac{1}{2^{k-m}} + O\left(\frac{1}{4^{k-m}}\right)$$

and get an even smaller error of order $O\left(4^{-k}\right)$.

Summing up we get

$$\begin{aligned} \mathbb{E}\, B_{n,k} &= 2^k \sum_{m=0}^{k} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma_m \gamma_{k-m}} \left(1-2^{m-k}\right)^n \\ &= 2^k \sum_{m\le k/3} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma \gamma_m} \left(1 - \frac{1}{2^{k-m}}\right) e^{-n2^{m-k}} \\ &+ O\left(2^k \frac{n}{4^k}\right) + O\left(2^k 2^{-\binom{\lfloor k/3 \rfloor}{2}}\right) \\ & 2^k \sum_{m=0}^{\infty} \frac{(-1)^m 2^{-\binom{m}{2}}}{\gamma \gamma_m} \left(1 - \frac{2^m}{2^k}\right) e^{-n2^{m-k}} + O\left(\frac{n}{2^k}\right) \\ &= 2^k F'(n2^{-k}) + F''(n2^{-k}) + O\left(n2^{-k}\right). \end{aligned}$$

This completes the proof of (7.46). The proof of (7.47) is exactly the same.

8

Recursive Algorithms and the Contraction Method

Recursive Algorithms follow the Divide-and-Conquer-Principle. The idea behind this principle is to break a problem into two or several subproblems which are easier to solve (heuristically because they are just smaller problems) and then to construct the solution of the original problem by combining the solutions of the subproblems appropriately. If this idea is iteratively (or recursively) applied then one speaks of a *recursive algorithm* and, moreover, these kinds of algorithms give rise to a (hidden) tree structure.

The most prominent example in this context is the widely used sorting algorithm *Quicksort* which was invented by C. A. R. Hoare [100, 101]. It is the standard sorting procedure in Unix systems, and the basic idea can be described easily:

> A list of n real numbers $A = (x_1, x_2, \ldots, x_n)$ is given. Select a pivot element x_j from this list. Divide the remaining numbers into sets $A_\le, A_>$ of numbers smaller (or equal) and larger than x_j. Next apply the same procedure to each of these two sets if they contain more than one element. Finally, we end up with a sorted list of the original numbers.

This sorting procedure can be encoded with a binary tree with n (internal) nodes. The first selected element x_j is put to the root, whereas recursively $A_\le$ produces a left subtree of x_j and $A_>$ the right subtree of x_j. (An empty string produces an empty tree which is usually encoded as an external node.)

For example, if one always uses the first element of the list as the pivot element then the tree structure behind Quicksort is precisely the binary search tree associated to the same input data (see Figure 8.1 and compare with Section 6.1.2). In particular, the number of comparisons needed for Quicksort equals the internal path length of the corresponding binary search tree, and the maximum number of recursive calls equals the height of the tree. The analysis of Quicksort is therefore – despite of a reformulation – the same as the analysis of binary search trees.

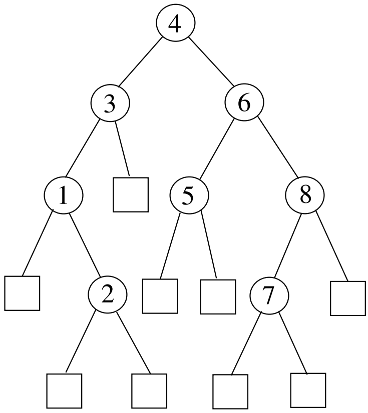


Fig. 8.1. Tree representation of Quicksort applied to the input list $(4, 6, 3, 5, 1, 8, 2, 7)$

In Computer Science one mainly does a "worst case" analysis or an "average case" analysis. While for the worst case one just takes the supremum of the complexity over all possible inputs (of a given size), for an average case analysis a certain probability measure is assumed on the set of possible inputs. This is often the uniform distribution if the possible inputs form a finite set. For example, the standard model for Quicksort is to assume that all permutations of the input data are equally likely.

At this point one should also mention that various tree types serve for the organisation of data (for example, digital search trees or binary search trees). Many of them indeed have a recursive structure. This indicates that there is a close relation between these recursive algorithms and algorithms on tree structures. Anyway, the analysis of recursive algorithms covers the analysis of many interesting random tree structures.

The purpose of this chapter is to describe a very powerful method for the average case analysis of recursive algorithms, the so-called *contraction method.* The method is tailored to derive convergence in distribution for parameters of recursive structures. It was introduced in Rösler [186] and later independently extended in Rösler [187] and Rachev and Rüschendorf [179] and has been developed to a fairly general tool during the last years.

The procedure of this method runs a follows. One starts with a recurrence satisfied by the quantities of interest and, based on information on the first two moments, one does a proper normalisation of the quantities. The recurrence for the scaled quantities leads to a fixed-point equation and a potential limit distribution is characterised as the fixed-point of a measure valued map.

This chapter was very much inspired by Neininger's habilitation thesis [160].

8.1 The Number of Comparisons in Quicksort

In order to introduce the contraction method we consider the complexity of Quicksort. This was actually the first analysis that was based on the contraction method (see Rösler [186]).

Let Y_n denote the number of comparisons which are needed to sort a random permutation of $\{1, 2, \ldots, n\}$. The recursive description of Quicksort immediately translates to

$$Y_n \stackrel{\mathrm{d}}{=} Y_{I_n} + Y^*_{n-1-I_n} + n - 1, \qquad n \ge 2, \tag{8.1}$$

where $Y_0 = Y_1 = 0$, $Y_2 = 1$, I_n is uniformly distributed on $\{0, 1, \ldots, n-1\}$, $Y_j \stackrel{\mathrm{d}}{=} Y_j^*$, and I_n, Y_j, Y_j^* $(1 \le j \le n)$ are independent.

Equivalently we can consider the generating functions

$$y_n(u) = \mathbb{E}\, u^{Y_n}$$

for which we get the recurrence

$$y_n(u) = \frac{u^{n-1}}{n} \sum_{j=1}^{n} y_{j-1}(u) y_{n-j}(u).$$

Consequently the differential equation for the double generating function $Y(x,u) = \sum_{n\ge 0} y_n(u) x^n$ is

$$\frac{\partial Y(x,u)}{\partial x} = Y(xu, u)^2$$

with the side conditions

$$\frac{\partial Y(0,u)}{\partial x} = 1 \quad \text{and} \quad Y(x,1) = \frac{1}{1-x}.$$

We will not analyse Quicksort using these relations. In fact, due to the argument xu on the right-hand-side and due to the non-linearity of the equation it is not clear whether this partial differential equation can be analysed directly (in both variables). Of course, one can consider derivatives with respect to u and derive asymptotics for moments. Nevertheless, we want to note that the corresponding generating functions for the profile and the height of binary search trees have a similar structure (compare with Section 6.5).

Instead of using generating functions, we work with (8.1). For example, it is easy to obtain explicit representations for the expected value $\mathbb{E} Y_n$. The recurrence

$$\mathbb{E}\, Y_n = n - 1 + \frac{1}{n} \sum_{j=1}^{n} (\mathbb{E} Y_{j-1} + \mathbb{E} Y_{n-j}) \tag{8.2}$$

can be explicitly solved and one also obtains an asymptotic expansion:

$$\begin{aligned}\mathbb{E}\, Y_n &= 2(n+1)\sum_{h=1}^{n+1}\frac{1}{h} - 4(n+1) + 2 \\ &= 2n\log n + n(2\gamma - 4) + 2\log n + 2\gamma + 1 + O\left((\log n)/n\right) \qquad (8.3)\end{aligned}$$

with $\gamma = 0.57721...$ being Euler's constant.

In fact, much more is known about this random variable.

Theorem 8.1. *Let Y_n denote the number of comparisons which are needed to sort a random permutation of $\{1, 2, \ldots, n\}$. Then we have*

$$\frac{Y_n - \mathbb{E}\, Y_n}{n} \xrightarrow{d} X,$$

where the distribution of the random variable X is defined by the fixed point equation

$$X \stackrel{\mathrm{d}}{=} UX + (1-U)X^* + b(U), \qquad (8.4)$$

where U is uniformly distributed on $[0,1]$, $X \stackrel{\mathrm{d}}{=} X^*$, U, X, X^* are independent, and

$$b(x) = 2x\log x + 2(1-x)\log(1-x) + 1.$$

The existence of a limiting distribution X (the *Quicksort distribution*) was first observed by Régnier [182] via a martingale approach, whereas the characterisation of X with a fixed point equation is due to Rösler [186]. It is now also known that there exists a density ([201]), which is a bounded C^∞ function; tail estimates are available, and orders of convergence are estimated (compare with [77, 78, 79, 125]). However, no explicit representations for the limiting distribution are known.

From the fixed point equation (8.4) it also possible to calculate moments step by step, e.g. the variance of X is given by

$$\mathbb{V}\mathrm{ar}\, X = 7 - \frac{2}{3}\pi^2.$$

We briefly describe Rösler's approach. His main observation was that (8.4) has a unique fixed point (because it constitutes a contraction with respect to the Wasserstein metric ℓ_2).

Let $\mathcal{M}_2$ denote the space of measures on $\mathbb{R}$ with finite second moment and zero first moment. Then the Wasserstein metric (or L_2 metric) ℓ_2 is defined as

$$\ell_2(\mu,\nu) = \inf \|X - Y\|_2,$$

where $\|\cdot\|_2$ denotes the L_2-norm and the infimum is taken over all random variables X with law $\mu = \mathcal{L}(X)$ and all Y with law $\nu = \mathcal{L}(Y)$. It is well known that $(\mathcal{M}_2, \ell_2)$ constitutes a Polish space and that a sequence μ_n converges to μ in $\mathcal{M}_2$, if and only if μ_n converges weakly to μ and if the second moments of μ_n converge to the second moment of μ.

Let us consider the normalised random variables $X_n = (Y_n - \mathbb{E}Y_n)/n$. From (8.1) we get

$$X_n \stackrel{\mathrm{d}}{=} X_{I_n}\frac{I_n}{n} + X^*_{n-1-I_n}\frac{n-1-I_n}{n} + b_n(I_n), \qquad n \geq 2, \tag{8.5}$$

where $X_0 = X_1 = 0$, I_n is uniformly distributed on $\{0, 1, \ldots, n-1\}$, and $X_j = X_j^*$, and I_n, X_j, X_j^* $(1 \leq j \leq n)$ are independent. Furthermore,

$$\begin{aligned} b_n(j) &= \frac{n-1}{n} + \frac{1}{n}\left(\mathbb{E}\, Y_j + \mathbb{E}\, Y_{n-1-j} - \mathbb{E}\, Y_n\right) \\ &= 1 + 2\frac{j}{n}\log\frac{j}{n} + 2\left(1 - \frac{j}{n}\right)\log\left(1 - \frac{j}{n}\right) + O\left(\frac{\log n}{n}\right). \end{aligned}$$

Thus, if X_n has a limiting distribution X then it has to satisfy (8.4).

The first step is to show that (8.4) actually has a unique solution with $\mathbb{E}Y = 0$.

Lemma 8.2. *Let $T : \mathcal{M}_2 \to \mathcal{M}_2$ be a map defined by*

$$T(\mu) = \mathcal{L}(UX + (1-U)X^* + b(U)),$$

where X, X^*, U are independent, $\mathcal{L}(X^*) = \mathcal{L}(X) = \mu$, and U is uniformly distributed on $[0, 1]$. Then T is a contraction with respect to the Wasserstein metric ℓ_2 and, thus, there is a unique fixed point $\mu \in \mathcal{M}_2$ with $T(\mu) = \mu$.

Proof. Let $\mu, \nu \in \mathcal{M}_2$ and suppose that $\mathcal{L}(X^*) = \mathcal{L}(X) = \mu$, $\mathcal{L}(Y^*) = \mathcal{L}(Y) = \nu$, and U is uniformly distributed on $[0, 1]$ such that U, X, X^* and U, Y, Y^* are independent. Then $T(\mu) = \mathcal{L}(UX + (1-U)X^* + b(U))$, $T(\nu) = \mathcal{L}(UY + (1 - U)Y^* + b(U))$, and consequently

$$\begin{aligned} \ell_2^2(T(\mu), T(\nu)) &\leq \|UX + (1-U)X^* - UY - (1-U)Y^*\|_2^2 \\ &= \|U(X-Y) + (1-U)(X^* - Y^*)\|_2^2 \\ &= \mathbb{E}(X-Y)^2 \cdot \mathbb{E}U^2 + \mathbb{E}(X^* - Y^*)^2 \cdot \mathbb{E}(1-U)^2 \\ &= \frac{2}{3}\mathbb{E}(X-Y)^2. \end{aligned}$$

Taking the infimum over all possible X, Y we obtain

$$\ell_2(T(\mu), T(\nu)) \leq \sqrt{\frac{2}{3}} \cdot \ell_2(\mu, \nu),$$

which completes the proof of the lemma.

The next step is to show that the X_n actually converge to X. (Recall that $\ell_2(\mathcal{L}(X_n), \mathcal{L}(X)) \to 0$ implies $X_n \xrightarrow{d} X$.)

Lemma 8.3. *We have*

$$\ell_2^2(\mathcal{L}(X_n), \mathcal{L}(X)) \le \frac{2}{n}\sum_{j=1}^{n}\left(\frac{j-1}{n}\right)^2 \ell_2^2(\mathcal{L}(X_{j-1}), \mathcal{L}(X)) + O\left(\frac{\log n}{n}\right)$$

and consequently $\lim_{n\to\infty} \ell_2(\mathcal{L}(X_n), \mathcal{L}(X)) = 0.$

Proof. Let $\nu_n = \mathcal{L}(X_n)$, where X_n is the normalised number of comparisons, and let Y and Y^* be independent with $\mathcal{L}(Y) = \mathcal{L}(Y^*) = \mu$, where μ is the unique fixed point of $T(\mu) = \mu$. Next choose versions X_j, X_j^* (which are independent for $1 \le j \le n-1$) with

$$\mathbb{V}\mathrm{ar}(X_j - X) = \ell_2^2(\nu_j - \mu) \quad \text{and} \quad \mathbb{V}\mathrm{ar}(X_j^* - X^*) = \ell_2^2(\nu_j - \mu)$$

and set $V_x = X_j$ and $V_x^* = X_j^*$ for $x \in (\frac{j}{n}, \frac{j+1}{n}]$. Then, for U uniformly distributed on $[0,1]$ and independent of X_j and X_j^*, we have

$$\nu_n = \mathcal{L}\left(\frac{\lceil nU\rceil - 1}{n} V_U + \frac{n - \lceil nU\rceil}{n} V_{1-U}^* + b_n(\lceil nU\rceil)\right).$$

Observe that

$$\sup_{0\le x\le 1} |b_n(\lceil nx\rceil) - b(x)| = O\left(\frac{\log n}{n}\right). \tag{8.6}$$

By definition of the Wasserstein metric we have

$$\begin{aligned}
\ell_2^2(\nu_n, \mu) \le \mathbb{E}\Bigg(&\frac{\lceil nU\rceil - 1}{n} V_U - UX + \frac{n - \lceil nU\rceil}{n} V_{1-U}^* \\
&- (1-U)X^* + b_n(\lceil nU\rceil) - b(U)\Bigg)^2 \\
= {}&\mathbb{E}\left(\frac{\lceil nU\rceil - 1}{n} V_U - UX\right)^2 + \mathbb{E}\left(\frac{n - \lceil nU\rceil}{n} V_{1-U}^* - (1-U)X^*\right)^2 \\
&+ \mathbb{E}\left(b_n(\lceil nU\rceil) - b(U)\right)^2 \\
&+ 2\,\mathbb{E}\left(\left(\frac{\lceil nU\rceil - 1}{n} V_U - UX\right)\left(\frac{n - \lceil nU\rceil}{n} V_U^* - (1-U)X^*\right)\right) \\
&+ 2\,\mathbb{E}\left(\left(\frac{\lceil nU\rceil - 1}{n} V_U - UX\right)\left(b_n(\lceil nU\rceil) - b(U)\right)\right) \\
&+ 2\,\mathbb{E}\left(\left(\frac{n - \lceil nU\rceil}{n} V_{1-U}^* - (1-U)X^*\right)\left(b_n(\lceil nU\rceil) - b(U)\right)\right).
\end{aligned}$$

The first term can be rewritten to

$$
\begin{aligned}
&\mathbb{E}\left(\frac{\lceil nU\rceil - 1}{n}V_U - UX\right)^2 \\
&= \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{(\frac{j-1}{n},\frac{j}{n}]}(U)\left(\frac{j-1}{n}X_{j-1} - UX\right)\right)^2 \\
&= \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{(\frac{j-1}{n},\frac{j}{n}]}(U)\left(\frac{j-1}{n}\left(X_{j-1} - X\right) + O\left(\frac{|X|}{n}\right)\right)\right)^2 \\
&= \sum_{j=1}^{n} \mathbb{E}\left(\mathbf{1}_{(\frac{j-1}{n},\frac{j}{n}]}(U)\left(\frac{j-1}{n}\left(X_{j-1} - X\right)\right)\right)^2 \\
&+ O\left(\sum_{j=1}^{n} \mathbb{E}\left(\mathbf{1}_{(\frac{j-1}{n},\frac{j}{n}]}(U)\frac{j-1}{n}\left|X_{j-1} - X\right|\frac{|X|}{n}\right)\right) + O\left(\frac{1}{n^2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\frac{(j-1)^2}{n^2}\ell_2^2(\nu_{j-1},\mu) + O\left(\frac{1}{n^2}\sum_{j=1}^{n}\frac{j-1}{n}\ell_2(\nu_{j-1},\mu)\right) + O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

The same holds for the second term. The third term is estimated using (8.6):

$$
\mathbb{E}\left(b_n(\lceil nU\rceil) - b(U)\right)^2 = O\left(\frac{(\log n)^2}{n^2}\right)
$$

By using the independence assumptions and $\mathbb{E}\,X_j = \mathbb{E}\,X = 0$ we also get

$$
\begin{aligned}
&\mathbb{E}\left(\left(\frac{\lceil nU\rceil - 1}{n}V_U - UX\right)\left(\frac{n - \lceil nU\rceil}{n}V_U^* - (1-U)X^*\right)\right) \\
&\qquad = O\left(\frac{1}{n^2}\sum_{j=1}^{n}\frac{j-1}{n}\ell_2(\nu_{j-1},\mu)\right) + O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

Finally we have for the fifth (and similarly for the sixth) term

$$
\begin{aligned}
&\mathbb{E}\left(\left(\frac{\lceil nU\rceil - 1}{n}V_U - UX\right)\left(b_n(\lceil nU\rceil) - b(U)\right)\right) \\
&\qquad = O\left(\frac{\log n}{n^2}\sum_{j=1}^{n}\frac{j-1}{n}\ell_2(\nu_{j-1},\mu)\right) + O\left(\frac{\log n}{n^2}\right).
\end{aligned}
$$

Thus, with $a_j = \ell_2^2(\nu_j,\mu)$ one has

$$
\begin{aligned}
a_n = \frac{2}{n}\sum_{j=1}^{n}\left(\frac{j-1}{n}\right)^2 a_{j-1} + O\left(\frac{\log n}{n^2}\right)\sum_{j=1}^{n}\frac{j-1}{n}\sqrt{a_{j-1}} \\
+ O\left(\frac{(\log n)^2}{n^2}\right)
\end{aligned}
$$

and thus

$$a_n \leq \frac{2}{3} \max_{0 \leq j \leq n-1} a_j + C_1 \frac{\log n}{n} \max_{0 \leq j \leq n-1} \sqrt{a_j} + C_2 \frac{\log^2 n}{n^2}.$$

for some positive constants C_1, C_2. It is now easy to show that $a_n \to 0$. First it follows that $a_n \leq \overline{A}$, where $\overline{A}$ satisfies $a_1 \leq \overline{A}$ and the inequality

$$\overline{A} \leq \frac{2}{3}\overline{A} + C_1 \frac{\log n}{n} \sqrt{\overline{A}} + C_2 \frac{\log^2 n}{n^2}$$

for all $n \geq 2$. Consequently we also have

$$a_n \leq \frac{2}{3} \max_{0 \leq j \leq n-1} a_j + C_3 \frac{\log n}{n}$$

for a proper constant C_3 which shows that $a_n \to 0$, as proposed.

8.2 The L_2 Setting of the Contraction Method

The method that has been discussed in Section 8.1 can be generalised in several ways. For this purpose we specify a class of recursive sequences of distributions that can be analysed by using L_2 techniques. In order to be a bit more general we develop the theory for random vectors in $\mathbb{R}^d$.

8.2.1 A General Type of Recurrence

We consider a sequence of random, d-dimensional vectors $(Y_n)_{n \geq 0}$ which satisfy the recurrence

$$Y_n \stackrel{\mathrm{d}}{=} \sum_{r=1}^{K} A_r(n) Y^{(r)}_{I_r^{(n)}} + b_n, \qquad (n \geq n_0), \tag{8.7}$$

where $(A_1(n), \ldots, A_K(n), b_n, I^{(n)}), (Y_n^{(1)}), \ldots, (Y_n^{(K)})$ are independent, $A_1(n), \ldots, A_K(n)$ are random $d \times d$ matrices, b_n is a random d-dimensional vector, $I^{(n)}$ is a vector of random integers $I_r^{(n)} \in \{0, \ldots, n\}$, and $(Y_n^{(1)}), \ldots, (Y_n^{(K)})$ are identical distributed as (Y_n). We have $n_0 \geq 1$ and $Y_0, \ldots, Y_{n_0-1}$ are given initialising random vectors. The number $K \geq 1$ is deterministic.

We rescale the random vector Y_n in (8.7) by

$$X_n = C_n^{-1/2}(Y_n - M_n) \qquad (n \geq 0), \tag{8.8}$$

where $M_n \in \mathbb{R}^d$ and C_n is a symmetric, positive-definite $d \times d$ matrix. If the first two moments of Y_n are finite then M_n and C_n are typically of the order of the expectation and the covariance matrix of Y_n respectively. From recurrence (8.7) we obtain for X_n the modified recurrence

$$X_n \stackrel{\mathrm{d}}{=} \sum_{r=1}^{K} A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \tag{8.9}$$

with

$$A_r^{(n)} = C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \tag{8.10}$$

$$b^{(n)} = C_n^{-1/2} \left(b_n - M_n + \sum_{r=1}^{K} A_r(n) M_{I_r^{(n)}} \right), \tag{8.11}$$

and independence properties as in (8.7).

The contraction method aims to provide assertions of the following type:

> Suppose that we (X_n) is characterised by (8.9). Then appropriate convergence of the coefficients
>
> $$A_r^{(n)} \xrightarrow{d} A_r^*, \quad b^{(n)} \xrightarrow{d} b^*, \quad (n \to \infty) \tag{8.12}$$
>
> implies convergence in distribution of the quantities (X_n) to a limit X. The limit distribution $\mathcal{L}(X)$ is characterised by a fixed-point equation, which is obtained from the modified recurrence by formally letting $n \to \infty$:
>
> $$X \stackrel{\mathrm{d}}{=} \sum_{r=1}^{K} A_r^* X^{(r)} + b^*. \tag{8.13}$$
>
> Here, $(A_1^*, \ldots, A_K^*, b^*)$, $X^{(1)}, \ldots, X^{(K)}$ are independent and $X^{(r)} \stackrel{\mathrm{d}}{=} X$ for $r = 1, \ldots, K$.

To reformulate the fixed-point property, we denote by $\mathcal{M}^d$ the space of all probability measures on $\mathbb{R}^d$ and by T the measure valued map

$$T : \mathcal{M}^d \to \mathcal{M}^d, \quad \mu \mapsto \mathcal{L}\left(\sum_{r=1}^{K} A_r^* Z^{(r)} + b^* \right), \tag{8.14}$$

where $(A_1^*, \ldots, A_K^*, b^*)$, $Z^{(1)}, \ldots, Z^{(K)}$ are independent and $\mathcal{L}(Z^{(r)}) = \mu$ for $r = 1, \ldots, K$. Then, X is a solution of the fixed-point equation (8.13), if and only if its distribution $\mathcal{L}(X)$ is a fixed-point of the map T.

Remark 8.4 *Maps of type (8.14) do not often have unique fixed-points in the space of all probability distributions, and the characterisation of the set of all fixed-points is – up to a few special cases – an open and important problem (see Section 8.3).*

In order to have a more precise study we define the following subsets of $\mathcal{M}^d$:

$$\mathcal{M}_s^d = \{\mu \in \mathcal{M}^d \ : \ \|\mu\|_s < \infty\}, \quad s > 0, \tag{8.15}$$

$$\mathcal{M}_s^d(M) = \{\mu \in \mathcal{M}_s^d \ : \ \mathbb{E}\mu = M\}, \quad s \geq 1, \tag{8.16}$$

$$\mathcal{M}_s^d(M, C) = \{\mu \in \mathcal{M}_s^d(M) \ : \ \mathbb{C}\mathrm{ov}(\mu) = C\}, \quad s \geq 2, \tag{8.17}$$

where $M \in \mathbb{R}^d$ and C is a symmetric, positive-definite $d \times d$ matrix, and where $\|\mu\|_s$, $\mathbb{E}\mu$ and $\mathbb{C}\mathrm{ov}(\mu)$ denote the s-th absolute moment, expectation and covariance matrix, respectively, of a random variable with distribution μ.

The core of the method is to endow an appropriate subset $\mathcal{M}^* \subset \mathcal{M}^d$, e.g., on one of the sets in (8.15)–(8.17), with a complete metric δ such that the restriction of T to $\mathcal{M}^*$ is a contraction on the metric space $(\mathcal{M}^*, \delta)$ in the sense of Banach's fixed-point theorem. This implies the existence of a fixed-point $\mathcal{L}(X)$ of T, being unique in $\mathcal{M}^*$. In a second step one shows convergence of the rescaled quantities $\mathcal{L}(X_n)$ to $\mathcal{L}(X)$ in the metric δ, $\delta(\mathcal{L}(X_n), \mathcal{L}(X)) \to 0$ for $n \to \infty$, based on an appropriate convergence of the coefficients as in (8.12). If δ is chosen such that convergence in δ implies weak convergence then the desired convergence in distribution follows.

8.2.2 A General L_2 Convergence Theorem

One crucial point is to choose a proper metric for which one can show that the mapping T is a contraction. A natural choice is the minimal L_p metric ℓ_p (for some $p > 0$) that is defined by

$$\ell_p(\mu, \nu) := \inf \{\|X - Y\|_p \ : \ \mathcal{L}(X) = \mu, \, \mathcal{L}(Y) = \nu\}, \quad \mu, \nu \in \mathcal{M}_p^d,$$

where $\|X\|_p := (\mathbb{E}\,\|X\|^p)^{\min\{(1/p),1\}}$ denotes the L_p norm of a random vector X and $\|X\|$ denotes its Euclidean norm. The Wasserstein metric ℓ_2 is a special but important case.

Furthermore, the spaces $(\mathcal{M}_p^d, \ell_p)$ for $p > 0$ as well as $(\mathcal{M}_p^d(M), \ell_p)$ for $M \in \mathbb{R}^d$, $p \geq 1$ are complete metric spaces and convergence in ℓ_p is equivalent to weak convergence plus convergence of the p-th absolute moment. For $\mu, \nu \in \mathcal{M}_p^d$ there always exist vectors X, Y on a joint probability space with $\mathcal{L}(X) = \mu$, $\mathcal{L}(Y) = \nu$ and $\ell_p(\mu, \nu) = \|X - Y\|_p$. Such vectors are called optimal couplings of μ and ν. For these and further properties of the minimal L_p metric ℓ_p see Dall'Aglio [43], Major [148], Bickel and Freedman [16], Rachev [178], and Rachev and Rüschendorf [180].

In order to obtain contraction properties of the map (8.14) we denote by

$$\|A\|_{\mathrm{op}} = \sup_{\|x\|=1} \|Ax\|$$

the operator norm of a square matrix A, by A^{T} the transpose of the matrix A, and by

$$\mathrm{lip}(T) = \inf_{\mu,\nu \in \mathcal{M}^*, \, \mu \neq \nu} \frac{\ell_p(T(\mu), T(\nu))}{\ell_p(\mu, \nu)}$$

the Lipschitz constant of T if it is finite. Actually, in the case $p = 2$ the Lipschitz constant can be estimated easily.

Lemma 8.5. *Let $(A_1^*, \ldots, A_K^*, b^*)$ be an L_2-integrable vector of random $d \times d$ matrices $A_1^* \ldots, A_K^*$ and a random d-dimensional vector b^* with $\mathbb{E} b^* = 0$ and assume that T is as in (8.14). Then the restriction of T to $\mathcal{M}_2^d(0)$ is Lipschitz continuous in ℓ_2, and for the Lipschitz constant* $\operatorname{lip}(T)$ *we have*

$$\operatorname{lip}(T) \leq \left\| \mathbb{E}\left(\sum_{r=1}^{K} (A_r^*)^{\mathrm{T}} A_r^* \right) \right\|_{\mathrm{op}}^{1/2}. \tag{8.18}$$

Proof. Clearly, if $\mu \in \mathcal{M}_2^d(0)$, then $T(\mu)$ has a finite second moment and $\mathbb{E}\, T(\mu) = 0$. We just have to apply the independence conditions and the assumption $\mathbb{E}\, b = 0$. Thus, $T : \mathcal{M}_2^d(0) \to \mathcal{M}_2^d(0)$ is a well-defined map. Let $\mu, \nu \in \mathcal{M}_2^d(0)$ be given and $(W^{(1)}, Z^{(1)}), \ldots, (W^{(K)}, Z^{(K)})$ be optimal couplings of (μ, ν) for $r = 1, \ldots, K$ so that $(A_1, \ldots, A_k, b)$, $(W^{(1)}, Z^{(1)}), \ldots, (W^{(K)}, Z^{(K)})$ are independent. Then

$$\begin{aligned}
\ell_2^2(T(\mu), T(\nu)) &\leq \mathbb{E}\left(\left\| \sum_{r=1}^{K} A_r (W^{(r)} - Z^{(r)}) \right\|^2 \right) \\
&= \mathbb{E}\left(\sum_{r=1}^{K} \left(A_r (W^{(r)} - Z^{(r)}) \right)^{\mathrm{T}} \cdot \left(A_r (W^{(r)} - Z^{(r)}) \right) \right) \\
&\quad + \mathbb{E}\left(\sum_{1 \leq r \neq s \leq K} \left(A_r (W^{(r)} - Z^{(r)}) \right)^{\mathrm{T}} \cdot \left(A_s (W^{(r)} - Z^{(r)}) \right) \right) \\
&= \mathbb{E}\left(\sum_{r=1}^{K} \left(W^{(r)} - Z^{(r)} \right)^{\mathrm{T}} \cdot \left(A_r^{\mathrm{T}} A_r (W^{(r)} - Z^{(r)}) \right) \right) \\
&= \mathbb{E}\left(\sum_{r=1}^{K} \left(W^{(r)} - Z^{(r)} \right)^{\mathrm{T}} \cdot \left(\mathbb{E}(A_r^{\mathrm{T}} A_r)(W^{(r)} - Z^{(r)}) \right) \right) \\
&= \mathbb{E}\left(\left(W^{(1)} - Z^{(1)} \right)^{\mathrm{T}} \cdot \left(\left(\sum_{r=1}^{K} \mathbb{E}(A_r^{\mathrm{T}} A_r) \right) (W^{(1)} - Z^{(1)}) \right) \right) \\
&\leq \left\| \sum_{r=1}^{K} \mathbb{E}(A_r^{\mathrm{T}} A_r) \right\|_{\mathrm{op}} \mathbb{E}\left(\left\| W^{(1)} - Z^{(1)} \right\|^2 \right) \\
&= \left\| \sum_{r=1}^{K} \mathbb{E}(A_r^{\mathrm{T}} A_r) \right\|_{\mathrm{op}} \ell_2^2(\mu, \nu).
\end{aligned}$$

Note that we have used the independence of $W^{(r)} - Z^{(r)}$ and $W^{(s)} - Z^{(s)}$ for $r \neq s$ together with the fact that $\mathbb{E}(W^{(r)} - Z^{(r)}) = 0$. The additional expectation $\mathbb{E}(A_r^{\mathrm{T}} A_r)$ is also justified by independence.

The second step of the contraction method is to show convergence in ℓ_2 for sequences $(\mathcal{L}(X_n))$ of the form (8.9). The following Theorem provides sufficient conditions (see [187, Theorem 3.1] and [158, Theorem 4.1]).

Theorem 8.6. *Assume (X_n) is a sequence of centred d-dimensional, L_2-integrable random vectors, satisfying the recurrence (8.9) with L_2-integrable random $d \times d$ matrices $A_r^{(n)}$, $1 \le r \le K$, a random L_2-integrable centred vector $b^{(n)}$, and a vector of random integers $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$ with values in $I_r^{(n)} \in \{0, 1, \dots, n\}$. Assume that we have*

$$\left(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}\right) \xrightarrow{L_2} (A_1^*, \dots, A_K^*, b^*) \quad (n \to \infty), \tag{8.19}$$

$$\mathbb{E}\left(\sum_{r=1}^K \left\|(A_r^*)^{\mathrm{T}} A_r^*\right\|_{\mathrm{op}}\right) < 1, \tag{8.20}$$

$$\mathbb{E}\left(\mathbf{1}_{\{I_r^{(n)} \le \ell\}} \left\|(A_r^{(n)})^{\mathrm{T}} A_r^{(n)}\right\|_{\mathrm{op}}\right) \to 0, \quad (n \to \infty), \tag{8.21}$$

for all integers $\ell \ge 0$ and $r = 1, \dots, K$. Then we have

$$\ell_2(\mathcal{L}(X_n), \mathcal{L}(X)) \to 0 \quad (n \to \infty),$$

where $\mathcal{L}(X)$ is the fixed-point of map T defined in (8.14), which is unique in $\mathcal{M}_2^d(0)$.

Before giving the proof we comment on the conditions of Theorem 8.6 and their applicability.

Condition (8.19) means that the convergence of the coefficients in (8.12) has to hold in L_2. For this reason we are allowed to construct $(A_1^*, \dots, A_K^*, b^*)$ according to $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ on a joint probability space, i.e., (8.19) means

$$\ell_2(\mathcal{L}(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}), \mathcal{L}(A_1^*, \dots, A_K^*, b^*)) \to 0. \tag{8.22}$$

Condition (8.20), by Jensen's inequality, is stronger than the contraction condition

$$\left\|\mathbb{E}\left(\sum_{r=1}^K (A_r^*)^{\mathrm{T}} A_r^*\right)\right\|_{\mathrm{op}} < 1 \tag{8.23}$$

from Lemma 8.5. Whether condition (8.20) in Theorem 8.6 can be replaced by the weaker condition (8.23) is unknown (compare with the discussion in [163]).

Condition (8.21) is a technical condition, which is usually easy to verify in application. In Section 8.1 we have not made explicit use of this condition but have checked the convergence $X_n \xrightarrow{d} X$ directly (see Lemma 8.3).

For the application of Theorem 8.6 to recursive sequences (Y_n) as in (8.7) one has to note, that for the scaling in (8.8) we have to choose $M_n = \mathbb{E}Y_n$ in order to guarantee the conditions $\mathbb{E}X_n = 0$ and $\mathbb{E}b^{(n)} = 0$. Since, on the other hand $b^{(n)}$ in (8.11) contains the quantities M_n and in (8.19) we need to derive a limit for $b^{(n)}$, this implies that for the application of Theorem 8.6 an asymptotic expansion of the mean $\mathbb{E}Y_n$ has to be known. In contrast, the covariance matrix $\mathbb{C}\mathrm{ov}(Y_n)$ can be guessed in its first order asymptotic expansion such that Theorem 8.6 applies. Since convergence in ℓ_2 implies convergence of the second moment, Theorem 8.6 then automatically implies an asymptotic expansion of the covariance matrix $\mathbb{C}\mathrm{ov}(Y_n)$.

Proof. We already observed that Jensen's inequality (8.20) implies (8.23), that is, we have a contraction. Namely, by the definition of $b^{(n)}$ we have $\mathbb{E}\, b^{(n)} = 0$. Thus, the L_2-convergence of $(b^{(n)})$ implies $\mathbb{E}\, b^* = 0$. Therefore we can apply Lemma 8.5 so that there is a unique distributional fixed-point X of map T in (8.14). Let $X_n^{(r)} \stackrel{\mathrm{d}}{=} X_n$, $X^{(r)} \stackrel{\mathrm{d}}{=} X$ so that $(X_n^{(r)}, X^{(r)})$ are optimal couplings of (X_n, X) for all $n \in \mathbb{N}$ and $r = 1, \dots, K$ and that $(A_1, \dots, A_K, b_n, I^{(n)})$, $(X_n^{(1)}, X^{(1)}), \dots, (X_n^{(K)}, X^{(K)})$ are independent.

In a first step we derive an estimate of $\ell_2^2(X_n, X)$ in terms of $\ell_2^2(X_0, X), \dots,$ $\ell_2^2(X_{n-1}, X)$. This inequality will be sufficient to deduce that $\ell_2(X_n, X) \to 0$. From (8.9) and (8.14) and from the independence conditions it follows that (for $n \geq n_0$)

$$\begin{aligned}\ell_2^2(X_n, X) &\leq \left\| \sum_{r=1}^{K} \left(A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right) + b^{(n)} - b^* \right\|^2 \\ &= \sum_{r=1}^{K} \left\| A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right\|^2 + \left\| b^{(n)} - b^* \right\|^2 .\end{aligned}$$

By assumption we have $\|b^{(n)} - b^*\| \xrightarrow{d} 0$, so we only have to take care of the first term:

$$\begin{aligned}&\sum_{r=1}^{K} \left\| A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right\|^2 \\ &\quad = \sum_{r=1}^{K} \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) + \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|^2 \\ &\quad = \sum_{r=1}^{K} \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|^2 + \sum_{r=1}^{K} \left\| \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|^2 \qquad (8.24) \\ &\qquad + 2 \sum_{r=1}^{K} \left(A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right)^{\mathrm{T}} \cdot \left(\left(A_r^{(n)} - A_r^* \right) X^{(r)} \right) .\end{aligned}$$

By (8.19), independence, and $\|X\|_2 < \infty$ it follows that, as $n \to \infty$, and for $r = 1, \dots, K$,

$$\left\| \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|^2 \xrightarrow{d} 0.$$

For the first term in (8.24) we have the estimate

$$\begin{aligned}
&\left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|^2 \\
&\quad = \mathbb{E} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right)^{\mathrm{T}} \cdot \left((A_r^{(n)})^{\mathrm{T}} A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right) \\
&\quad \leq \mathbb{E} \left(\left\| (A_r^{(n)})^{\mathrm{T}} A_r^{(n)} \right\|_{\mathrm{op}}^2 \cdot \left\| X_{I_r^{(n)}}^{(r)} - X^{(r)} \right\| \right).
\end{aligned}$$

Since the operator norm is a Lipschitz continuous map and by the L_2 convergence of $(A_r^{(n)})$, we have $\mathbb{E} \left\| A_r^{(n)} - A_r^* \right\|_{\mathrm{op}}^2 \to 0$ and consequently

$$\lim_{n \to \infty} \mathbb{E} \left\| (A_r^{(n)})^{\mathrm{T}} A_r^{(n)} \right\|_{\mathrm{op}}^2 = \mathbb{E} \left\| (A_r^*)^{\mathrm{T}} A_r^* \right\|_{\mathrm{op}}^2$$

for $r = 1, \ldots, K$.

For the third term in (8.24) we use the estimate

$$\begin{aligned}
&\mathbb{E} \left(A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right)^{\mathrm{T}} \cdot \left(\left(A_r^{(n)} - A_r^* \right) X^{(r)} \right) \\
&\quad \leq \mathbb{E} \left(\left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\| \cdot \left\| \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\| \right) \\
&\quad \leq \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2 \cdot \left\| \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|_2 \\
&\quad = o \left(\left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2 \right) \\
&\quad = o \left(\left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2^2 \right) + o(1) \\
&\quad = \eta_{r,n} \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2^2 + o(1),
\end{aligned}$$

where $\eta_{r,n} \to 0$ as $n \to \infty$. Putting things together and by using the notation

$$M_{r,n} = \left\| (A_r^*)^{\mathrm{T}} A_r^* \right\|_{\mathrm{op}}^2 + \eta_{r,n}$$

we get

$$\begin{aligned}
\ell_2^2(X_n, X) &\leq \sum_{r=1}^{K} \mathbb{E} \left(M_{r,n} \left\| X_{I_r^{(n)}}^{(r)} - X^{(r)} \right\|^2 \right) + o(1) \\
&= \sum_{r=1}^{K} \mathbb{E} \left(\sum_{i=0}^{n} \mathbf{1}_{\{I_r^{(n)} = i\}} M_{r,n} \left\| X_i^{(r)} - X^{(r)} \right\|^2 \right) + o(1) \\
&= \sum_{i=0}^{n} \left(\sum_{r=1}^{K} \mathbb{E} \left(\mathbf{1}_{\{I_r^{(n)} = i\}} M_{r,n} \right) \right) \ell_2^2(X_i, X) + o(1).
\end{aligned}$$

By assumption there exist n_0 and $\xi < 1$ such that

$$\sum_{r=1}^{K} \mathbb{E}\, M_{r,n} = \sum_{r=1}^{K} \mathbb{E}\, \left\|(A_r^*)^{\mathrm{T}} A_r^*\right\|_{\mathrm{op}}^2 + \sum_{r=1}^{K} \eta_{r,n} \leq \xi$$

for $n \geq n_0$. Thus, with $a_n = \ell_2^2(X_n, X)$ and $A_n = \max_{n_0 \leq j \leq n} a_j$ we obtain for $n \geq n_0$ and some constant $R > 0$

$$a_n \leq \xi A_n + R.$$

Since A_n is monotone, this also implies $A_n \leq \xi A_n + R$ and consequently

$$A_n \leq \frac{R}{1-\xi} = O(1).$$

Hence, the sequence a_n is bounded. Set $\overline{A} = \sup_{n \geq 0} a_n < \infty$ and $L = \limsup_{n\to\infty} a_n$. Then for every $\epsilon > 0$ there exists $n_1 \geq n_0$ with $a_n \leq L + \epsilon$ for $n \geq n_1$. Now for $n \geq n_1$ we obtain (as above and by using (8.21))

$$a_n \leq \sum_{r=1}^{K} \mathbb{E}\left(\mathbf{1}_{\{I_r^{(n)} \leq \ell\}} \left\|(A_r^{(n)})^{\mathrm{T}} A_r^{(n)}\right\|_{\mathrm{op}} + \eta_{r,n} \right) \overline{A} + \xi(L+\epsilon) + o(1)$$

and consequently

$$L \leq \xi(L+\epsilon).$$

If $L > 0$, this is a contradiction for $0 < \epsilon < L(1-\xi)/\xi$. Hence, $L = 0$. This completes the proof of the theorem.

8.2.3 Applications of the L_2 Setting

We apply Theorem 8.6 to some examples from computer science and related areas in detail.

Quicksort

We already discussed the number Y_n of key comparisons needed by the sorting algorithm Quicksort when applied to a uniform random permutation of length n. The recurrence (8.7) applies here with $d = 1$, $K = 2$, $A_1(n) = A_2(n) = 1$, $I_1^{(n)}$ uniformly distributed on $\{0, \ldots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$ and $b_n = n - 1$. Recall that we have to work in the space $\mathcal{M}_2(0)$, that is, we have to know something of the first moment in order to reduce (8.7) to (8.9). This is provided by (8.3). It then is natural to normalise by the $1/n$:

$$X_n = \frac{Y_n - \mathbb{E} Y_n}{n}.$$

As we have observed in Section 8.1 this leads to the recurrence (8.5) that is of the form (8.9). The limit equation for X is given by (8.4). Thus, $A_1^* = U$, $A_2^* = 1 - U$ and $b^* = 1 + 2U \log U + 2(1-U)\log(1-U)$. The conditions of Theorem 8.6 can be checked directly. For example, the Lipschitz constant can be estimated by

$$\left(\mathbb{E}\|(A_1^*)^2\|_{\mathrm{op}} + \mathbb{E}\|(A_2^*)^2\|_{\mathrm{op}}\right)^{1/2} = \left(2\int_0^1 u^2\,du\right)^{1/2} = \sqrt{\frac{2}{3}}.$$

Thus, everything works precisely as in Section 8.1.

We apply the general concept to a slightly different parameter, the number of key exchanges K_n. Of course, key exchanges are only necessary if the keys are not in the right order. We again get a recurrence of the form (8.7), where the parameters d, K, $A_1(n)$, $A_2(n)$, $I^{(n)}$ are given as for the key comparisons, however b_n now depends on $I^{(n)}$ as

$$\mathbb{P}(b_n = j \mid I_1^{(n)} = k) = \frac{\binom{k}{j}\binom{n-1-k}{j}}{\binom{n-1}{k}}, \quad 0 \le j \le k < n,$$

that is, the probability that there are exactly j exchanges when the rank of the pivot element is $k+1$ (compare also with Sedgewick [190, p. 55] and Hwang and Neininger [107, section 6]). We have

$$\mathbb{E}\,K_n = \frac{n+1}{3}H_n - \frac{7}{9}n + \frac{1}{18}, \qquad (n \ge 2).$$

The limiting fixed point equation for the normalised random variable $(K_n - \mathbb{E}\,K_n)/n$ equals

$$X \stackrel{\mathrm{d}}{=} UX^{(1)} + (1-U)X^{(2)} + U(1-U) + \frac{1}{3}U\log U + \frac{1}{3}(1-U)\log(1-U).$$

Again all conditions are satisfied, the only difference is in the term b^*.

Wiener Index of Random Binary Search Trees

The Wiener index of a connected graph is the sum of the distances between all pairs of nodes in the graph, where the distance is the minimal number of edges connecting the nodes in the graph. The Wiener index has its origin in mathematical chemistry but has independently been investigated in graph theory (see [97, 55]). The Wiener index W_n of a random binary search tree does not satisfy the recurrence (8.7) with dimension $d = 1$. However, it can be covered by (8.7) in dimension $d = 2$ as follows (see [159]). Suppose that Y_n denotes the internal path length (which is precisely the number of key comparisons in Quicksort). Then we have

$$\begin{aligned} W_n &\stackrel{\mathrm{d}}{=} W^{(1)}_{I_n} + (n-I_n)Y_{I_n} + W^{(2)}_{n-1-I_n} + (I_n+1)Y_{n-I_n-1} \\ &\quad + 2I^{(n)}(n-1-I^{(n)}) + n - 1, \\ Y_n &\stackrel{\mathrm{d}}{=} Y^{(1)}_{I_n} + Y^{(2)}_{n-1-I_n} + n - 1. \end{aligned}$$

This means, that we choose $d = 2$, $K = 2$, $I_1^{(n)}$ uniformly distributed on $\{0, \dots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$ as well as

$$A_1(n) = \begin{pmatrix} 1 & n - I_1^{(n)} \\ 0 & 1 \end{pmatrix}, \quad A_2(n) = \begin{pmatrix} 1 & n - I_2^{(n)} \\ 0 & 1 \end{pmatrix},$$

$$b_n = \begin{pmatrix} 2I_1^{(n)} I_2^{(n)} + n - 1 \\ n-1 \end{pmatrix}.$$

The expectation can be computed by

$$\begin{aligned} \mathbb{E}W_n &= 2n^2 H_n - 6n^2 + 8nH_n - 10n + 6H_n \\ &= 2n^2 \log n + (2\gamma - 6)n^2 + o(n^2). \end{aligned} \tag{8.25}$$

This leads, together with the expansion of the mean of Y_n in (8.3), to the applicability of Theorem 8.6. In particular we use the normalised random vector

$$X_n = \left(\frac{W_n - \mathbb{E}\, W_n}{n^2}, \frac{Y_n - \mathbb{E}\, Y_n}{n} \right).$$

The corresponding limit equation (8.13) and (8.14), respectively, are given by

$$A_1^* = \begin{pmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{pmatrix}, \quad A_2^* = \begin{pmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{pmatrix},$$

$$b^* = \begin{pmatrix} 6U(1-U) + 2U\log U + 2(1-U)\log(1-U) \\ 1 + 2U \log U + 2(1-U)\log(1-U) \end{pmatrix},$$

where U is uniformly distributed in $[0, 1]$.

Interestingly, the contraction method can only be applied in this case, since we have the relation $2\gamma - 6 = (2\gamma - 4) - 2$ for the constants in the linear resp. in the quadratic terms of (8.3) and (8.25).

Profile of Random Recursive Trees and Binary Search Trees

In Sections 6.3 and 6.5 we have discussed the profiles of recursive trees and binary search trees and indicated that the contraction method is a proper tool to obtain limiting relations.

Let $X_{n,k}$ denote the number of nodes of distance k to the root in random recursive trees of size n. Then $X_{n,0} = 1 - \delta_{n,0}$ (with Kronecker's δ) and

$$X_{n,k} \stackrel{\mathrm{d}}{=} X^{(1)}_{I_1^{(n)},k-1} + X^{(2)}_{I_2^{(n)},k}, \quad n \geq 1,\ 1 \leq k \leq n, \tag{8.26}$$

where $I_1^{(n)}$ is uniformly distributed on $\{0, \ldots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$ and independence properties as in (8.7).

Similarly the external profile of random binary search trees satisfies $B_{n,0} = \delta_{n,0}$ and

$$B_{n,k} \stackrel{\mathrm{d}}{=} B^{(1)}_{I_1^{(n)},k-1} + B^{(2)}_{I_2^{(n)},k-1}, \quad n \geq 1,\ 1 \leq k \leq n, \tag{8.27}$$

where $I_1^{(n)}$ and $I_2^{(n)} = n - 1 - I_1^{(n)}$ are as above.

The expectations are given by

$$\mathbb{E}\, X_{n,k} = \frac{|s_{n,k+1}|}{(n-1)!} \sim \frac{(\log n)^k}{k!\Gamma\left(\frac{k}{\log n} + 1\right)} \tag{8.28}$$

and by

$$\mathbb{E}\, B_{n,k} = \frac{2^k}{n!}|s_{n,k}| \sim \frac{(2\log n)^k}{n\, k!\Gamma\left(\frac{k}{\log n}\right)} \tag{8.29}$$

where $|s_{n,k}|$ are the (sign-less) Stirling numbers of first kind (see Section 6.1.1).

By considering the normalised profiles

$$\frac{X_{n,k}}{\mathbb{E}\, X_{n,k}} \quad \text{and} \quad \frac{B_{n,k}}{\mathbb{E}\, B_{n,k}} \tag{8.30}$$

and setting $k \sim \alpha \log n$ one observes that the recurrences (8.26) and (8.27), rewritten to the normalised profiles, stabilise to stochastic fixed point equations

$$X(\alpha) \stackrel{\mathrm{d}}{=} \alpha U^\alpha X(\alpha)^{(1)} + (1-U)^\alpha X(\alpha)^{(2)}$$

for recursive trees, respectively to

$$Y(\alpha) \stackrel{\mathrm{d}}{=} \frac{\alpha}{2} U^{\alpha-1} Y(\alpha)^{(1)} + \frac{\alpha}{2}(1-U)^{\alpha-1} Y(\alpha)^{(2)}$$

for binary search trees (see Theorems 6.18 and 6.41 and Section 6.3.3).

The first problem is to check whether these fixed point equations have unique solutions. By using the L_2 setting the Lipschitz constant in the recursive tree case is given by

$$\mathrm{lip}(T_1) \leq \frac{\alpha^2 + 1}{2\alpha + 1}$$

and for the binary search tree case by

$$\mathrm{lip}(T_2) \leq \frac{\alpha^2}{4\alpha - 2}.$$

This means that we have contraction for $0 < \alpha < 2$ in the recursive tree case and for $2 - \sqrt{2} < \alpha < 2 + \sqrt{2}$ in the binary search tree case.

Actually it can be shown that the normalised profiles (8.30) converge to $X(\alpha)$ resp. to $Y(\alpha)$ in these ranges in the L_2 sense (compare with [87]). In this setting it is impossible to do it better. There are, however, solution $X(\alpha)$ for $0 < \alpha < e$ and $Y(\alpha)$ for $\alpha_- < \alpha < \alpha_+$ (where $\alpha_- = 0.373...$ and $\alpha_+ = 4.311...$ are the solution of the equation $\alpha \log\left(\frac{2e}{\alpha}\right) = 1$) but $X(\alpha)$ and $Y(\alpha)$ have finite second moments only for $0 < \alpha < 2$ resp. for $2 - \sqrt{2} < \alpha < 2 + \sqrt{2}$ (for details see [87]).

In order to extend these limit theorems to the maximum ranges $0 < \alpha < e$ resp. $\alpha_- < \alpha < \alpha_+$ one has to use a different setting. For example, by using the Zolotarev metric ζ_s for a suitable $s > 1$ (see Section 8.3.1) it is possible to prove convergence of the normalised profiles in the ranges of interest (for details see again [87]).

8.3 Limitations of the L_2 Setting and Extensions

The L_2 setting is relatively easy but has several limitations. For example, it does not apply if (for simplicity we just consider the case $d = 1$)

$$\mathbb{E}\left(\sum_{r=1}^{K} (A_r^*)^2\right) = 1, \tag{8.31}$$

that is, the contraction condition (8.23) resp. (8.20) is not satisfied. In particular, there are many applications that lead to the limit equation

$$X \stackrel{\mathrm{d}}{=} \sum_{r=1}^{K} A_r^* X^{(r)}, \tag{8.32}$$

with

$$\sum_{r=1}^{K} (A_r^*)^2 = 1. \tag{8.33}$$

From the convolution property of the normal distribution it follows that normally distributed random variables with law $N(0, \sigma^2)$ are solutions of the limit equation (8.32) under (8.33). Thus, it seems, that one cannot prove a central limit theorem even if there is one. Since the limit equation has no unique fixed point, there is definitely no metric for which T is a contraction.

Another principal problem of the contraction method is the occurrence of degenerate limit equations. For example, if we rescale a recurrence of the form

$$Y_n \stackrel{\mathrm{d}}{=} Y_{I_n} + b_n, \qquad (n \geq n_0), \tag{8.34}$$

we might end up with the degenerate limit equation

$$X \stackrel{\mathrm{d}}{=} X. \tag{8.35}$$

This limit equation does not give any information about a potential limit distribution and the concept of the contraction method needs to be significantly extended to deal with such cases.

8.3.1 The Zolotarev Metric

In order to overcome the problem of limit equations with no unique solution one has to choose a proper subspace. For example, if we work with $\mathcal{M}_s^d(M, C)$ instead of $\mathcal{M}_s^d(M)$ then the (co-)variance is fixed, and equation (8.32) under (8.33) will have a unique limit. Furthermore one has to use another metric that makes T a contraction on that subspace.

Actually, this program was realised by Neininger and Rüschendorf [161]. They used the Zolotarev metric ζ_s (introduced by Zolotarev [209]) which is defined for random vectors X, Y in $\mathbb{R}^d$ by

$$\zeta_s(X, Y) = \sup_{f \in \mathcal{F}_s} |\mathbb{E}[f(X) - f(Y)]| \tag{8.36}$$

where $s = m + \alpha$ with $0 < \alpha \leq 1$, $m \in \mathcal{N}_0$, and

$$\mathcal{F}_s := \{f \in \mathrm{C}^m(\mathbb{R}^d, \mathbb{R}) : \|f^{(m)}(x) - f^{(m)}(y)\| \leq \|x - y\|^\alpha\},$$

denotes the space of m times continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}$ such that the m-th derivative is Hölder continuous of order α. We use the short notation $\zeta_s(\mathcal{L}(X), \mathcal{L}(Y)) = \zeta_s(X, Y)$. We have $\zeta_s(X, Y) < \infty$, if all mixed moments of orders $1, \ldots, m$ of X and Y are equal and if the s-th absolute moments of X and Y are finite. Furthermore, $(\mathcal{M}_s^d, \zeta_s)$ for $0 < s \leq 1$, $(\mathcal{M}_s^d(M), \zeta_s)$ for $1 < s \leq 2$ and $(\mathcal{M}_s^d(M, C), \zeta_s)$ for $2 < s \leq 3$ are complete metric spaces with $M \in \mathbb{R}^d$, and C being a symmetric, positive definite $d \times d$ matrix. Furthermore, the Lipschitz constant of the mapping T can be estimated by

$$\mathrm{lip}(T) \leq \mathbb{E}\left(\sum_{r=1}^{K} \|A_r^*\|_{\mathrm{op}}^s\right).$$

For example, if (8.33) is satisfied then $T : \mathcal{M}_s^1(0, 1) \to \mathcal{M}_s^1(0, 1)$ is a contraction for all $2 < s \leq 3$.

There is a direct extension of Theorem 8.6 to the Zolotarev setting with many applications to statistics of recursive algorithms and related subjects (see [161, 162]) also leading to central limit theorems under very general conditions. It is even possible to work in a suitable Banach space setting (see [68]).

8.3.2 Degenerate Limit Equations

Degenerate limit equations $X \stackrel{\mathrm{d}}{=} X$ do not give any information on the nature of the potential limiting distribution. Nevertheless, in a quite general class of problems (see [162]) it is possible to prove a central limit theorem. The idea is to keep track of the recurrence and to show that an accompanying sequence of normal distributions is close to X_n. It applies to equations of the form (8.34) under some regularity conditions on I_n and under the assumption that the variance of Y_n is of logarithmic order.

9

Planar Graphs

At first sight planar graphs and trees have nothing in common despite the trivial fact that trees are planar. Nevertheless there are strong similarities in the combinatorial and asymptotic analysis of trees and planar graphs. Planar graphs contain several (hidden) tree or tree-like structures. The tree-decomposition into cut-vertices and blocks (2-connected components) is the most prominent one. The reduction to 3-connected components is more involved but uses so-called series-parallel networks that are obtained from series-parallel extensions of trees. Thus, it is natural that properly extended tree counting techniques and analytic techniques that have been developed for trees apply. However, these extensions are not straight forward. Several graph theoretical and and also topological concepts have to be combined with *combinatorics on trees*. There are also different levels of complexity in the asymptotics analysis. From this point of view outerplanar graphs and series-parallel graphs – these are two subclasses of planar graphs that we will study first – are more tree-like than the class of all planar graphs, since the singularity structure of the corresponding generating functions is of square root type $\sqrt{R-x}$, whereas the class of all planar graphs has a dominant singularity of the form $(R-x)^{3/2}$. Geometrically this indicates that outerplanar graphs and series-parallel graphs are more or less governed by a one-dimensional topology (as trees) but the class of all planar graphs by a two-dimensional one.

In this chapter we focus on labelled planar graphs $\mathcal{R}$ and use a generating function approach. As in the case of trees we then consider the uniform random model on the subclasses $\mathcal{R}_n$ with n vertices. We will shortly call them random planar graphs. There are of course different models that could be used for a random model for planar graphs. For example, there are many results on random planar maps that we will not discuss in detail (this goes back to Tutte [30, 203, 204, 205], see also [137] and the references therein). The difference between planar maps and planar graphs is that a planar map is an already embedded planar graph. If a planar graph has several non-equivalent embeddings on the 2-sphere then each of them corresponds to different planar maps although the underlying planar graph is the same. Of course, it is also interesting to consider unlabelled planar graphs. However, this model is actually much more involved than the labelled one – at least from the combinatorial point of view. At the moment there are only some results on outerplanar unlabelled graphs (see [19]).

The counting problem of several classes of planar graphs and planar maps goes back to Tutte [30, 203, 204, 205]. Interestingly the study of random planar graphs is a recent one. Random planar graphs were introduced by Denise et al [47], and since then they have been widely studied. Several natural parameters defined on $\mathcal{R}_n$ have been studied, starting with the number of edges, which is probably the most basic one. Partial results where obtained in [22, 47, 89, 167], until it was shown by Giménez and Noy [91] that the number of edges in random planar graphs asymptotically obeys a normal limit law with linear expectation and variance. The expectation is asymptotically κn, where $\kappa \approx 2.21326$ is a well-defined analytic constant. This implies that the average degree of the vertices is $2\kappa \approx 4.42652$. McDiarmid et al showed that with high probability a planar graph has a linear number of vertices of degree k, for each $k \ge 1$.

In this chapter we present a systematic combinatorial and asymptotic study of labelled random planar graphs. As already indicated the (asymptotic) counting problem is solved by using generating functions and singularity analysis. We also focus on the degree distribution. More precisely, for every k we will determine the limiting probability d_k that a random vertex in a random planar graph has degree k. This is based on recent work by Drmota, Giménez and Noy [63, 64] and extends the results of [150]. Related results have also been obtained by Bernasconi, Panagiotou and Steger [13, 14]. Interestingly the degree distribution of labelled planar graphs is different from the degree distribution of labelled trees, where we have observed a Poisson law (see Section 3.2.1). For planar graphs the degree distribution behaves as $d_k \sim ck^{-1/2}q^k$ as $k \to \infty$ (for some constant $q < 1$; see Theorem 9.46).

9.1 Basic Notions

There are several ways to introduce planar graphs. Historically they are considered as graphs that can be embedded into the plane (or into the 2-sphere) without edge crossings. For example, Kuratowski's theorem says that planar graphs are characterised by the property that no minor[1] is isomorphic to K_5 or to $K_{3,3}$.[2]

---

[1] H is a minor of G if it can be obtained from G by contracting edges, removing edges, and removing isolated nodes.

[2] A complete graph K_n is an undirected graph with n vertices where each vertex is connected to all other vertices by an edge. A complete bipartite graph $K_{m,n}$ consists of 2 vertex sets of m and n vertices, where each vertex of the first vertex set is connected to all vertices of the second vertex set by an undirected edge.

Nevertheless, it is of special interest to look at (one of) the embedding(s) $\tilde{G}$ of a planar graph G into the plane $\mathbb{R}^2$. The complement of the embedding $\mathbb{R}^2 \setminus \tilde{G}$ consists then of several connected components, the so-called *faces* of $\tilde{G}$. For example, if G is connected then the number of vertices v, the number of edges e and the number of faces f satisfy the (Euler) relation

$$v - e + f = 2.$$

In particular this means that the number of faces is independent of the embedding.

If there is a unique embedding into the plane then one can introduce the *dual graph* G^*. The vertices of G^* are the faces of G, and two faces of G are joined by an edge (in G^*) if they have a common edge (in G). G^* is then another planar graph with a unique embedding and G^{**} is isomorphic to G.

We will also distinguish between connected, 2-connected, and 3-connected (planar) graphs. A graph is 2-connected if it is connected and one has to remove at least two vertices (and all incident edges) to disconnect it. Similarly, a graph is 3-connected if it is 2-connected and one has to remove at least three vertices to disconnect it. (There is only the triangle K_3 which has to be considered separately. It is defined to be 2- but not 3-connected.[3] Figure 9.1 shows a connected, a 2-connected and a 3-connected planar graph. 3-connected planar graphs are of special interest since Whitney's theorem says that they have a unique embedding. Thus the counting problem of 3-connected planar maps and the counting problem of 3-connected planar graphs is the same.

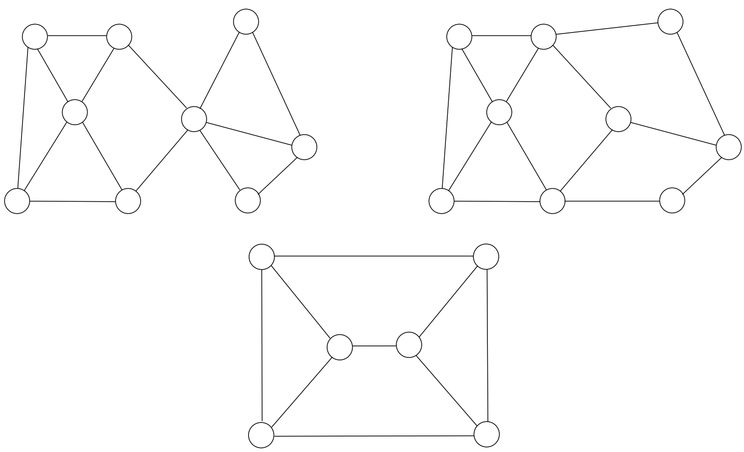


Fig. 9.1. A connected, a 2-connected and a 3-connected planar graph

[3] Generally the complete graph K_n is defined to be $(n-1)$-connected but not n-connected.

There is an intimate relation between connected, 2-connected and 3-connected planar graphs as we will see in the sequel. In fact, if one is interested in the number of (labelled) planar graphs then one should first study 3-connected graphs (or equivalently 3-connected maps), knowing that there is a general procedure to generate 2-connected ones and finally connected planar graphs.

We will also consider two subclasses of planar graphs, namely so-called outerplanar graphs and series-parallel graphs.

An outerplanar graph is a planar graph that can be embedded into the plane so that all vertices are incident to the external face. Equivalently they are characterised as those graphs not containing a minor isomorphic to K_4 or $K_{2,3}$.

A graph is series-parallel if it does not contain the complete graph K_4 as a minor. Since both K_5 and $K_{3,3}$ have a minor isomorphic to K_4, by Kuratowski's theorem a series-parallel graph is planar. Furthermore, every outerplanar graph is series-parallel. The notation *series-parallel* comes from the fact that this kind of graphs can be seen as the result of consecutive series-parallel edge extensions applied to a tree. In a series extension an edge is replaced by a chain of arbitrary finite length, whereas in a parallel extension an edge is replaced by a multiple edge. Since we are only interested in simple graphs we restrict ourselves to those graphs where the final result of this procedure has no multiple edges.

In what follows we will first solve the counting problem for labelled outerplanar, series-parallel and planar graphs. The complexity of the counting problem of these three classes increases. In particular series-parallel networks play an important role in describing the relation between 2-connected and 3-connected planar graphs.

However, the main goal of this chapter is to characterise the degree distribution of these three classes of planar graphs.

For all these problems we use (again) the concept of generating functions and singularity analysis.

9.2 Counting Planar Graphs

In this section we present the solution of the counting problem for labelled planar graphs. On the level of generating functions this is classical. However, the corresponding asymptotics are not obvious at all. In particular the cases of 2-connected and connected planar graphs have been resolved quite recently by Bender, Gao and Wormald [9] and Giménez and Noy [91].

9.2.1 Outerplanar Graphs

As explained above an outerplanar graph is a planar graph that can be embedded into the plane so that all vertices are incident to the external face.

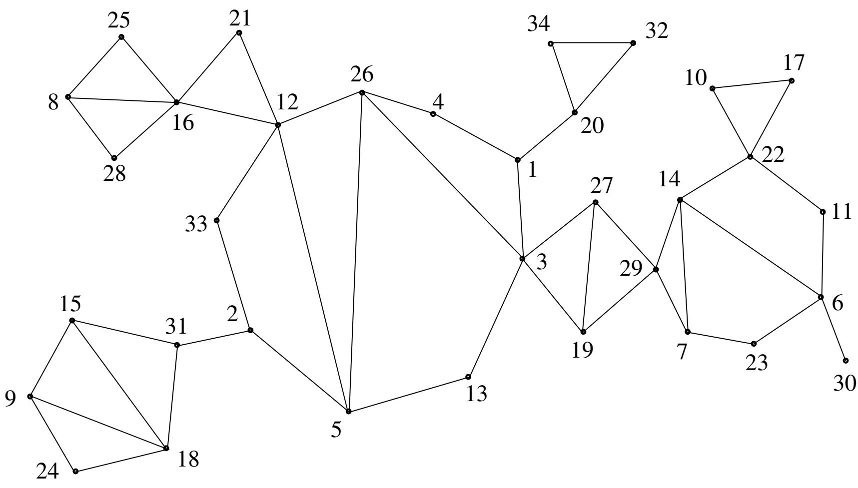


Fig. 9.2. Connected outerplanar graph

We first solve the counting problem on the level of generating functions.

Theorem 9.1. *Let b_n denote the number of 2-connected labelled outerplanar graphs, c_n the number of connected labelled outerplanar graphs and g_n the number of all labelled outerplanar graphs. Furthermore, let*

$$B(x) = \sum_{n\geq 0} b_n \frac{x^n}{n!}, \quad C(x) = \sum_{n\geq 0} c_n \frac{x^n}{n!}, \quad \textit{and} \quad G(x) = \sum_{n\geq 0} g_n \frac{x^n}{n!}$$

be the corresponding (exponential) generating functions. These functions are determined by the following system of equations:

$$G(x) = e^{C(x)}, \tag{9.1}$$

$$C'(x) = e^{B'(xC'(x))}, \tag{9.2}$$

$$B'(x) = x + \frac{1}{2}x\,A(x), \tag{9.3}$$

$$\begin{aligned} A(x) &= x(1+A(x))^2 + x(1+A(x))A(x) \qquad (9.4)\\ &= \frac{1-3x-\sqrt{1-6x+x^2}}{4x}. \end{aligned}$$

Proof. Equation (9.1) is obvious, since every (labelled) outerplanar graph uniquely decomposes into a system of (labelled) connected outerplanar graphs. Thus, the exponential generating functions are related by $G(x) = e^{C(x)}$.

The second equation (9.2) follows a general principle, too (compare with [184]). First of all note that

$$B'(x) = \sum_{n\ge 1} b_{n+1}\frac{x^n}{n!}$$

can be considered as the generating function of the 2-connected outerplanar graphs with $n+1$ vertices, where one vertex (without label) is distinguished as a root. Similarly we interprete $C'(x)$.

The right hand side

$$e^{B'(xC'(x))} = \sum_{k=0}^{\infty} \frac{1}{k!} B'(xC'(x))^k$$

can be interpreted as the exponential generating function of a finite set of rooted 2-connected outerplanar graphs, where the root vertices are identified to form a new connected rooted outerplanar graph, and every vertex different from the root is replaced by a rooted connected outerplanar graph (see Figure 9.3). It is clear that every graph of that kind constitutes a rooted connected outerplanar graph. Furthermore every rooted connected outerplanar graph G can be decomposed uniquely in the above way, as we indicate next.

Recall that a cut-vertex v of a graph G is defined by the property that the number of components of $G\setminus\{v\}$ is larger than that of G ($G\setminus\{v\}$ means that we remove v and all incident edges). Of course, a graph is 2-connected, if and only if it is connected and contains no cut-vertices. In what follows we will make use of the following reduction procedure. Let G be a connected graph and v a vertex of G and let $G_1,\ldots,G_J$ denote the connected components of $G\setminus\{v\}$. Then the subgraphs

$$G_j' = G \setminus ((G_1 \cup \cdots \cup G_{j-1} \cup G_j \cup \cdots \cup G_J) \setminus \{v\}) \quad (1 \le j \le J)$$

can be considered as subgraphs of G that can be glued together at the (cut-)point v to recover G. Furthermore, set $G_j'' = G \setminus G_j$ which is again a connected graph. Then we can also recover G by gluing G_j' and G_j'' at v. Of course, if v is not a cut-point then $J=1$, $G_1 = G\setminus\{v\}$, $G_1' = G$, and $G_1'' = \{v\}$.

Let v_{root} denote the root vertex of a connected outerplanar graph. If we delete an arbitrary vertex $v \ne v_{root}$ (in the above sense) then the graph decomposes into $k \ge 1$ components $G_1, G_2, \ldots G_J$, where we assume that the root v_{root} is contained in G_1. Without loss of generality we can assume that v is not a cut-vertex. We reduce G to G_1' and have in mind that we can recover G by gluing G_1' and G_1'' at v. We can also think of cutting off $G_2,\ldots,G_J$. Note that G_1'' is a connected graph that can be considered as rooted at v. By repeating this reduction procedure we end up with a graph $\tilde{G}$ that contains the root v_{root} and has no cut-vertex different from v_{root}. Finally we delete the root v_{root} and obtain $k \ge 1$ connected (outerplanar) graphs $\tilde{G}_1,\ldots,\tilde{G}_k$. Set $B_j = \tilde{G}_j'$, $1 \le j \le k$. Then B_j, $1\le j\le k$, is a 2-connected planar graph that is rooted at v_{root}, and $\tilde{G}$ is obtained by identifying all B_j at the root vertices.

We now restart the reduction procedure but we only cut at points $v \neq v_{root}$ that are contained in $\tilde{G}$. Then we again end up at $\tilde{G}$ and this provides the required (unique) decomposition (see Figure 9.3).

This decomposition procedure recovers an underlying tree structure of G that consists of 2-connected subgraphs which are linked at cut-vertices. Figure 9.3 can be seen from this point of view, too. Accordingly the equation $C'(x) = e^{B'(xC'(x))}$ has an interpretation in terms of simply generated trees or equivalently of Galton-Watson branching processes (see Remark 9.2).

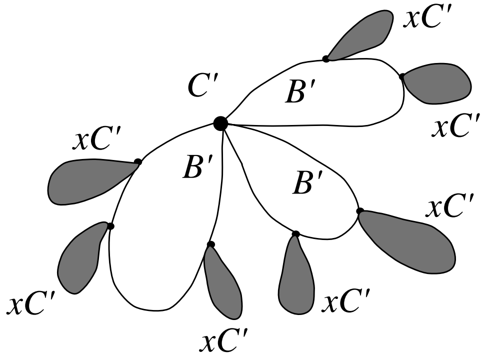


Fig. 9.3. Connection between 2-connected and connected outerplanar graphs

In order to prove the next relations (9.3) and (9.4) we introduce so-called dissections. A dissection is a convex polygon together with a set of non-crossing diagonals. In our context we will further assume that one edge of the polygon is rooted (or marked, see Figure 9.4). Alternatively we can interpret a dissection as a 2-connected outerplanar graph with a rooted edge on the infinite face. Let $\mathcal{A}$ denote the set of dissections with at least 3 vertices and let a_n, $n \geq 1$, be the number of dissections with $n+2$ vertices, that is, the vertices of the marked edge are not counted. Further, let

$$A(x) = \sum_{n \geq 1} a_n x^n$$

denote the corresponding generating function.

Dissections have a very easy recursive structure. One considers the face f that contains the rooted edge and looks at the valency ℓ of f which has to be at least 3. The dissection can be then *decomposed* into the original rooted edge plus $\ell - 1$ dissections (or just single edges) situated around f (see Figure 9.5). In terms of generating functions this reads as

$$A(x) = x(1 + A(x))^2 + x^2(1 + A(x))^3 + \cdots = \frac{x(1 + A(x))^2}{1 - x(1 + A(x))}.$$

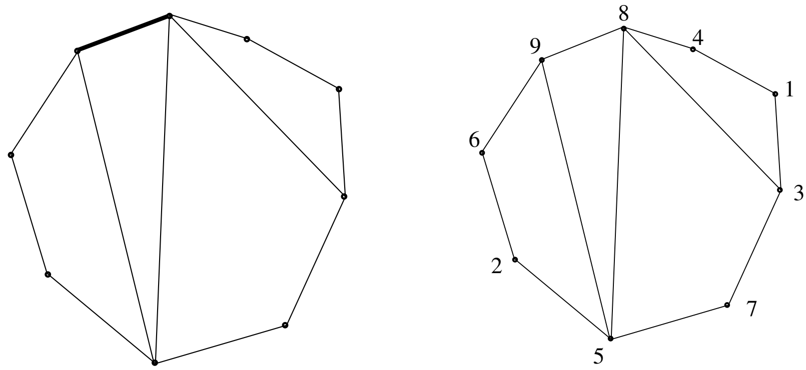


Fig. 9.4. Dissection of a convex polygon and a 2-connected outerplanar graph

Obviously this is the same equation as (9.4).

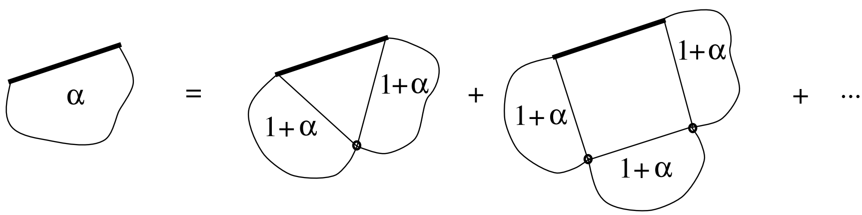


Fig. 9.5. Recursive decomposition of dissections

Finally we consider 2-connected labelled outerplanar graphs. There is exactly one 2-connected outerplanar graph with two vertices, namely a single edge. However, if $n \ge 3$ then we have

$$b_n = \frac{(n-1)!}{2} a_{n-2}.$$

First, it is clear that b_n can also be considered as the number of 2-connected outerplanar graphs with n vertices, where one vertex is marked (or rooted) and the remaining $n-1$ vertices are labelled by $1, 2, \ldots, n-1$. We just have to identify the vertex with label n with the marked vertex. Next consider a dissection with n vertices. There are a_{n-2} dissections of that kind. We mark the vertex v_1 of the root edge $e = (v_1, v_2)$ (where the vertices are numbered counter clockwise). Then there are exactly $(n-1)!$ ways to label the remaining $n-1$ vertices by $1, 2, \ldots, n-1$. Finally since the direction of the outer circle is irrelevant, we have to divide the resulting number $(n-1)! a_{n-2}$ by 2 to get back b_n. Of course, this is exactly the relation (9.4).

Remark 9.2 *Note that the above description of outerplanar graphs contains two tree-like constructions.*

First, the relation (9.2) between rooted connected graphs and rooted 2-connected corresponds to a tree-like decomposition. Set $y(x) = xC'(x)$ *and* $\Phi(x) = e^{B'(x)}$. *Then (9.2) rewrites to* $y(x) = x\Phi(y(x))$ *which is precisely the functional equation of simply generated (or Galton-Watson) trees. Figure 9.3 depicts the tree-structure at the root.*

Second, the recursive description of dissections could be alternatively given by a corresponding tree approach. We start with a planted root outside the dissection and connect it with an internal node inside f *by an edge that cuts the rooted edge. In a similar way (see Figure 9.6) we continue. For example, if the face* f *has valency* 4 *then the node inside* f *has* $3 = 4 - 1$ *subtrees that correspond to the* 3 *sub-dissections around* f.

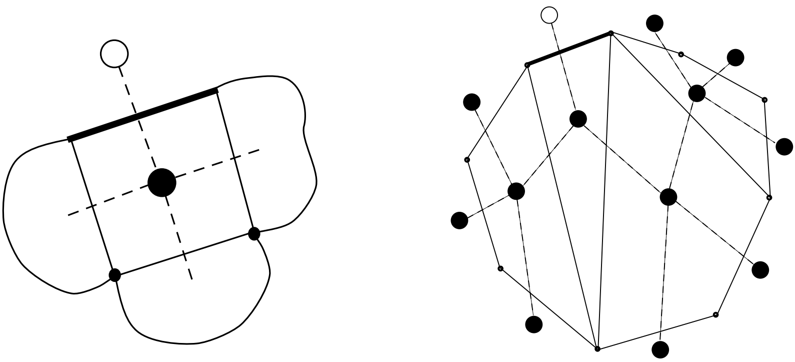


Fig. 9.6. Dissections and trees

This gives a bijection between dissections and planted plane trees, where all internal nodes have out-degree ≥ 2. *However, one has to be careful by transferring statistics between these two objects. A dissection with* r *faces and* $n+2$ *(non-rooted) edges corresponds to a tree with* r *internal nodes (different from the root) and* $n+1$ *leaves.*

By using the system of equations (9.1)–(9.4) we derive asymptotic expansions for b_n, c_n, and g_n.

Theorem 9.3. *The numbers* b_n, c_n *and* g_n *of 2-connected, connected and all labelled outerplanar graphs are asymptotically given by*

$$b_n = b \cdot n^{-\frac{5}{2}} (3 + 2\sqrt{2})^n n! \left(1 + O\left(\frac{1}{n}\right)\right),$$

$$c_n = c \cdot n^{-\frac{5}{2}} \rho^n n! \left(1 + O\left(\frac{1}{n}\right)\right),$$
$$g_n = g \cdot n^{-\frac{5}{2}} \rho^n n! \left(1 + O\left(\frac{1}{n}\right)\right),$$

where $\rho = y_0 e^{-B'(y_0)} = 0.1365937...$ *and* $y_0 = 0.1707649...$ *satisfies the equation* $1 = y_0 B''(y_0)$, *and*

$$b = \frac{1}{8\sqrt{\pi}} \sqrt{114243\sqrt{2} - 161564} = 0.000175453...,$$
$$c = 0.0069760...,$$
$$g = 0.017657...$$

are positive constants.

Proof. First note that $A(x)$ and $B'(x)$ can be explicitly computed:

$$A(x) = \frac{1 - 3x - \sqrt{1 - 6x + x^2}}{4x},$$
$$B'(x) = \frac{1 + 5x - \sqrt{1 - 6x + x^2}}{8}.$$

Both of them have a radius of convergence $\rho = 3 - 2\sqrt{2} = 0.171572875....$ Furthermore, the asymptotic expansion for b_n is immediate.

Next observe that the function $v(x) = xC'(x)$ satisfies the functional equation $v = xe^{B'(v)}$. By applying Theorem 2.19 it follows that the solution $v(x)$ gets singular at $\rho = 0.1365937...$, which is given by $\rho = v_0 e^{-B'(v_0)}$, and $v_0 = 0.1707649...$ satisfies the equation $1 = v_0 B''(v_0)$. Note that $v_0 = v(\rho) = \rho C'(\rho) < \rho_1 = 3 - 2\sqrt{2}$ which ensures that the singularity of $B'(x)$ has no influence to the singular behaviour of $C'(x)$. In fact we obtain from Theorem 2.19 that

$$[x^n]\, xC'(x) = \frac{nc_n}{n!} = c \cdot n^{-\frac{3}{2}} \rho^n \left(1 + O\left(\frac{1}{n}\right)\right),$$

where

$$c = \sqrt{\frac{1}{2\pi(B''(y_0)^2 + B'''(y_0))}} = 0.0069760...$$

Furthermore the function $xC'(x)$ has a square root type singular expansion of the form

$$xC'(x) = g(x) - h(x)\sqrt{1 - \frac{x}{\rho}},$$

where $h(\rho) = 2\sqrt{\pi} \cdot c$. Consequently we get a singular expansion for $C(x)$ of the form

$$C(x) = g_1(x) + h_1(x)\left(1 - \frac{x}{\rho}\right)^{\frac{3}{2}}$$

with $g_1(\rho) = C(\rho)$ and $h_1(\rho) = \frac{2}{3}h(\rho)$.

Finally, since $G(x) = e^{C(x)}$, the function $G(x)$ has (up to constants) the same singular behaviour as $C(x)$

$$G(x) = g_2(x) + h_2(x)\left(1 - \frac{x}{\rho}\right)^{\frac{3}{2}}$$

with $h_2(\rho) = e^{g_1(\rho)}h_1(\rho)$, which leads directly to the asymptotic expansion for g_n.

In order to evaluate the constants we note that $C(x)$ can be represented as

$$C(x) = xC'(x)\left(1 - \log(C'(x))\right) + B(xC'(x)), \tag{9.5}$$

where $B(x)$ is explicitly given by

$$B(x) = \frac{x}{8} + \frac{5}{16}x^2 - \frac{1}{32}(-6 + 2x)\sqrt{1 - 6x + x^2}$$
$$+ \frac{1}{2}\log\left(-3 + x + \sqrt{1 - 6x + x^2}\right).$$

By using these representations the constants $g_1(\rho) = C(\rho)$, $h_2(\rho) = e^{g_1(\rho)}h_1(\rho)$, and consequently $g = 0.017657...$ can be computed.

In order to prove (9.5) we set $F(x) = xC'(x)$ and get

$$C(x) = \int_0^x \frac{F(s)}{s}\,ds = F(x)\log x - \int_0^x F'(s)\log s\,ds.$$

With help of the substitution $t = F(s)$, where we also have $s = te^{-B'(t)}$, we can evaluate the last integral to

$$\begin{aligned}\int_0^x F'(s)\log s\,ds &= \int_0^{F(x)} \log\left(te^{-B'(t)}\right)\,dt\\ &= \int_0^{F(x)} \left(\log t - B'(t)\right)\,dt\\ &= F(x)\log F(x) - F(x) - B(F(x)).\end{aligned}$$

This directly provides (9.5).

Note that the above counting procedure is quite flexible. For example, we can take care of the number of edges, too. If $g_{n,m}$ denotes the number of outerplanar graphs with n vertices and m edges, then we define

$$G(x, y) = \sum_{n,m\ge 0} g_{n,m}\frac{x^n}{n!}y^m.$$

Similarly let $A(x,y)$, $B(x,y)$, $C(x,y)$ be the corresponding generating functions for dissections, 2-connected outerplanar graphs and 3-connected outerplanar graphs. Note that in the case of dissections we make the convention not to count the rooted edge. By using exactly the same counting procedure as above we get the following relations:

$$G(x,y) = e^{C(x,y)}, \tag{9.6}$$

$$\frac{\partial C(x,y)}{\partial x} = \exp\left(\frac{\partial B}{\partial x}\left(x\frac{\partial C(x,y)}{\partial x}, y\right)\right), \tag{9.7}$$

$$\frac{\partial B(x,y)}{\partial x} = xy + \frac{1}{2}xy\,A(x,y), \tag{9.8}$$

$$\begin{aligned} A(x,y) &= xy^2(1+A(x,y))^2 + xy(1+A(x,y))A(x,y) \\ &= \frac{1-xy2xy^2-\sqrt{1-2xy-4xy^2+x^2y^2}}{2xy(1+y)}. \end{aligned} \tag{9.9}$$

With the help of these equations and by precisely the same analytic considerations as above, where we consider y as an additional parameter, we get, for example, a representation for $G(x,y)$ of the form

$$G(x,y) = g_1(x,y) + h_1(x,y)\left(1-\frac{x}{\rho(y)}\right)^{\frac{3}{2}}.$$

Hence, by Theorem 2.25 we derive that the number of edges X_n in a of size n satisfies a central limit theorem. In particular we obtain

$$\mathbb{E}\,X_n = \mu n + O(1) \quad \text{and} \quad \mathbb{V}\,X_n = \sigma^2 n + O(1),$$

where $\mu = 1.56251...$ and $\sigma^2 = 0.22399...$ (compare with [20]).

9.2.2 Series-Parallel Graphs

We recall that a graph is series-parallel if it does not contain the complete graph K_4 as a minor. Equivalently a connected series-parallel graph can be seen as the result of consecutive series-parallel edge extensions applied to a tree. Thus, the basic element of a series-parallel graph is the result of a series-parallel edge extensions of a single edge. Such graphs are also called series-parallel networks. They have two distinguished vertices (or roots) that are called *poles*. The consecutive series-parallel extension induces a recursive description of series-parallel networks: they are either a parallel composition of series-parallel networks or a series decomposition of series-parallel networks or just the smallest network consisting of the two poles and an edge joining them. Figure 9.7 shows the parallel decomposition of a series-parallel network that occurs if the first step of the series-parallel edge extension was a parallel extension. In a similar way the series decomposition works (see Figure 9.8).

It is also easy to characterise 2-connected series-parallel graphs. They can be either described as the result of consecutive series-parallel edge extensions applied to an initial double edge or by a series parallel network where one additional edge is used to join the poles.

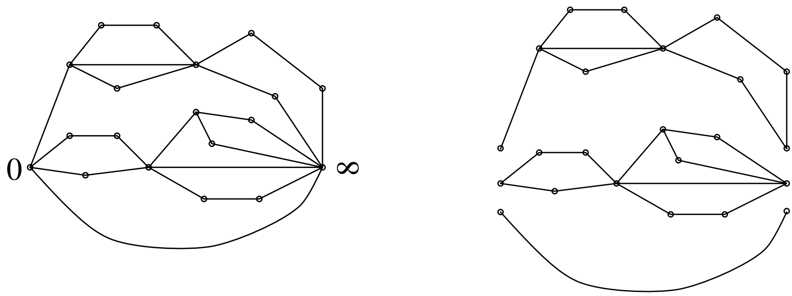


Fig. 9.7. Series-parallel network and its parallel decomposition

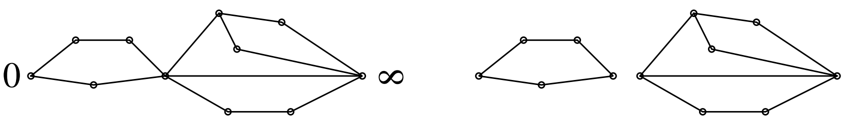


Fig. 9.8. Series decomposition of a series-parallel network

Thus, it is relatively easy to count 2-connected series-parallel graphs using generating function (see Theorem 9.4). But then a decomposition of connected graphs into 2-connected part (that works as in the outerplanar case, see also Figure 9.9) makes it possible to count connected series-parallel graphs on the level of generating functions, too. However, in contrast to outerplanar graphs we have to keep track of the number of edges from the very beginning.

Theorem 9.4. *Let $b_{n,m}$ denote the number of 2-connected labelled series-parallel graphs, $c_{n,m}$ the number of connected labelled series-parallel graphs and $g_{n,m}$ the number of all labelled series-parallel graphs with n vertices and m edges. Furthermore, let*

$$B(x,y) = \sum_{m,n\ge 0} b_{n,m} \frac{x^n}{n!} y^m, \;\; C(x,y) = \sum_{m,n\ge 0} c_{n,m} \frac{x^n}{n!} y^m$$

and

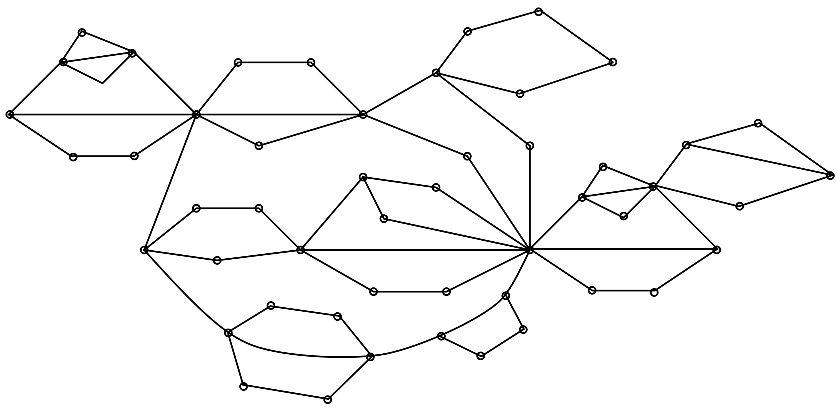


Fig. 9.9. Connected Series-parallel graph

$$G(x,y) = \sum_{m,n\geq 0} g_{n,m}\,\frac{x^n}{n!}\,y^m$$

be the corresponding generating functions. Then these functions are determined by the following system of equations:

$$G(x,y) = e^{C(x,y)} \tag{9.10}$$

$$\frac{\partial C(x,y)}{\partial x} = \exp\left(\frac{\partial B}{\partial x}\left(x\frac{\partial C(x,y)}{\partial x}, y\right)\right), \tag{9.11}$$

$$\frac{\partial B(x,y)}{\partial y} = \frac{x^2}{2}\,\frac{1+D(x,y)}{1+y} = \frac{x^2}{2}e^{S(x,y)} \tag{9.12}$$

$$D(x,y) = (1+y)e^{S(x,y)} - 1, \tag{9.13}$$

$$S(x,y) = (D(x,y) - S(x,y))xD(x,y). \tag{9.14}$$

Proof. The first two relations (9.10) and (9.11) can be proved in completely the same way as in the proof of Theorem 9.1. We only have to be aware of the number of edges (that are additive) and the fact that the tree-like decomposition of connected series-parallel graphs into 2-connected series-parallel graphs works, too.

Next we denote by

$$D(x,y) = \sum_{n,m} d_{n,m}\frac{x^n}{n!}y^m$$

the exponential generating function of all series-parallel networks, more precisely, $d_{n,m}$ is the number of series-parallel networks with $n+2$ vertices and m edges, where the n internal vertices (that are those vertices that are different from the poles) are labelled by $\{1, 2, \ldots, n\}$. In the same way we define $S(x, y)$ for series-parallel networks that have a series decomposition into at least two series-parallel networks and $P(x, y)$ for series-parallel networks that have a parallel decomposition into at least two series-parallel networks. For short, we call these networks s- and p-networks. By definition we have

$$D(x, y) = S(x, y) + P(x, y) + y.$$

Furthermore, observe that the series decomposition of a s-network (and the parallel decomposition of a p-network) is unique if we require that the components are either edges or p-networks (or at most one edge and s-networks). This unique decomposition translates directly into the system of equations

$$\begin{aligned} S(x, y) &= x(P(x, y) + y)^2 + x^2(P(x, y) + y)^3 + \cdots \qquad (9.15) \\ &= \frac{x(P(x, y) + y)^2}{1 - x(P(x, y) + y)}, \\ P(x, y) &= (e^{S(x,y)} - 1 - S(x, y)) + y(e^{S(x,y)} - 1). \qquad (9.16) \end{aligned}$$

Of course, these two equations can be rewritten to the system (9.13) and (9.14) if we use $D - S$ instead of $P + y$.

Alternatively we can interprete the equations (9.13) and (9.14). The first equation (9.13) expresses the fact that a series-parallel network is a parallel composition of series networks (this is the exponential term), to which we may add or not the edge connecting the two poles. The second equation (9.14) means that a series network is formed by taking at first a non-series network (this is the term $D - S$), and concatenating to it an arbitrary network.

Finally we have (9.12), since a series-parallel network with non-adjacent poles (which is counted by $e^{S(x,y)}$) can be obtained by distinguishing, orienting and then deleting any edge of an arbitrary 2-connected series-parallel graph. The factor x^2 takes into account the two poles, the factor $1/2$ the two possible orientations and the derivative with respect to y the choice and deletion of an arbitrary edge.

Remark 9.5 *Note that the system (9.15) and (9.16) can be interpreted as a tree-like composition with two kinds of nodes (S and P) and two different reproductions rules.*

Furthermore, we want to note that $D(x, y)$ solves a single equation

$$\log\left(\frac{1 + D(x, y)}{1 + y}\right) = \frac{xD(x, y)^2}{1 + xD(x, y)}, \qquad (9.17)$$

which will make the singularity analysis slightly easier.

Theorem 9.6. *The numbers b_n, c_n and g_n of 2-connected, connected and all series-parallel graphs are asymptotically given by*

$$b_n = b \cdot n^{-\frac{5}{2}} \rho_1^n n! \left(1 + O\left(\frac{1}{n}\right)\right),$$
$$c_n = c \cdot n^{-\frac{5}{2}} \rho_2^n n! \left(1 + O\left(\frac{1}{n}\right)\right),$$
$$g_n = g \cdot n^{-\frac{5}{2}} \rho_2^n n! \left(1 + O\left(\frac{1}{n}\right)\right),$$

where $\rho_1 = 0.1280038...$, $\rho_2 = 0.11021...$ *and*

$$b = 0.0010131...,$$
$$c = 0.0067912...,$$
$$g = 0.0076388...$$

are positive constants.

Proof. Since $D(x, y)$ satisfies a single equation (9.17), it follows that there is a singular expansion of the form

$$D(x, y) = g(x, y) - h(x, y)\sqrt{1 - \frac{x}{\rho(y)}},$$

where $\rho(1) = \rho_1 = 0.12800....$ Since $e^{S(x,y)} = (D(x, y) + 1)/(1 + y)$, we get a corresponding expansion for

$$\frac{\partial B(x, y)}{\partial y} = g_1(x, y) - h_1(x, y)\sqrt{1 - \frac{x}{\rho(y)}}.$$

By Theorem 2.30 this implies that $B(x, y)$ has a representation of the form

$$B(x, y) = g_2(x, y) + h_2(x, y)\left(1 - \frac{x}{\rho(y)}\right)^{\frac{3}{2}}. \quad (9.18)$$

Of course, by setting $y = 1$ this leads to the asymptotic expansion for b_n with $\rho_1 = \rho(1)$.

From (9.18) we also obtain a representation for

$$\frac{\partial B(x, y)}{\partial x} = g_3(x, y) - h_3(x, y)\sqrt{1 - \frac{x}{\rho(y)}},$$

where we can safely set $y = 1$. For notational simplicity we set $B'(x) = \frac{\partial B(x,1)}{\partial x}$ and $C'(x) = \frac{\partial C(x,1)}{\partial x}$. Then the function $v(x) = xC'(x)$ satisfies the functional equation $v(x) = xe^{B'(v(x))}$, that is, we are in a similar situation as in the proof of Theorem 9.3. Again we observe that there exists $v_0 = 0.1279695... < \rho_1$ with $v_0 B''(v_0) = 1$ (compare with [20]). Thus, by Theorem 2.19, the function $v(x) = xC'(x)$ has a local representation of the form

$$xC'(x) = g_4(x) - h_4(x)\sqrt{1 - \frac{x}{\rho_2}},$$

where $\rho_2 = v_0 e^{-B'(v_0)}) = 0.11021....$ Consequently we obtain corresponding representations for

$$C(x) = g_5(x) + h_5(x)\left(1 - \frac{x}{\rho_2}\right)^{\frac{3}{2}},$$

and for

$$G(x) = e^{C(x)} = g_6(x) + h_6(x)\left(1 - \frac{x}{\rho_2}\right)^{\frac{3}{2}},$$

that induce corresponding asymptotic representations for c_n and g_n.

Remark 9.7 *We want to remark that one has to be careful with the evaluation of constants since we have been working with derivatives with respect to x and y. Nevertheless it is not difficult to solve the above system of (differential) equations numerically and check all necessary conditions, for example, that there exists $v_0 < \rho_1$ with $v_0 B''(v_0) = 1$.*

However, due to the simplicity of the above equations, one can solve some of the integrals explicitly. For example, in [20] it is shown that $B(x, y)$ can be represented in terms of $D = D(x, y)$:

$$B(x, y) = \frac{1}{2}\log(1 + xD) - \frac{xD(x^2D^2 + xD + 2 - 2x)}{4(1 + xD)},$$

and $C(x, y)$ can be represented in terms of $\frac{\partial C(x,1)}{\partial x}$ and $B(x, y)$:

$$C(x, y) = x\frac{\partial C(x, 1)}{\partial x}\left(1 - \log\frac{\partial C(x, 1)}{\partial x}\right) - B\left(x\frac{\partial C(x, 1)}{\partial x}, y\right).$$

Remark 9.8 *As in the case of outerplanar graphs we also get a central limit theorem for the number X_n of edges in a random series-parallel graph of size n, with*

$$\mathbb{E}\, X_n = \mu n + O(1) \quad \textit{and} \quad \mathbb{V}\, X_n = \sigma^2 n + O(1),$$

where $\mu = 1.61673...$ and $\sigma^2 = 0.55347...$ (compare with [20]). By a slight variation of the proof of Theorem 9.6 we obtain a local representation of $G(x, y)$ of the form

$$G(x, y) = g_6(x, y) + h_6(x, y)\left(1 - \frac{x}{\rho_2(y)}\right)^{\frac{3}{2}}.$$

Thus, a direct application of Theorem 2.25 gives the result.

9.2.3 Quadrangulations and Planar Maps

Before we can count planar graphs we have to know something on the relation between quadrangulations and planar maps. A planar map is – informally spoken – an embedded planar graph, and a quadrangulation is a planar map where every face has valency 4 (also the external face).

Interestingly, there is a bijective relation between quadrangulations and 2-connected maps. Suppose a 2-connected map M is given. Let M^* denote the dual map, that is, the vertices of M^* are the faces of M and two vertices of M^* are linked by an edge (of M^*) if the corresponding faces of M have an edge in common. For simplicity the vertices of M^* are represented by points contained in the corresponding faces of M. Now we construct a quadrangulation Q in the following way. The vertices of Q are the vertices of M together with the vertices of M^* (represented as points in the plane). Finally we connect a vertex v of M with a vertex v^* of M^* by an edge (in Q) if v is contained on the boundary of the face of M corresponding to v^*. Figure 9.10 shows an example of this procedure.

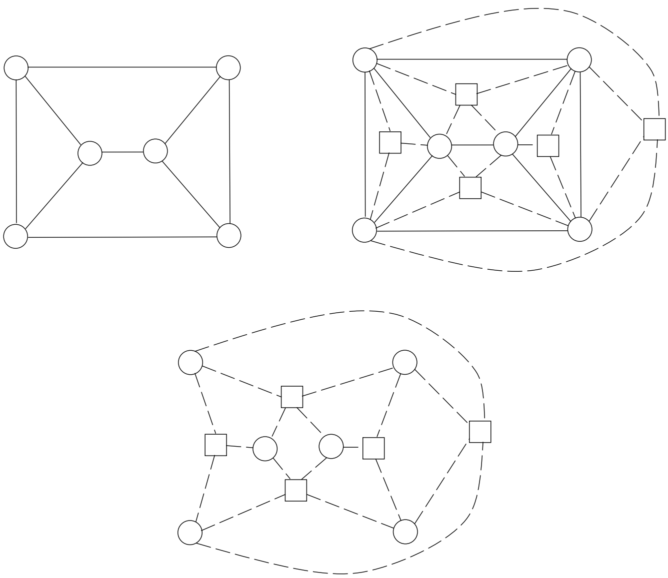


Fig. 9.10. Quadrangulations and maps

A *diagonal* in a quadrangulation is an internal path of length 2 joining two opposite vertices of the external face. A quadrangulation is *simple* if it has no diagonal, if every cycle of length 4 (other than the external cycle that encircles the whole graph) defines a face, and if it is not the trivial map reduced to a single quadrangle. For example, the quadrangulation depicted in Figure 9.10

is a simple one. It is a well known fact and an easy exercise (see [29, 156]) that the above described procedure leads to a simple quadrangulation, if and only if the corresponding map is 3-connected.

Furthermore, it is easy to see that edge-rooted maps correspond to edge rooted quadrangulations, where the rooted edge is directed and always on the external face. These kinds of objects can now be counted with the help of generating functions. Note further, that quadrangulations are bipartite graphs so that we can distinguish between black and white vertices. This is uniquely defined by assuming that the rooted (directed) edge starts from a black vertex. With this convention the above bijection transfers vertices of the 2-connected map to black vertices of the quadrangulation, edges to faces, faces to white vertices, and the root degree (of the first vertex of the root edge) stays the same.

Let $f_{i,j,k}$ be the number of edge-rooted 2-connected maps with $i+1$ vertices and $j+1$ faces where the (first) vertex of the rooted edge has degree $k+1$. Then $f_{i,j,k}$ is also the number of edge-rooted quadrangulations with $i+1$ black vertices, and $j+1$ white vertices where the (first) vertex of the rooted edge has degree $k+1$. The following lemma provides an explicit representation of the trivariate generating function of these numbers.

Lemma 9.9. *The generating function*

$$F(x,y,w) = \sum_{i,j,k} f_{i,j,k} x^i y^j w^k$$

is given by

$$\begin{aligned} F(x,y,w) &= -\frac{1}{2w}(1-(1+v-u+uv-2v^2u)w + v(1-u)^2w^2) \\ &\quad + \frac{1}{2w}(1-(1-u)w)\sqrt{1-2v(1+u+2uv)w+v^2(1-u)^2w^2}, \end{aligned}$$

where $u = u(x,y)$ and $v = v(x,y)$ are determined by the system

$$x = u(1-v)^2, \tag{9.19}$$

$$y = v(1-u)^2. \tag{9.20}$$

Remark 9.10 *Note that the function $F(x,y,1)$ is given by*

$$F(x,y,1) = uv(1-u-v). \tag{9.21}$$

This representation is due to Brown and Tutte [30]. In particular, it can be used to prove an explicit formula for

$$f_{i,j} = \sum_{k\ge 0} f_{i,j,k} = \frac{(2i+j-2)!(2j+i-2)!}{i!j!(2i-1)!(2j-1)!}.$$

For example, this can be derived from (9.21) with help of Lagrange's inversion formula (see [30]).

Interestingly there is also a bijection to two different tree classes, to so-called description trees and skew ternary trees (see [109]).

Proof. Let $w_{n,j,m}$ denote the number of 2-connected edge-rooted planar maps with n edges, and $j+1$ faces, where the external face has valency m. Furthermore set

$$w_m(x,y) = \sum_{n\ge 2}\sum_{j\ge 1} w_{n,m,j} x^n y^j,$$
$$w(x,y,z) = \sum_{m\ge 2} w_m(x,y)\, z^m = \sum_{n,j,m} w_{n,j,m} x^n y^j z^m.$$

Now consider the rooted edge e of the external face of the 2-connected map. Of course, e belongs to the external face and to another face f. Denote by γ_1 the remaining edges on the external face and by γ_2 the remaining edges on the face f. If one deletes the edge e, then $M \setminus e$ might have $k \ge 0$ cut-points $a_1, \ldots, a_k$ that are exactly common points of γ_1 and γ_2 (different from the two vertices incident to the rooted edge e, see Figure 9.11).

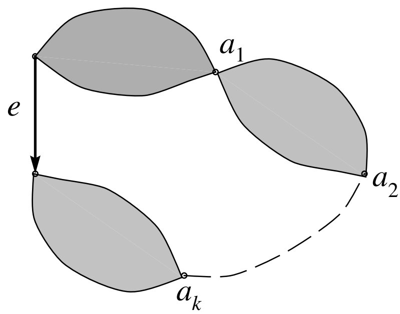


Fig. 9.11. 2-connected maps

A careful look at this recursive structure leads to the relation

$$w(x,y,z) = xyz \sum_{k\ge 0} \left(\sum_{m\ge 2} w_m(x,y)(z + z^2 + \cdots z^{m-1}) + xz \right)^{k+1}$$
$$= xyz \frac{\sum_{m\ge 2} w_m(x,y)(z+z^2+\cdots z^{m-1}) + xz}{1 - \sum_{m\ge 2} w_m(x,y)(z+z^2+\cdots z^{m-1}) - xz}.$$

This rewrites to

$$w(x,y,z)^2 + ((1-z)(1-xz) + xyz - zw(x,y,1))w(x,y,z) \\ -xz^2y(x(1-z) + w(x,y,1)) = 0.$$

Now let $h_{i,j,m}$ be the number of edge-rooted 2-connected planar maps with $i+1$ vertices, and $j+1$ faces with an external face of valency m (where we exclude the map that consists just of one edge). Then by Euler's relation we have

$$h_{i,j,m} = w_{i+j,j,m}$$

and consequently the corresponding generating function $h(x,y,z)$ is given by

$$h(x,y,z) = \sum_{i,j,m} h_{i,j,m} x^i y^j z^m = w(x,y/x,z)$$

and satisfies the relation

$$h(x,y,z)^2 + ((1-z)(1-xz) + yz - zh(x,y,1))h(x,y,z) \\ -yz^2(x(1-z) + h(x,y,1)) = 0. \tag{9.22}$$

It is easy to check that (9.22) has a unique analytic solution if one assumes that $h(x,0,z) = 0$. Namely, if we set

$$h(x,y,z) = \sum_{j\ge 0} h_j(x,z) y^j$$

(with $h_0(x,z) = h(x,0,z) = 0$) then (9.22) rewrites to a recurrence for $h_j(x,z)$. Now observe that

$$h(x,y,z) = -\frac{1}{2}(1 - (1+u-v+uv-2u^2v)z + u(1-v)^2z^2) \\ + \frac{1}{2}(1-(1-v)z)\sqrt{1-2u(1+v+2uv)z + u^2(1-v)^2z^2},$$

where $u = u(x,y)$ and $v = v(x,y)$ are defined by (9.19) and (9.20) is actually an analytic solution of (9.22) that meets these conditions.

Finally we apply the above enumeration result to the dual maps. Here vertices are transferred to faces, faces to vertices, and the valency of the outer face to the degree of the degree of the root vertex. As a consequence we have that $h(x,y,w) = wF(x,y,w)$, where the extra factor w appears, because in F we are counting the degree of the root vertex minus one. Of course, this completes the proof of the lemma.

We now turn to simple triangulations or equivalently to 3-connected edge-rooted maps. Let $q_{i,j,k}$ denote the number of 3-connected edge-rooted maps with $i+1$ vertices and $j+1$ faces, where the (first) vertex of the root edge has degree $k+1$. Equivalently this is the number of simple quadrangulations with $i+1$ black vertices and $j+1$ white vertices, where the (first) vertex of the root edge has degree $k+1$.

Lemma 9.11. *The generating function*

$$Q(x,y,w) = \sum_{i,j,k\geq 1} q_{i,j,k} x^i y^j w^k \tag{9.23}$$

is given by

$$Q(x,y,w) = xyw\left(\frac{1}{1+x} + \frac{1}{1+wy} - 1\right) - \frac{rs}{(1+r+s)^3}\frac{-w_1(r,s,w) + (r-w+1)\sqrt{w_2(r,s,w)}}{2(s+1)^2(sw+r^2+2r+1)},$$

where $w_1(r,s,w)$ *and* $w_2(r,s,w)$ *are polynomials given by*

$$\begin{aligned} w_1(x,r,s) =& -rsw^2 + w(1+4s+3rs^2+5s^2+r^2+2r+2s^3+3r^2s+7rs) \\ &+ (r+1)^2(r+2s+1+s^2), \end{aligned} \tag{9.24}$$

$$\begin{aligned} w_2(x,r,s) =& r^2s^2w^2 - 2wrs(2r^2s+6rs+2s^3+3rs^2+5s^2+r^2+2r+4s+1) \\ &+ (r+1)^2(r+2s+1+s^2)^2 \end{aligned} \tag{9.25}$$

and $r = r(x,y)$, *and* $s = s(x,y)$ *are determined by the system*

$$r = x(s+1)^2, \tag{9.26}$$

$$s = y(r+1)^2. \tag{9.27}$$

Observe, too, that the function $Q(x,y,1)$ is given by

$$Q(x,y,1) = xy\left(\frac{1}{1+x} + \frac{1}{1+y} - 1\right) - \frac{rs}{(r+s+1)^3}.$$

This formula is due to Mullin and Schellenberg [156]. In passing we also obtain a representation of the function $M(x,y)$ that counts the number of 3-connected edge-rooted planar maps according to the (total) number of vertices and edges:

$$M(x,y) = x^2y^2\left(\frac{1}{1+xy} + \frac{1}{1+y} - 1 - \frac{(1+U)^2(1+V)^2}{(1+U+V)^3}\right), \tag{9.28}$$

where $U = U(x,y)$ and $V = V(x,y)$ are defined by $U = xy(1+V)^2$ and $V = y(1+U)^2$.

Proof. Let $F(x,y,w)$ be the generating function used in Lemma 9.9. However, we now use it as the generating function of rooted quadrangulations, where the variables x, y and w mark, respectively, the number of black vertices minus one, the number of white vertices minus one, and the degree of the first vertex of the rooted edge minus one. We recall that a diagonal is an internal path of length 2 joining two opposite vertices of the external face. We will call a diagonal black or white, according to the colour of the external vertices they join.

We associate generating functions F_N, F_B and F_W, respectively, to quadrangulations with no diagonal, to those with at least one black diagonal (at the root vertex), and to those with at least one white diagonal (not at the root vertex). By planarity only one of the two kinds of diagonals can appear in a quadrangulation; it follows that

$$F(x,y,w) = F_N(x,y,w) + F_B(x,y,w) + F_W(x,y,w).$$

A quadrangulation with a diagonal can be decomposed into two quadrangulations, by considering the maps to the left and to the right of this diagonal. In order to make this decomposition unique we assume that we decompose at the leftmost diagonal when there are several of them. This gives raise to the equations

$$\begin{aligned} F_B(x,y,w) &= (F_N(x,y,w) + F_W(x,y,w))\,\frac{F(x,y,w)}{x}, \\ F_W(x,y,w) &= (F_N(x,y,w) + F_B(x,y,w))\,\frac{F(x,y,1)}{y}. \end{aligned}$$

In the second case, only one of the two quadrangulations contributes to the degree of the root vertex; this is the reason why the term $F(x,y,1)$ appears. The x and the y in the denominators appear because the three vertices of the diagonal are common to the two quadrangulations. Since we are considering vertices minus one, we only need to correct the colour that appears twice at the diagonal. Incidentally, no term w appears in the equations for the same reason.

Let us write $F = F(x,y,w)$ and $F(1) = F(x,y,1)$. From the previous equations we deduce that

$$\begin{aligned} F = F_N + F_B + F_W &= (F_N + F_B)(1 + \frac{F}{x}), \\ F = F_N + F_B + F_W &= (F_N + F_W)(1 + \frac{F(1)}{y}), \end{aligned}$$

so that

$$F + F_N = (F_N + F_B) + (F_N + F_W) = F\left(\frac{1}{1+\frac{F}{x}} + \frac{1}{1+\frac{F(1)}{y}}\right),$$

and finally

$$F_N = F\left(\frac{1}{1+\frac{F}{x}} + \frac{1}{1+\frac{F(1)}{y}} - 1\right). \tag{9.29}$$

Now we proceed to count simple quadrangulations. We use the following combinatorial decomposition of quadrangulations with no diagonals, in terms of simple quadrangulations: all quadrangulations with no diagonals, with the only exception of the trivial one, can be decomposed uniquely into a simple quadrangulation q and as many quadrangulations as internal faces q has. More precisely, we replace each internal face f of a simple triangulation by an arbitrary quadrangulation q_f by identifying the external edges of q_f with the edges of f.

Next we translate the combinatorial decomposition of simple quadrangulations into generating functions as follows.

$$\begin{aligned} F_N(x,y,w) - xyw &= \sum_{i,j,k} q_{i,j,k} x^i y^j \left(\frac{F}{xy}\right)^k \left(\frac{F(1)}{xy}\right)^{i+j-1-k} \\ &= \sum_{i,j,k} q_{i,j,k} \frac{xy}{F(1)} \left(\frac{F(1)}{y}\right)^i \left(\frac{F(1)}{x}\right)^j \left(\frac{F}{F(1)}\right)^k = \\ &= \frac{xy}{F(1)} Q\left(\frac{F(1)}{y}, \frac{F(1)}{x}, \frac{F}{F(1)}\right), \end{aligned} \tag{9.30}$$

where we are using the fact that a quadrangulation counted by $q_{i,j,k}$ has $i+j+2$ vertices, $i+j-1$ internal faces, and k of them are incident to the root vertex.

At this point we change variables as $X = F(1)/y$, $Y = F(1)/x$ and $W = F/F(1)$. Then the equations (9.29) and (9.30) can be rewritten as

$$\frac{xy}{F(1)} Q(X,Y,Z) = F_N - xyw = F\left(\frac{1}{1+\frac{F}{x}} + \frac{1}{1+\frac{F(1)}{y}} - 1\right) - xyw,$$

$$Q(X,Y,W) = XYW\left(\frac{1}{1+WY} + \frac{1}{1+X} - 1\right) - F(1)w. \tag{9.31}$$

The last equation would be an explicit expression of Q in terms of X, Y, W, if there was not the term $F(1)w = F(x,y,1)w$.

Note that $F(1) = F(x,y,1) = uv(1-u-v)$ was already determined. When we substitute

$$R = \frac{u}{1-u-v} \quad \text{and} \quad S = \frac{v}{1-u-v}$$

then we have

$$F(1) = \frac{RS}{(1+R+S)^3}, \tag{9.32}$$

where $R = R(X,Y)$ and $S(X,Y)$ are algebraic functions defined by

$$R = X(S+1)^2, \qquad S = Y(R+1)^2. \tag{9.33}$$

Recall that we have also substituted $X = F(1)/y$, $Y = F(1)/x$.

Hence it remains only to obtain an expression for $w = w(X, Y, W)$ to get an explicit expression for Q.

Recall that we have the $h(y, x, w) = wF(x, y, w)$ (see the proof of Lemma 9.9). Hence $F = F(x, y, w)$ and $F(1) = F(x, y, 1)$ satisfy the equation

$$(1-w)(1-yw)wF = -w^2F^2 + (-xw + wF(1))wF + xw^2(y(1-w) + F(1)).$$

By dividing both sides by $F(1)^2$, and rewriting in terms of $X = F(1)/y$, $Y = F(1)/x$ and $W = F/F(1)$, we obtain

$$(1-w)\left(\frac{1}{F(1)} - \frac{w}{X}\right) wW = -w^2W^2 + \left(1 - \frac{1}{Y}\right) w^2W + \frac{w^2}{Y}\left(\frac{1}{X}(1-w) + 1\right),$$

$$\frac{(1-w)(X - wF(1))wW}{XF(1)} = \frac{w^2(-XYW^2 + XYW - XW + 1 - w + X}{XY},$$

$$Y(1-w)(X - wF(1))W = wF(1)(-XYW^2 + XYW - XW + 1 - w + X). \quad (9.34)$$

Observe that this is a quadratic equation in w. Solving for w in (9.34) and using (9.32) and (9.33) we get

$$w = \frac{-w_1(R, S, W) + (R - W + 1)\sqrt{w_2(R, S, W)}}{2(S+1)^2(SW + R^2 + 2R + 1)}, \quad (9.35)$$

where $w_1(R, S, W)$ and $w_2(R, S, W)$ are those polynomials stated in (9.24) and (9.25).

Thus, together with equations (9.31) and (9.32), we have finally obtained an explicit expression for the generating function $Q(X, Y, W)$ of simple quadrangulations in terms of W and algebraic functions $R(X, Y)$ and $S(X, Y)$.

9.2.4 Planar Graphs

The counting problem for labelled planar graphs is the most complex one. We have to use results on 3-connected maps and network constructions as we will see next.

Theorem 9.12. *Let $b_{n,m}$ denote the number of 2-connected labelled planar graphs, $c_{n,m}$ the number of connected labelled planar graphs and $g_{n,m}$ the number of all labelled planar graphs with n vertices and m edges. Furthermore, let*

$$B(x, y) = \sum_{m,n \geq 0} b_{n,m} \frac{x^n}{n!} y^m, \; C(x, y) = \sum_{m,n \geq 0} c_{n,m} \frac{x^n}{n!} y^m,$$

and

$$G(x, y) = \sum_{m,n \geq 0} g_{n,m} \frac{x^n}{n!} y^m$$

the corresponding generating functions. Then these functions are determined by the following system of equations:

$$G(x,y) = \exp\left(C(x,y)\right), \tag{9.36}$$

$$\frac{\partial C(x,y)}{\partial x} = \exp\left(\frac{\partial B}{\partial x}\left(x\frac{\partial C(x,y)}{\partial x}, y\right)\right), \tag{9.37}$$

$$\frac{\partial B(x,y)}{\partial y} = \frac{x^2}{2}\,\frac{1+D(x,y)}{1+y}, \tag{9.38}$$

$$\frac{M(x,D(x,y))}{2x^2D(x,y)} = \log\left(\frac{1+D(x,y)}{1+y}\right) - \frac{xD(x,y)^2}{1+xD(x,y)}, \tag{9.39}$$

$$M(x,y) = x^2y^2\left(\frac{1}{1+xy} + \frac{1}{1+y} - 1 - \frac{(1+U(x,y))^2(1+V(x,y))^2}{(1+U(x,y)+V(x,y))^3}\right), \tag{9.40}$$

$$U(x,y) = xy(1+V(x,y))^2, \tag{9.41}$$

$$V(x,y) = y(1+U(x,y))^2. \tag{9.42}$$

Proof. As in the previous cases of outerplanar and series-parallel graphs the first two relations (9.36) and (9.37) follow general principles.

Furthermore we already know that $M(x,y)$ denotes the generating function that counts the number of 3-connected edge-rooted planar maps according to the (total) number of vertices and edges (compare with (9.28)). Thus, it remains to explain the relations (9.38) and (9.39) that provide a relation between 2- and 3-connected planar graphs. Here we follow the arguments given by Walsh [207] and Bender, Gao and Wormald [9].

First of all, by Whitney's theorem every 3-connected planar graph has a unique embedding into the plane. In particular, if we take into account the rooted edge and its direction, we thus obtain that the (exponential) generating function

$$T^\bullet(x,y) = \frac{1}{2}M(x,y)$$

counts 3-connected labelled edge-rooted planar graphs. More precisely, every rooted planar map with n vertices corresponds to $n!$ labelled rooted graphs and there are precisely two ways of rooting an embedding of a directed edge-rooted graph in order to get a rooted map (compare with [9]).

Next we introduce so-called *networks* that extend the concept of series-parallel networks. A network N is a (multi-)graph with two distinguished vertices, called its poles (usually labelled 0 and ∞) such that the (multi-)graph $\hat{N}$ obtained from N by adding an edge between the poles of N is 2-connected.

Let M be a network and $X = (N_e, e \in E(M))$ a system of networks indexed by the edge-set $E(M)$ of M. Then $N = M(X)$ is called the superposition with core M and components N_e and is obtained by replacing all edges $e \in E(M)$ by the corresponding network N_e (and, of course, by identifying the poles of N_e with the end vertices of e accordingly). A network N is called an h-network if it can be represented by $N = M(X)$, where the core M has the property that the graph $\hat{M}$ obtained by adding an edge joining the poles is 3-connected and has at least 4 vertices. Similarly $N = M(X)$ is called a p-network if M consists of 2 or more edges that connect the poles, and it is called an s-network if M consists of 2 or more edges that connect the poles in series.

Now Trakhtenbrot's canonical network decomposition theorem [202] says that any network with at least 2 edges belongs to exactly one of the 3 classes of h-, p- or s-networks. Furthermore, any h-network has a unique decomposition of the form $N = M(x)$, and a p-network (or any s-network) can be uniquely decomposed into components which are not themselves p-networks (or s-networks).

Trakhtenbrot's theorem was formulated for general networks but it also applies to planar networks, since any composition $N = M(X)$ of planar networks $X = (N_e, e \in E(M))$ is planar again. Furthermore, we can also adapt it to a counting procedure for graphs without multi-edges (compare with [207]).

We introduce the following generating functions, where the exponent of x counts the number of vertices and the exponent of y the number of edges. Of course, the generating functions are exponential in x, since the vertices are labelled. Let $K(x, y)$ be the generating function corresponding to all planar networks where the two poles are not connected by an edge (and where the two poles are not counted), $D(x, y)$ the generating function corresponding to all planar networks with at least one edge, $S(x, y)$ the generating function corresponding to all s-networks, $F(x, y) = D(x, y) - S(x, y)$ the generating function corresponding to all non-s-networks (with at least one edge), and $N(x, y)$ the generating function corresponding to all non-p-networks.

These generating functions satisfy the following relations:

$$\frac{\partial B(x, y)}{\partial y} = \frac{x^2}{2} K(x, y), \tag{9.43}$$

$$D(x, y) = (1 + y)K(x, y) - 1, \tag{9.44}$$

$$K(x, y) = e^{N(x,y)}, \tag{9.45}$$

$$S(x, y) = xD(x, y)(D(x, y) - S(x, y)), \tag{9.46}$$

$$\frac{T^\bullet(x, D(x, y))}{x^2 D(x, y))} = N(x, y) - S(x, y). \tag{9.47}$$

The first relation (9.43) is just a rewriting of the definition of (planar) networks. We get a planar network (with no edge connecting the two poles) by distinguishing, orienting and then deleting any edge from a 2-connected planar graph. Obviously we have $D(x, y) = (K(x, y) - 1) + yK(x, y)$, since $K(x, y) - 1$ corresponds to all networks (with at least one edge), where the poles are not connected by an edge and $yK(x, y)$ corresponds to those, where the poles are connected by an edge. This proves (9.44). Furthermore, we have the decomposition

$$K(x,y) = 1 + \sum_{k \ge 2} \frac{1}{k!} N(x,y)^k + N(x,y).$$

The 1 corresponds to the network without edges, the sum $\sum_{k\ge 2} \frac{1}{k!} N(x,y)^k$ to all p-networks (by Trakhtenbrot's theorem) and the remaining ones are non-p-networks counted by $N(x,y)$. Of course, this gives (9.45). Next we have (again by Trakhtenbrot's theorem)

$$S(x,y) = xF(x,y)^2 + x^2 F(x,y)^3 + \cdots = \frac{xF(x,y)^2}{1 - xF(x,y)}.$$

After substituting $F(x,y) = D(x,y) - S(x,y)$ this rewrites to (9.46) which has also a natural interpretation. Finally, we have (9.47), since both sides correspond to h-networks. This is immediately clear for $U(x,y) - S(x,y)$, and for the left-hand-side we again apply Trakhtenbrot's theorem saying that every h-network has a unique representation as a 3-connected graph (minus the rooted edge), where every (remaining) edge is replaced by a network (with at least one edge).

It is now immediate that (9.38) and (9.39) follow from (9.43)–(9.47).

Theorem 9.13. *The numbers b_n, c_n and g_n of 2-connected, connected and all planar graphs are asymptotically given by*

$$b_n = b \cdot n^{-\frac{7}{2}} \rho_1^n \, n! \left(1 + O\left(\frac{1}{n}\right)\right),$$
$$c_n = c \cdot n^{-\frac{7}{2}} \rho_2^n \, n! \left(1 + O\left(\frac{1}{n}\right)\right),$$
$$g_n = g \cdot n^{-\frac{7}{2}} \rho_2^n \, n! \left(1 + O\left(\frac{1}{n}\right)\right),$$

where $\rho_1 = 0.03819...$, $\rho_2 = 0.03672841...$ *and*

$$b = 0.3704247487... \cdot 10^{-5},$$
$$c = 0.4104361100... \cdot 10^{-5},$$
$$g = 0.4260938569... \cdot 10^{-5}$$

are positive constants.

Proof. In a first step we consider the two equations for U and V. For notational convenience we substitute y by z. Obviously, the two equations for U and V can be reduced to a single one, say

$$U = xz\left(1 + z(1+U)^2\right)^2.$$

Then by Theorem 2.19 $U(x,z)$ has a singular representation of the form

$$U(x,z) = g(x,z) - h(x,z)\sqrt{1 - \frac{z}{\tau(x)}},$$

that we rewrite now (also for $V(x,z)$) as

$$\begin{aligned} U(x,z) &= u_0(x) + u_1(x)Z + u_2(x)Z^2 + u_3(x)Z^3 + O(Z^3), \\ V(x,z) &= v_0(x) + v_1(x)Z + v_2(x)Z^2 + v_2(x)Z^3 + O(Z^3), \end{aligned}$$

where Z abbreviates

$$Z = \sqrt{1 - \frac{z}{\tau(x)}}.$$

Observe that $u_0(x)$ is the solution of the equation

$$x = \frac{(1+u)(3u-1)^3}{16u}$$

and $\tau(x)$ is then given by

$$\tau(x) = \frac{1}{\left(4x^2(1+u_0(x))\right)^{2/3}}.$$

The functions $u_j(x)$ and $v_j(x)$ are also analytic and can be explicitly given in terms of $u = u_0(x)$. In particular we have

$$\begin{aligned} u_0(x) &= u & v_0(x) &= \frac{1+u}{3u-1}, \\ u_1(x) &= -\sqrt{2u(u+1)} & v_1(x) &= -\frac{2\sqrt{2u(u+1}}{3u-1}, \\ u_2(x) &= \frac{(1+u)(7u+1)}{2(1+3u)} & v_2(x) &= \frac{2u(3+5u)}{(3u-1)(1+3u)}, \\ u_3(x) &= -\frac{(1+u)(67u^2+50u+11)u}{4(1+3u)^2\sqrt{2u^2+2u}} & v_3(x) &= -\frac{\sqrt{2}u(1+u)(79u^2+42u+7)}{4(1+3u)^2(3u-1)\sqrt{u(1+u)}}. \end{aligned}$$

With the help of these expansions it follows that there is a cancellation of the coefficient of Z in the expansion of

$$\frac{(1+U)^2(1+V)^2}{(1+U+V)^3} = E_0 + E_2Z^2 + E_3Z^3 + O(Z^4),$$

where

$$\begin{aligned} E_0 &= \frac{16(3u-1)}{27u(u+1)}, \quad E_2 = \frac{16(3u^2+1)(3u-1)}{81u^2(u+1)^2}, \\ E_3 &= 2\frac{\sqrt{2}p_1(u) + \sqrt{u(u+1)}p_2(u)}{729\sqrt{u(u+1)}u^2(1+3u)^2(u+1)^2} \end{aligned}$$

with polynomials

$$p_1(u) = 324u^5 + 297u^4 + 54u^3 - 1296u^2 - 1146u - 281,$$
$$p_2(u) = 8748u^4 - 5832u^3 - 3888u^2 + 648u + 324.$$

Thus, $M(x, z)$ can be represented as

$$M(x, z) = M_0(x) + M_2(x)Z^2 + M_3(x)Z^3 + O(Z^4).$$

Next we rewrite (9.39) to

$$D = (1 + y) \exp\left(\frac{xD^2}{1 + xD} + \frac{M(x, D)}{2x^2 D}\right) - 1 = \Phi(x, y, D)$$

and suppose that y equals 1 (or is very close to 1). We observe that Theorem 2.19 does not apply here, since the region of analyticity of the right hand side is not large enough to find a solution of the system of equations $d_0 = G(x_0, 1, d_0)$, $1 = G_d(x_0, 1, d_0)$. However, we can apply Theorem 2.31 and obtain a local expansion for $D = D(x, y)$ of the form

$$D(x, y) = D_0(y) + D_2(y)X + D_3(y)X^3 + O(X^4), \tag{9.48}$$

where

$$X = \sqrt{1 - \frac{x}{R(y)}}$$

for some function $R(y)$.

In fact, we can be much more precise (compare with [9, 91]). Let $t = t(y)$ be defined by the equation

$$y = \frac{1 + 2t}{(1 + 3t)(1 - t)} \exp\left(-\frac{t^2(1 - t)(18 + 36t + 5t^2)}{2(3 + t)(1 + 2t)(1 + 3t)^2}\right) - 1 \tag{9.49}$$

that exists in a suitable neighbourhood of $y = 1$. Then $R(y)$ is given by

$$R(y) = \frac{(1 + 3t(y))(1 - t(y))^3}{16t(y)^3},$$

in particular $R = R(1) = 0.038191...$, and the coefficients in (9.48) are given by

$$D_0 = \frac{3t^2}{(1 - t)(1 + 3t)},$$
$$D_2 = -\frac{48t^2(1 + t)(1 + 2t)^2(18 + 6t + t^2)}{(1 + 3t)\beta},$$
$$D_3 = 384t^2(1 + t)^2(1 + 2t)^2(3 + t)^2\alpha^{3/2}\beta^{-5/2},$$

with

$$\alpha = 144 + 592t + 664t^2 + 135t^3 + 6t^4 - 5t^5,$$
$$\beta = 3t(1+t)(400 + 1808t + 2527t^2 + 1155t^3 + 236t^4 + 17t^5).$$

In view of (9.38) the representation (9.48) provides a local expansion for $\frac{\partial B(x,y)}{\partial y}$ of the form

$$\frac{\partial B(x,y)}{\partial y} = \overline{B}_0(y) + \overline{B}_2(y)X + \overline{B}_3(y)X^3 + O(X^4)$$
$$= g_1(x,y) + h_1(x,y)X^{3/2}$$

with certain analytic functions $g_1(x,y)$ and $h_1(x,y)$. Hence, by Theorem 2.30 $B(x,y)$ and consequently $\frac{\partial B(x,y)}{\partial x}$ have an expansions of the form

$$B(x,y) = g_2(x,y) + h_2(x,y)X^{5/2},$$
$$\frac{\partial B(x,y)}{\partial x} = g_3(x,y) + h_3(x,y)X^{3/2}$$

with certain analytic functions $g_2(x,y)$, $g_3(x,y)$ and $h_2(x,y)$, $h_2(x,y)$.

Finally we have to solve (9.37). For simplicity set $y = 1$. Since $R\,B''(R) \approx 0.0402624 < 1$, Theorem 2.19 does not apply (see also [91]). The singularity of the right-hand-side induces the singular behaviour of the solution $xC'(x)$. Actually we just have to apply Theorem 2.31 and obtain a local expansion for $C'(x)$ of the form

$$C'(x) = g_3(x) + h_3(x)\left(1 - \frac{x}{\rho}\right)^{\frac{3}{2}},$$

where $\rho = Re^{-B'(R)} = 0.0367284...$, and consequently we obtain corresponding representations for

$$C(x) = g_4(x) + h_4(x)\left(1 - \frac{x}{\rho}\right)^{\frac{5}{2}}$$

and for

$$G(x) = g_5(x) + h_5(x)\left(1 - \frac{x}{\rho}\right)^{\frac{5}{2}}.$$

Using these representations the asymptotic expansion for b_n, c_n, and g_n follow immediately. Again one has to be careful with the evaluation of constants (compare with Remark 9.7). They can be either evaluated by numerical integration or by using explicit representations for $B(x,y)$ and $C(x,y)$ (compare with [91]).

Remark 9.14 *As in the previous cases of outerplanar and series-parallel graphs the number X_n of edges in a random planar graph satisfies a central limit theorem. For all planar graphs we have*

$$\mathbb{E}\, X_n = \mu n + O(1) \quad \textit{and} \quad \mathbb{V}\, X_n = \sigma^2 n + O(1),$$

where $\mu = 2.2132652...$ and $\sigma^2 = 0.4303471...$ (compare with [91]).

9.3 Outerplanar Graphs

9.3.1 The Degree Distribution of Outerplanar Graphs

We already determined the degree distribution of several classes of random trees. We next show that such distributions exist for planar graphs, too. Recall, that we are interested in the limits

$$d_k = \lim_{n\to\infty} d_{k;n},$$

where $d_{k;n}$ denotes the probability that a randomly chosen node in a random planar graph has degree k.

In fact there are two different points of view to this problem. First, if $X_n^{(k)}$ denotes the (random) number of vertices of degree k in a random planar graph of size n then

$$d_{k;n} = \frac{\mathbb{E}\, X_n^{(k)}}{n}.$$

Second, we can change the model. Recall that we are considering unrooted labelled combinatorial objects. Suppose that g_n is the number of objects of size n. Then for each of these objects there are precisely n possible ways to root these objects and we, thus, obtain ng_n corresponding rooted objects. Now the probability that a random vertex (in the unrooted model) has degree k is precisely the same as the probability that the root vertex (in the rooted model) has degree k. Thus, if we are only interested in the degree distribution it is sufficient to look at the degree of the root. In what follows we will make use of both points of view. We will start by looking at the degree of the root (which is actually easier) and then we will have a closer look at the random variable $X_n^{(k)}$. As a matter of fact these random variables (usually) obey a central limit theorem with mean and variance asymptotically proportional to n. Of course, this is a much more precise statement than just the existence of a degree distribution.

Before we start with the calculations we make some general remarks on the methods we use. Again, they are based on generating functions. Let $C_k(x)$ be the exponential generating function of a certain class[4] of rooted vertex labelled connected graphs, where the root bears no label and has degree k; that is, the coefficient $[x^n/n!]C_k(x)$ equals the number of rooted connected graphs with $n+1$ vertices, in which the root has no label and has degree k. Analogously we define $B_k(x)$ for 2-connected graphs. Also, let

$$B^\bullet(x,w) = \sum B_k(x) w^k, \qquad C^\bullet(x,w) = \sum C_k(x) w^k.$$

Suppose that we have the relation

[4] In this context we can confine ourselves to outerplanar graphs, series-parallel graphs or planar graphs.

$$C'(x) = e^{B'(xC'(x))}$$

between the generating function for 2-connected and connected graphs (which is certainly true for outerplanar graphs, series-parallel graphs or planar graphs). If we introduce the degree of the root, then the equation becomes

$$C^\bullet(x,w) = e^{B^\bullet(xC'(x),w)}. \tag{9.50}$$

The reason is that only the 2-connected components containing the root vertex contribute to its degree.

Our goal in each case is to estimate $[x^n]C_k(x)$, since the limit probability that a given fixed vertex has degree k is equal to

$$d_k = \lim_{n\to\infty} \frac{[x^n]C_k(x)}{[x^n]C'(x)}. \tag{9.51}$$

Another observation is that the asymptotic degree distribution is the same for connected members of a class than for all members in the class. Let $G(x)$ be the generating function for all members of the class, and let $G_k(x)$ be the generating function of all rooted graphs of the class, where the root has degree k. Then we have

$$G(x) = e^{C(x)}, \qquad G_k(x) = C_k(x)e^{C(x)}.$$

The first equation is standard, and in the second equation the factor $C_k(x)$ corresponds to the connected component containing the root, and the second factor to the remaining components. The functions $G(x)$ and $C(x)$ have the same dominant singularity. Given the singular expansions of $G(x)$ and $C(x)$ at the dominant singularity in each of the considered cases, it follows that

$$\lim_{n\to\infty} \frac{[x^n]G_k(x)}{[x^n]G'(x)} = \lim_{n\to\infty} \frac{[x^n]C_k(x)}{[x^n]C'(x)}.$$

Hence, in each case we only need to determine the degree distribution for connected graphs. A more intuitive explanation is that the largest component in random planar graphs eats up almost everything. It is a fact that the expected number of vertices not in the largest component is constant [149].

As in the counting procedure of outerplanar graphs we discuss first dissections and can then derive the results for 2-connected outerplanar graphs and finally for connected graphs.

Recall that the generating function for dissections is given by

$$A(x) = \frac{1 - 3x - \sqrt{1 - 6x + x^2}}{4x}, \tag{9.52}$$

and let $A_k(x)$ denote the generating function, where the root vertex (the first vertex of the rooted edge) has degree k and where the root vertices are not counted by x. Then by using the recursive structure of dissections we immediately get

$$\begin{aligned} A_2(x) &= x(A(x)+1) + (x(A(x)+1))^2 + (x(A(x)+1))^3 + \cdots \\ &= \frac{x(A(x)+1)}{1-x(A(x)+1)} \\ &= x(2A(x)+1). \end{aligned}$$

Furthermore, we have inductively

$$\begin{aligned} A_{k+1}(x) &= A_k(x)x(A(x)+1) + A_k(x)(x(A(x)+1))^2 + A_k(x)(x(A(x)+1))^3 + \cdots \\ &= A_k(x)x(2A(x)+1). \end{aligned}$$

Consequently we obtain for $k \ge 2$

$$A_k(x) = (x(2A(x)+1))^{k-1}.$$

Due to the general relation between dissections and 2-connected outerplanar graphs we have

$$B_k(x) = \frac{x}{2}A_k(x),\ k \ge 2, \qquad B_1 = x.$$

By summing a geometric series we have an explicit expression for $B^\bullet$, namely

$$\begin{aligned} B^\bullet(x,w) &= xw + \sum_{k=2}^{\infty} \frac{x^k}{2}(2A(x)+1)^{k-1}w^k \\ &= xw + \frac{xw^2}{2}\frac{x(2A(x)+1)}{1-x(2A(x)+1)w}. \end{aligned} \tag{9.53}$$

Theorem 9.15. *For every $k \ge 2$ the limiting probability d_k that a vertex of a two-connected outerplanar graph has degree k exists and we have*

$$p(w) = \sum_{k\ge 2} d_k w^k = \frac{2(3-2\sqrt{2})w^2}{(1-(\sqrt{2}-1)w)^2} = \sum_{k\ge 2} 2(k-1)(\sqrt{2}-1)^k w^k.$$

Moreover $p(1)=1$, so that the d_k are indeed a probability distribution.

Remark 9.16 *The degree distribution of dissections is precisely the same as of outerplanar graphs (compare also with [13]).*

Proof. Since $B^\bullet(x,1) = B'(x)$ and $A(x) = 2(B'(x)-x)/x$, we can represent $B^\bullet(x,w)$ as

$$B^\bullet(x,w) = xw + \frac{xw^2}{2}\frac{4B'(x)-3x}{1-(4B'(x)-3x)w}.$$

Hence, by applying Lemma 2.26 with $f(x) = B'(x)$ and

$$H(x,z) = xw + \frac{xw^2}{2}\frac{4z-3x}{1-(4z-3x)w}$$

we obtain

$$p(w) = \lim_{n\to\infty}\frac{[x^n]B^\bullet(x,w)}{[x^n]B'(x)} = \frac{2(3-2\sqrt{2})w^2}{(1-(\sqrt{2}-1)w)^2}.$$

Note that $\rho = 3-2\sqrt{2}$ and that w is considered as an additional (complex) parameter.

Theorem 9.17. *Let d_k be the limit probability that a vertex of a connected outerplanar graph has degree k. Then*

$$p(w) = \sum_{k\ge 1} d_k w^k = \rho\cdot\frac{\partial}{\partial x}e^{B^\bullet(x,w)}\big|_{x=\rho C'(\rho)},$$

where $B^\bullet$ is given by equations (9.52) and (9.53).

Moreover $p(1) = 1$, so that the d_k are indeed a probability distribution and we have asymptotically, as $k\to\infty$

$$d_k \sim c_1 k^{1/4} e^{c_2\sqrt{k}} q^k,$$

where $c_1 = 0.667187...$, $c_2 = 0.947130...$, and $q = 0.3808138....$

Proof. We have

$$\sum_k C_k(x)w^k = C^\bullet(x,w) = e^{B^\bullet(xC'(x),w)}.$$

The radius of convergence $3-2\sqrt{2}$ of $B(x)$ is larger than $\rho C'(\rho) = \tau \approx 0.17076$. Hence we can apply Lemma 2.26 $f(x) = xC'(x)$ and $H(x,w,z) = xe^{B^\bullet(z,w)}$, where w is considered as a parameter. Then we have

$$\begin{aligned}\left[\frac{\partial}{\partial x}e^{B^\bullet(x,w)}\right]_{x=\rho C'(\rho)} &= \lim_{n\to\infty}\frac{[x^n]C^\bullet(x,w)}{[x^n]xC'(x)}\\ &= \lim_{n\to\infty}\sum_{k\ge 1}\rho^{-1}\frac{[x^n]C_k(x)}{[x^n]C'(x)}w^k\\ &= \rho^{-1}\sum_{k\ge 1}d_k w^k,\end{aligned}$$

and the result follows.

For the second assertion let us note that $B^\bullet(x,1) = B'(x)$. If we recall that $\rho C'(\rho) = v_0$ and $v_0 B''(v_0) = 1$, then

$$p(1) = \rho\, e^{B'(v_0)}B''(v_0) = \rho C'(\rho)v_0^{-1} = 1.$$

In order to get an asymptotic expansion for d_k we have to compute $p(w)$ explicitly:

$$p(w) = \rho \frac{v_0^2(2A(v_0)+1)(2A+1+2v_0A'(v_0))w^2}{2(1-v_0(2A(v_0)+1)w)^2} \times \exp\left(v_0 w + \frac{v_0^2(2A(v_0)+1)w^2}{2(1-v_0(2A(v_0)+1)w)} \right).$$

This is a function that is admissible in the sense of Hayman [99]. Hence, it follows that

$$d_k \sim \frac{p(r_k)r_k^{-k}}{\sqrt{2\pi b(r_k)}},$$

where r_k is given by the equation $r_k p'(r_k)/p(r_k) = k$ and $b(w) = w^2 p''(w)/p(w) + wp'(w)/p(w) - (wp'(w)/p(w))^2$. A standard calculation gives the asymptotic expansion for the coefficients d_k.

9.3.2 Vertices of Given Degree in Dissections

In Section 9.3.1 we have demonstrated that the limiting probabilities d_k of the probabilities that a random vertex has degree k exist. Equivalently this says that the number $X_n^{(k)}$ of vertices of degree d (in a random graph of size n) satisfies

$$\mathbb{E}\, X_n^{(k)} \sim d_k n.$$

The purpose of this (and the following) sections is to obtain more precise information on $X_n^{(k)}$. In fact, we show that $X_n^{(k)}$ satisfies a central limit theorem which implies that $X_n^{(k)}$ is concentrated around its mean.

Theorem 9.18. *For $k \ge 2$, let $X_n^{(k)}$ denote the number of vertices of degree k in random dissection with $n+2$ vertices. Then $X_n^{(k)}$ satisfies a central limit theorem with linear expected value and variance.*

Actually we already know that the expected value is asymptotically given by

$$\mathbb{E}\, X_n^{(k)} = 2(k-1)(\sqrt{2}-1)^k\, n + O(1),$$

since the degree distribution of dissections is the same as that of 2-connected outerplanar graphs.

Before we start with the proof of Theorem 9.18 we reprove the relation (9.4)

$$A(x) = x(1+A(x))^2 + x(1+A(x))A(x)$$

by using a different recursive construction (see Figure 9.12) that provides this relation and will be useful for the analysis of $X_n^{(k)}$.

As in the proof of Theorem 9.1 we observe that every dissection α has a unique face f that contains the root edge. Suppose first that f contains exactly three vertices, that is, f is a triangle, and denote the three edges of f by $e = (v_1, v_2)$, the root edge, and by $e_1 = (v_2, v_3)$ and $e_2 = (v_3, v_1)$. (We

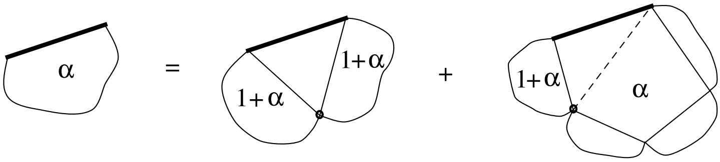


Fig. 9.12. Recursive decomposition of dissections

always label the vertices in counterclockwise order.) We can then decompose α into three parts. We cut α at the three vertices v_1, v_2, v_3 of f and get first the root edge e, then a part α_1 of α that contains e_1, and finally a third part α_2 that contains e_2. Obviously α_1 and α_2 are again connected planar graphs, where e_1 and e_2 can be viewed as rooted edges. Since all vertices of α are on the infinite face, the same holds for α_1 and α_2, however, α_1 and α_2 might consist just of e_1 or e_2. Thus, α_1 and α_2 are either just one (rooted) edge or again a dissection. By counting the number of vertices in the way described above this case corresponds to the generating function $x(1 + A(x))^2$.

If f contains more than three edges then we first cut f into two pieces f_1 and f_2, where f_1 consists of the root edge $e = (v_1, v_2)$, the adjacent edge $e_1 = (v_2, v_3)$, and a new edge $e_{new} = (v_3, v_1)$. We again cut α at the vertices v_1, v_2, v_3 and get, first, the root edge e, then a part α_1 of α that contains e_1 and a third part $\tilde{\alpha}_2$ that contains the new edge e_{new}. As above α_1 is either e_1 or is a dissection rooted at e_1. Since f has more than three edges, $\tilde{\alpha}_2$ has at least three vertices. Hence, we can consider $\tilde{\alpha}_2$ as a dissection rooted at e_{new}. Similarly to the above, this case corresponds to the generating function $x(1 + A(x))A(x)$.

The method of the proof of Theorem 9.18 is to provide a system of functional equations for proper generating functions from which we can read off the central limit theorem for $X_n^{(k)}$. For this purpose we need an extension of the above generating function counting procedure, using variables $x, z_1, z_2, \ldots, z_k, z_\infty$, where the variable z_ℓ, $1 \le \ell \le k$, marks vertices of degree ℓ, and z_∞ marks vertices of degree greater than k. Furthermore, we consider the degrees i, j of the vertices v_1 and v_2 of the root edge $e = (v_1, v_2)$. More precisely, if

$$a_{i,j;n,n_1,n_2,\ldots,n_k,n_\infty}$$

is the number of dissections with $2 + n = 2 + n_1 + n_2 + \cdots + n_k + n_\infty \ge 3$ vertices such that the two vertices v_1, v_2 of the marked edge $e = (v_1, v_2)$ have degrees $d(v_1) = i$ and $d(v_2) = j$, and that for $1 \le \ell \le k$ there are n_ℓ vertices $v \ne v_1, v_2$ with $d(v) = \ell$, and there are n_∞ vertices $v \ne v_1, v_2$ with $d(v) > k$. The corresponding generating functions are then defined by

$$A_{i,j}(x, z_1, z_2, \ldots, z_k, z_\infty) = \sum_{n,n_1,\ldots,n_k,n_\infty} a_{i,j;n,n_1,n_2,\ldots,n_k,n_\infty} x^n z_1^{n_1} \cdots z_k^{n_k} z_\infty^{n_\infty}. \tag{9.54}$$

Similarly we define $A_{i,\infty}$, $A_{\infty,j}$ and $A_{\infty,\infty}$, if one (or both) of the vertices of the root edge have degree(s) greater than k.

Note that z_1 is not necessary, since there are no vertices of degree one in dissections. However, we use z_1 for later purposes. Further observe that $A_{ij}(z_1, z_2, \ldots, z_k, z_\infty) = A_{ji}(z_1, z_2, \ldots, z_k, z_\infty)$. Thus, it is sufficient to consider A_{ij} for $i \le j$.

In order to state the following lemma in a more compact form we use the convention that ∞ means $> k$, and $\infty - 1$ means $> k-1$. In particular we set $\ell + \infty = \infty$ for all positive integers ℓ.

Lemma 9.19. *The generating functions $A_{ij} = A_{ji} = A_{ij}(x, z_1, z_2, \ldots, z_k, z_\infty)$, $i, j \in \{2, 3, \ldots, k, \infty\}$ satisfy the following strongly connected positive system of equations:*

$$\begin{aligned} A_{ij} = x \sum_{\ell_1+\ell_2 \le k} z_{\ell_1+\ell_2} A_{i-1,\ell_1} A_{j-1,\ell_2} + x z_\infty \left(\sum_{\ell_1+\ell_2 > k} A_{i-1,\ell_1} A_{j-1,\ell_2} \right) \\ + x \sum_{\ell_1+\ell_2 \le k+1} z_{\ell_1+\ell_2-1} A_{i-1,\ell_1} A_{j,\ell_2} + x z_\infty \left(\sum_{\ell_1+\ell_2 > k+1} A_{i-1,\ell_1} A_{j,\ell_2} \right). \end{aligned}$$

One has to be cautious in writing down the equations explicitly. For example we have

$$\begin{aligned} A_{i,\infty} = x &\sum_{\ell_1+\ell_2 \le k} z_{\ell_1+\ell_2} A_{i-1,\ell_1} (A_{k,\ell_2} + A_{\infty,\ell_2}) \\ &+ x z_\infty \left(\sum_{\ell_1+\ell_2 > k} A_{i-1,\ell_1} (A_{k,\ell_2} + A_{\infty,\ell_2}) \right) \\ &+ x \sum_{\ell_1+\ell_2 \le k+1} z_{\ell_1+\ell_2-1} A_{i-1,\ell_1} A_{\infty,\ell_2} \\ &+ x z_\infty \left(\sum_{\ell_1+\ell_2 > k+1} A_{i-1,\ell_1} A_{\infty,\ell_2} \right). \end{aligned}$$

As an illustration, for $k = 3$ we have the following system:

$$\begin{aligned} A_{22} &= x z_2 \\ &\quad + x z_2 A_{22} + x z_3 A_{23} + x z_\infty A_{2\infty}, \\ A_{23} &= x z_3 A_{22} + x z_\infty (A_{23} + A_{2\infty}) \\ &= x z_2 A_{23} + x z_3 A_{33} + x z_\infty A_{3\infty}, \\ A_{2\infty} &= x z_3 A_{23} + x z_\infty (A_{33} + A_{3\infty}) + x z_\infty (A_{2\infty} + A_{3\infty} + A_{\infty,\infty}) \\ &\quad + x z_2 A_{2\infty} + x z_3 A_{3\infty} + x z_\infty A_{\infty,\infty}, \end{aligned}$$

$$\begin{aligned}
A_{33} &= xz_\infty(A_{22}+A_{23}+A_{2\infty})^2 \\
&\quad + xz_\infty(A_{22}+A_{23}+A_{2\infty})(A_{23}+A_{33}+A_{3\infty}), \\
A_{3\infty} &= xz_\infty(A_{23}+A_{33}+A_{3\infty})(A_{2\infty}+A_{3\infty}+A_{\infty,\infty}) \\
&\quad + xz_\infty(A_{22}+A_{23}+A_{2\infty})(A_{2\infty}+A_{3\infty}+A_{\infty,\infty}), \\
A_{\infty,\infty} &= xz_\infty(A_{23}+A_{33}+A_{3\infty}+A_{2\infty}+A_{3\infty}+A_{\infty,\infty})^2 \\
&\quad + xz_\infty(A_{23}+A_{33}+A_{3\infty}+A_{2\infty}+A_{3\infty}+A_{\infty,\infty}) \\
&\qquad \times (A_{2\infty}+A_{3\infty}+A_{\infty,\infty}).
\end{aligned}$$

Proof. The idea is to have a more detailed look at the (second version of the) recursive structure of $\mathcal{A}$ as described above (see Figure 9.12).

We only discuss the recurrence for A_{ij} for finite i, j. If $i = \infty$ or $j = \infty$ similar considerations apply. The root edge will be denoted by $e = (v_1, v_2)$. We assume that v_1 has degree j and v_2 has degree i. Again we have to distinguish between the case where the face f containing the root edge e has exactly three edges, and the case where it has more than three edges.

In the first case we cut a dissection α at the three vertices v_1, v_2, v_3 of f and get the root edge e, and two dissections α_1 and α_2 that are rooted at $e_1 = (v_2, v_3)$ and $e_2 = (v_3, v_1)$. After the cut, α_1 has degree $i-1$ at v_2, and α_2 has degree $j-1$ at v_1. Furthermore, the total degree of the common vertex v_3 is just the sum of the degrees coming from α_1 and α_2. Hence, if the degree of v_3 is smaller or equal than k, then this situation corresponds to the generating function

$$x \sum_{\ell_1+\ell_2\le k} z_{\ell_1+\ell_2} A_{i-1,\ell_1} A_{j-1,\ell_2}.$$

Since all possible cases for α_1 are encoded in $A_{i-1} = A_{i-1,2} + \cdots + A_{i-1,k} + A_{i-1,\infty}$, and all cases for α_2 are encoded in A_{j-1}, it follows that all situations where the total degree of v_3 is greater than k are given by the generating function

$$xz_\infty \left(\sum_{\ell_1+\ell_2>k} A_{i-1,\ell_1} A_{j-1,\ell_2} \right).$$

Similarly we argue in the case where f contains more than three edges. After cutting f into two pieces f_1 and f_2 with a new edge $e_{new} = (v_3, v_1)$, and cutting α at the vertices v_1, v_2, v_3, we get again the root edge e and two dissections α_1 and α_2 that are rooted at $e_1 = (v_2, v_3)$, and at the new edge $e_{new} = (v_3, v_1)$. After the cut, α_1 has degree $i-1$ at v_2 and α_2 has degree j at v_1, since the new edge e_{new} has to be taken into account. The total degree of the common vertex v_3 is the sum of the degrees coming from α_1 and α_2 minus 1, since the new edge e_{new} is used in the construction of α_2. As above these observations translate into the generating functions

$$x \sum_{\ell_1+\ell_2\le k+1} z_{\ell_1+\ell_2-1} A_{i-1,\ell_1} A_{j,\ell_2}$$

and

$$xz_\infty \left(\sum_{\ell_1+\ell_2>k+1} A_{i-1,\ell_1} A_{j,\ell_2} \right).$$

Finally, by using the definition of $A_{i,j}$ and A_∞, it follows that the above system of equations is a positive one, that is, all coefficients on the right hand side are non-negative. Further, it is easy to check that the corresponding dependency graph is strongly connected (which means that no subsystem can be solved before the whole system is solved).

The proof of Theorem 9.18 is now a direct application of the analytic central limit theorem related to systems of generating functions formulated in Theorem 2.35. For this purpose we have to check that the assumptions are satisfied. Note, that if we set all variables $z_1 = \cdots = z_k = z_\infty = 1$ and sum all function $A_{i,j}$ then they sum up to $A(x)$ which has a square root singularity at $\rho = 3 - 2\sqrt{2}$. Thus, we can proceed as in the proof of Theorem 3.18 to show that the assumptions of Theorem 2.35 are satisfied.

In particular the function

$$A_{d=k}(x,u) = \sum_{2\le i\le\infty} A_{i,i}(x,1,\ldots,1,u,1) + 2 \sum_{2\le i<j\le\infty} A_{i,j}(x,1,\ldots,1,u,1)$$

is the bivariate generating function for dissections with $n+2$ vertices, where the exponent of u counts the number of vertices different from the root vertex that has degree k. Of course, it is also possible to take the root vertex into account by considering the function

$$\begin{aligned}\tilde{A}_{d=k}(x,u) = x^2 &\sum_{2\le i\le\infty, i\ne k} A_{i,i}(x,1,\ldots,1,u,1) + x^2u^2 A_{k,k}(x,1,\ldots,1,u,1)\\ &+ 2x^2 \sum_{2\le i<j\le\infty, i\ne k, j\ne k} A_{i,i}(x,1,\ldots,1,u,1)\\ &+ 2x^2 u \sum_{2\le i\le\infty, i\ne k} A_{i,k}(x,1,\ldots,1,u,1).\end{aligned}$$

For both versions we can apply Theorem 2.35 and obtain a central limit theorem for the number of vertices of degree k.

9.3.3 Vertices of Given Degree in 2-Connected Outerplanar Graphs

We consider now vertex labelled 2-connected outerplanar graphs. The result we obtain is exactly the same as for dissections.

Theorem 9.20. *For $k \ge 2$, let $X_n^{(k)}$ denote the number of vertices of degree k in a random 2-connected outerplanar graph with n vertices. Then $X_n^{(k)}$ satisfies a central limit theorem with linear expected value and variance.*

Recall that we already know that the expected value is asymptotically given by

$$\mathbb{E}\, X_n^{(k)} = 2(k-1)(\sqrt{2}-1)^k\, n + O(1).$$

For the proof we have to adapt the corresponding generating functions. Note that the generating function

$$B'(x) = \sum_{n\ge 2} b_n \frac{x^{n-1}}{(n-1)!} = \sum_{n\ge 1} b_{n+1}\frac{x^n}{n!}$$

can also be interpreted as the exponential generating function $B^\bullet(x)$ of 2-connected outerplanar graphs, where one vertex is marked and is not counted.

Next we set

$$B_j^\bullet(x, z_1, z_2, \ldots, z_k, z_\infty) = \sum_{n,n_1,\ldots,n_k,n_\infty} b^\bullet_{j;n,n_1,\ldots,n_k,n_\infty} \frac{x^n z_1^{n_1}\cdots z_k^{n_k} z_\infty^{n_\infty}}{n!},$$

where $b^\bullet_{j;n,n_1,\ldots,n_k,n_\infty}$ is the number of 2-connected outerplanar graphs with $1+n = 1+n_1+\cdots+n_k+n_\infty$ vertices, where one vertex of degree j is marked and the remaining n vertices are labelled by $1, 2, \ldots, n$ and where n_ℓ vertices have degree ℓ, $1 \le \ell \le k$, and n_∞ vertices have degree greater than k.

Lemma 9.21. *Let* $A_{ij} = A_{ji} = A_{ij}(x, z_1, z_2, \ldots, z_k, z_\infty)$, $i, j \in \{1, 2, \ldots, k, \infty\}$ *be defined by (9.54). Then the functions* $B_j = B_j^\bullet(x, z_1, z_2, \ldots, z_k, z_\infty)$, $j \in \{1, 2, \ldots, k, \infty\}$, *are given by*

$$B_1^\bullet = xz_1,$$

$$B_j^\bullet = \frac{1}{2}\sum_{i=1}^{k} xz_i A_{ij} + \frac{1}{2} xz_\infty A_{j\infty},$$

$$B_\infty^\bullet = \frac{1}{2}\sum_{i=1}^{k} xz_i A_{j\infty} + \frac{1}{2} xz_\infty A_{\infty,\infty}.$$

Proof. The proof is immediate by using the relation between dissections and outerplanar graphs. However, we have to take care of the vertex degrees.

Now let

$$B_{d=k}(x,u) = \sum_{n,\nu} b_{n,\nu}^{(k)} \frac{x^n}{n!} u^\nu$$

denote the exponential generating function of the numbers $b_{n,\nu}^{(k)}$ of 2-connected outerplanar labelled graphs with n vertices, where ν vertices have degree k. Then we have

$$\frac{\partial B_{d=k}(x,u)}{\partial x} = \sum_{j=1}^{k-1} B_j^\bullet(x, 1, \ldots, 1, u, 1) + uB_k^\bullet(x, 1, \ldots, 1, u, 1) + B_\infty^\bullet(x, 1, \ldots, 1, u, 1). \qquad (9.55)$$

Since $B(0,u)=0$, this equation completely determines $B_{d=k}(x,u)$.

As in the case of dissections all assumptions of Theorem 2.35 are satisfied. In particular we get a singular representation of the form

$$\frac{\partial B_{d=k}(x,u)}{\partial x} = g(x,u) - h(x,u)\sqrt{1-\frac{x}{\rho(u)}}$$

and consequently a corresponding singular representation for $B_{d=k}(x,u)$:

$$B_{d=k}(x,u) = \overline{g}(x,u) + \overline{h}(x,u)\left(1-\frac{x}{\rho(u)}\right)^{\frac{3}{2}}.$$

Of course, this also implies a central limit theorem for $X_n^{(k)}$.

9.3.4 Vertices of Given Degree in Connected Outerplanar Graphs

The next theorem concerns connected outerplanar graphs. Since the result (even at the level of expectation and variance) is the same for all outerplanar graphs, we include this case, too.

Theorem 9.22. *For $k \ge 2$, let $X_n^{(k)}$ denote the number of vertices of degree k in random connected (or a general) outerplanar graph with n vertices. Then $X_n^{(k)}$ satisfies a central limit theorem with linear expected value and variance.*

The proof uses a refined version of the general relation between rooted 2-connected graphs and rooted connected graphs which states in terms of generating functions as

$$C^\bullet(x) = e^{B^\bullet(xC^\bullet(x))}. \tag{9.56}$$

For this purpose we introduce the generating functions

$$C_j^\bullet(x, z_1, z_2, \ldots, z_k, z_\infty) = \sum_{n,n_1,\ldots,n_k,n_\infty} c^\bullet_{j;n;n_1,\ldots,n_k,n_\infty} z_1^{n_1}\cdots z_k^{n_k} z_\infty^{n_\infty}\frac{x^n}{n!},$$

$j \in \{1,2,\ldots,k,\infty\}$, where $c^\bullet_{j;n;n_1,\ldots,n_k,n_\infty}$ is the number of connected outerplanar graphs with $1+n = 1+n_1+\cdots+n_k+n_\infty$ vertices, where one vertex of degree j is marked[5] and the remaining n vertices are labelled by $1,2,\ldots,n$ and where n_ℓ of these n vertices have degree ℓ, $1\le \ell \le k$, and n_∞ of these vertices have degree greater than k. For convenience, we also define

$$C_0^\bullet(x, z_1, z_2, \ldots, z_k, z_\infty) = 1,$$

which corresponds to the case of a graph with a single rooted vertex.

The main observation is that these functions satisfy a system of functional equations, a refined version of (9.56).

[5] If $j=\infty$ this has to be interpreted as a vertex of degree $> k$.

Lemma 9.23. *Let* $W_j = W_j(z_1, \ldots, z_k, z_\infty, C_1^\bullet, \ldots, C_k^\bullet, C_\infty^\bullet)$, $j \in \{1, 2, \ldots, k, \infty\}$ *be defined by*

$$W_j = \sum_{i=0}^{k-j} z_{i+j} C_i^\bullet(x, z_1, \ldots, z_k, z_\infty) + z_\infty \left(\sum_{i=k-j+1}^{k} C_i^\bullet(x, z_1, \ldots, z_k, z_\infty) + C_\infty^\bullet(x, z_1, \ldots, z_k, z_\infty) \right), \quad (1 \le j \le k),$$

$$W_\infty = z_\infty \left(\sum_{i=0}^{k} C_i^\bullet(x, z_1, \ldots, z_k, z_\infty) + C_\infty^\bullet(x, z_1, \ldots, z_k, z_\infty) \right).$$

Then the functions $C_1^\bullet, \ldots, C_k^\bullet, C_\infty^\bullet$ *satisfy the system of equations*

$$C_j^\bullet(x, z_1, \ldots, z_k, z_\infty) = \sum_{\ell_1 + 2\ell_2 + 3\ell_3 + \cdots j\ell_j = j} \prod_{r=1}^{j} \frac{B_r^\bullet(x, W_1, \ldots, W_k, W_\infty)^{\ell_r}}{\ell_r!} \quad (1 \le j \le k),$$

$$C_\infty^\bullet(x, z_1, \ldots, z_k, z_\infty) = \exp\left(\sum_{j=1}^{k} B_j^\bullet(x, W_1, \ldots, W_k, W_\infty) + B_\infty^\bullet(x, W_1, \ldots, W_k, W_\infty) \right) - 1 - \sum_{1 \le \ell_1 + 2\ell_2 + 3\ell_3 + \cdots k\ell_k \le k} \prod_{r=1}^{k} \frac{B_r^\bullet(x, W_1, \ldots, W_k, W_\infty)^{\ell_r}}{\ell_r!}.$$

Proof. As already indicated, the proof is a refined version of the proof of (9.56), which reflects the decomposition of a rooted connected graph into a finite set of rooted 2-connected graphs, where every vertex (different from the root) is substituted by a rooted connected graph. The functions W_j serve the purpose of marking (recursively) the degree of the vertices in the 2-connected blocks which are substituted by other graphs. If we look at the definition of W_j, the summation means that we are substituting a vertex of degree i, but since originally the vertex had degree j, we are creating a new vertex of degree $i + j$, which is marked accordingly by z_{i+j}. The same remark applies to W_∞.

Finally let

$$C_{d=k}(x, u) = \sum_{n,\nu} c_{n,\nu}^{(k)} \frac{x^n}{n!} u^\nu$$

denote the exponential generating function for the numbers $c_{n,\ell}^{(k)}$ of connected outerplanar vertex labelled graphs with n vertices, where ν vertices have degree k. Then we have

$$\frac{\partial C_{d=k}(x,u)}{\partial x} = \sum_{j=1}^{k-1} C_j^\bullet(x,1,\ldots,1,u,1) + uC_k^\bullet(x,1,\ldots,1,u,1) + C_\infty^\bullet(x,1,\ldots,1,u,1). \tag{9.57}$$

Since $C(0,u) = 0$, this equation completely determines $C_{d=k}(x,u)$.

Thus, the derivative $\frac{\partial C_{d=k}(x,u)}{\partial x}$ is a linear combination of functions $C_j^\bullet(x, z_1, \ldots, z_k, z_\infty)$ that satisfy a positive and strongly connected system of equations. Note that the system of equations given in Lemma 9.23 uses the functions $B_j^\bullet(x, z_1, \ldots, z_k, z_\infty)$, that have a square-root singularity if the z_j are sufficiently close to 1; recall that $B'(x)$ and $B_j^\bullet(x,1,\ldots,1)$ have a square-root singularity at $\rho_1 = 3 - 2\sqrt{2}$. In order to apply Theorem 2.35 we have to check that the *critical point* of the system (9.57) does not interfere with the square-root singularity of $B_j^\bullet$.

We already know that the radius of convergence of $C'(x)$ is given by $\rho_2 = 0.1366\cdots$ which satisfies $\rho_2 C'(\rho_2) < \rho_1$ so that the singularity of $B'(x)$ does not interfere with the singularity of $C'(x)$; recall that $xC'(x) = xe^{B'(xC'(x))}$.

We are again in a situation, where we can apply Theorem 2.35 (we can argue in the same way as in the proof of Theorem 3.18 or Theorem 9.18). In particular we get a singular representation of the form

$$\frac{\partial C_{d=k}(x,u)}{\partial x} = g(x,u) - h(x,u)\sqrt{1 - \frac{x}{\rho_2(u)}}$$

and consequently a corresponding singular representation for $C_{d=k}(x,u)$:

$$C_{d=k}(x,u) = \overline{g}(x,u) + \overline{h}(x,u)\left(1 - \frac{x}{\rho_2(u)}\right)^{\frac{3}{2}} \tag{9.58}$$

which implies a central limit theorem for $X_n^{(k)}$.

Finally, we obtain a corresponding representation for the generating function of $G_{d=k}(x,u)$ for all outerplanar graphs of the form

$$G_{d=k}(x,u) = e^{C_{d=k}(x,u)} = \tilde{g}(x,u) + \tilde{h}(x,u)\left(1 - \frac{x}{\rho_2(u)}\right)^{3/2} \tag{9.59}$$

and a central limit theorem in this case, too.

9.4 Series-Parallel Graphs

9.4.1 The Degree Distribution of Series-Parallel Graphs

In this section $C(x)$ and $B(x)$ denote the generating functions of connected and 2-connected, respectively, series-parallel graphs.

In order to study 2-connected series-parallel graphs, we need to consider series-parallel networks. We recall that a network is a graph with two distinguished vertices, called poles, such that the multi-graph obtained by adding an edge between the two poles is 2-connected. Let $D(x,y,w)$ be the exponential generating function of series-parallel networks, where x, y, w mark, respectively, vertices, edges, and the degree of the first pole. We define $S(x,y,w)$ analogously for series networks. Then we have

$$D(x,y,w) = (1+yw)e^{S(x,y,w)} - 1, \tag{9.60}$$

$$S(x,y,w) = (D(x,y,w) - S(x,y,w))\, xD(x,y,1). \tag{9.61}$$

The first equation reflects the fact that a network is a parallel composition of series networks, and the second one the fact that a series network is obtained by connecting a non-series network with an arbitrary network; the factor $D(x,y,1)$ appears because we only keep track of the degree of the first pole.

Remark 9.24 *It would not be necessary to take into account the number of edges and we could set $y = 1$ everywhere. However, in the case of planar graphs we do need generating functions according to all three variables and it is appropriate to present the full development already here. In the proof of the main result of this section, Theorem 9.29, we just set $y = 1$.*

We also set $E(x,y) = D(x,y,1)$ that is the generating function for series-parallel networks without marking the degree of the root. Then from (9.60)–(9.61) we obtain the equations

$$\log\left(\frac{1+E(x,y)}{1+y}\right) = \frac{xE(x,y)^2}{1+xE(x,y)} \tag{9.62}$$

and

$$\log\left(\frac{1+D(x,y,w)}{1+yw}\right) = \frac{xE(x,y)D(x,y,w)}{1+xE(x,y)}. \tag{9.63}$$

Let now $B_k^\bullet(x,y)$ be the generating function for 2-connected series-parallel graphs, where the root bears no label and has degree k, and where y marks edges. Then we have the following relation.

Lemma 9.25.

$$w\frac{\partial B^\bullet(x,y,w)}{\partial w} = \sum_{k\ge 1} kB_k^\bullet(x,y)w^k = xywe^{S(x,y,w)}.$$

Proof. The sum counts rooted 2-connected graphs with a distinguished edge incident to the root. By definition this is precisely the set of networks containing an edge between the poles. The degree of the root in a 2-connected graph corresponds to the degree of the first pole in the corresponding network, hence the equation follows.

From the previous equation it follows that

$$B^\bullet(x,y,w) = xy \int_0^w e^{S(x,y,s)}\,ds. \tag{9.64}$$

From an analytic point of view it would be sufficient to work with this equation. However, if we are interested in more or less explicit formulas for the degree distribution we have to remove the integral and to express $B^\bullet$ in terms of D. Recall that $E(x,y) = D(x,y,1)$.

Lemma 9.26. *The generating function of rooted 2-connected series-parallel graphs is equal to*

$$B^\bullet(x,y,w) = x\left(D(x,y,w) - \frac{xE(x,y)}{1+xE(x,y)} D(x,y,w)\left(1 + \frac{D(x,y,w)}{2}\right)\right).$$

Proof. We start by using (9.60) to obtain

$$\int e^S\,dw = \int \frac{1+D}{1+yw}\,dw = y^{-1}\log(1+yw) + \int \frac{D}{1+yw}\,dw.$$

Now we integrate by parts and derive at

$$\int \frac{D}{1+yw}\,dw = y^{-1}\log(1+yw)D - \int y^{-1}\log(1+yw)\frac{\partial D}{\partial w}\,dw.$$

For the last integral we change variables $t = D(x,y,w)$ and use the fact that $\log(1+yw) = \log(1+t) - xEt/(1+xE)$. Consequently we obtain

$$\int \log(1+yw)\frac{\partial D}{\partial w}\,dw = \int_0^D \log(1+t)\,dt - \frac{xE}{1+xE}\int_0^D t\,dt.$$

Now everything can be integrated in closed form and, after a simple manipulation, we obtain the result as claimed.

Next we analyse the singular structure of the above generating functions. We start with $E(x,y) = D(x,y,1)$ and $D(x,y,w)$.

Lemma 9.27. *For $|w| \le 1$ and for fixed y (sufficiently close to 1) the dominant singularity of the functions $E(x,y)$, $D(x,y,w)$, and $B^\bullet(x,y,w)$ (considered as functions in x) is given by $x = R(y)$, where $R(y)$ is an analytic function in y with $R = R(1) \approx 0.1280038$. Furthermore, we have the following local expansion:*

$$\begin{aligned} E(x,y) &= E_0(y) + E_1(y)X + E_2(y)X^2 + \cdots, \\ D(x,y,w) &= D_0(y,w) + D_1(y,w)X + D_2(y,w)X^2 + \cdots, \\ B^\bullet(x,y,w) &= B_0(y,w) + B_1(y,w)X + B_2(y,w)X^2 + \cdots, \end{aligned}$$

where $X = \sqrt{1 - x/R(y)}$.

The functions $R(y)$, $E_j(y)$, $D_j(y,w)$, *and* $B_j(y,w)$ *are analytic in* y *and* w *and satisfy the relations*

$$\begin{aligned}
\frac{E_0(y)^3}{E_0(y)-1} &= \left(\log \frac{1+E_0(y)}{1+R(y)} - E_0(y)\right)^2, \\
R(y) &= \frac{\sqrt{1-1/E_0(y)}-1}{E_0(y)}, \\
E_1(y) &= -\left(\frac{2R(y)E_0(y)^2(1+R(y)E_0(y))^2}{(2R(y)E_0(y)+R(y)^2E_0(y)^2)^2+2R(y)(1+R(y)E_0(y))}\right)^{1/2}, \\
D_0(y,w) &= (1+yw)\exp\left(\frac{R(y)E_0(y)}{1+R(y)E_0(y)}D_0(y,w)\right)-1, \\
D_1(y,w) &= -\frac{D_0(y,w)E_1(y)R(y)(D_0(y,w)+1)}{(R(y)E_0(y)D_0(y,w)-1)(1+R(y)E_0(y))}, \\
B_0(y,w) &= -\frac{R(y)D_0(y,w)(R(y)E_0(y)D_0(y,w)-2)}{2(1+R(y)E_0(y))}, \\
B_1(y,w) &= \frac{E_1(y)R(y)^2D_0(y,w)^2}{2(1+R(y)E_0(y))^2}.
\end{aligned}$$

Proof. Since $E(x,y)$ satisfies Equation (9.62), it follows that the dominant singularity of the mapping $x \mapsto E(x,y)$ is of square-root type and there is an expansion of the form $E(x,y) = E_0(y) + E_1(y)X + E_2(y)X^2 + \cdots$ (with $X = \sqrt{1-x/R(y)}$), where $R(y)$ and $E_j(y)$ are analytic in y. Furthermore, if we set

$$\Phi(x,y,z) = (1+y)\exp\left(\frac{xz^2}{1+xz}\right) - 1$$

then $R(y)$ and $E_0(y)$ satisfy the two equations

$$\Phi(R(y),y,E_0(y)) = E_0(y) \quad \text{and} \quad \Phi_z(R(y),y,E_0(y)) = 0$$

and $E_1(y)$ is then given by

$$E_1(y) = -\left(\frac{2R(y)\Phi_x(R(y),y,E_0(y))}{\Phi_{zz}(R(y),y,E_0(y))}\right)^{1/2}$$

(compare with Theorem 2.19).

Next notice that for $|w| \le 1$ the radius of convergence of the function $x \mapsto D(x,y,w)$ is surely $\ge |R(y)|$. However, $D(x,y,w)$ satisfies Equation (9.63), which implies that the dominant singularity of $E(x,y)$ leads to that of $D(x,y,w)$. Thus, the mapping $x \mapsto D(x,y,w)$ has dominant singularity $R(y)$ and it also follows that $D(x,y,w)$ has a singular expansion of the form $D(x,y,w) = D_0(y,w)+D_1(y,w)X+D_2(y,w)X^2+\cdots$. Hence, by Lemma 9.25 we also get an expansion for $B^\bullet(x,y,w)$ of that form.

Finally the relations for D_0, D_1 and B_0, B_1 follow by comparing coefficients in the corresponding expansions.

Now we are able to characterise the degree distribution for 2-connected series parallel graphs.

Theorem 9.28. *Let d_k be the limit probability that a vertex of a 2-connected series-parallel graph has degree k. Then*

$$p(w) = \sum_{k \geq 1} d_k w^k = \frac{B_1(1,w)}{B_1(1,1)}.$$

Obviously, $p(1) = 1$, so that the d_k are indeed a probability distribution. We have asymptotically, as $k \to \infty$,

$$d_k \sim c \cdot k^{-3/2} q^k,$$

where $c \approx 3.7340799$ is a computable constant and

$$q = \left((1 + 1/(R(1)E_0(1)))\, e^{-1/(1+R(1)E_0(1))} - 1 \right)^{-1} \approx 0.7620402.$$

Proof. First observe that

$$p(w) = \lim_{n \to \infty} \frac{[x^n]B^\bullet(x,1,w)}{[x^n]B^\bullet(x,1,1)}.$$

However, from the local expansion of $B^\bullet(x,1,w)$ that is given in Lemma 9.27 (and by the fact that $B^\bullet(x,1,w)$ can be analytically continued to a Δ-region) it follows that

$$[x^n]B^\bullet(x,1,w) = -\frac{B_1(1,w)}{2\sqrt{\pi}} n^{-3/2} R(1)^{-n} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Hence, $p(w) = B_1(w)/B_1(1)$.

Next observe that Lemma 9.27 provides $B_1(1,w)$ only for $|w| \leq 1$. However, it is easy to continue $B_1(1,w)$ analytically to a larger region and it is also possible to determine the dominant singularity of $B_1(1,w)$, from which we deduce an asymptotic relation for the coefficients of $p(w) = B_1(1,w)/B_1(1,1)$.

For this purpose first observe from Lemma 9.27 that $D_0(y,w)$ satisfies a functional equation which provides an analytic continuation of the mapping $w \mapsto D_0(y,w)$ to a region including the unit disc. Furthermore, it follows that there exists a dominant singularity $w_0(y)$ and a local expansion of the form

$$D_0(y,w) = D_{00}(y) + D_{01}(y)W + D_{02}(y)W^2 + \cdots,$$

where $W = \sqrt{1 - w/w_0(y)}$. Besides, if we set

$$\Psi(y,w,z) = (1 + yw) \exp\left(\frac{R(y)E_0(y)}{1 + R(y)E_0(y)} z\right) - 1$$

then $w_0(y)$ and $D_{00}(y)$ satisfy the equations

$$\Psi(y, w_0(y), D_{00}(y)) = D_{00}(y) \quad \text{and} \quad \Psi_z(y, w_0(y), D_{00}(y)) = 0.$$

Hence

$$D_{00}(y) = \frac{1}{R(y)E_0(y)}$$

and

$$w_0(y) = \frac{1}{y}\left(1 + \frac{1}{R(y)E_0(y)}\right)\exp\left(-\frac{1}{1+R(y)E_0(y)}\right) - \frac{1}{y}.$$

Finally, with the help of Lemma 9.27 it also follows that this local representation of $D_0(y, w)$ provides similar local representations for D_1, B_0, and B_1:

$$\begin{aligned} D_1(y, w) &= D_{1,-1}(y)W^{-1} + D_{10}(y) + D_{11}(y)W + \cdots, \\ B_0(y, w) &= B_{00}(y) + B_{02}(y)W^2 + B_{03}(y)W^3 + \cdots, \\ B_1(y, w) &= B_{10}(y) + B_{11}(y)W + B_{12}(y)W^2 + \cdots, \end{aligned}$$

where $W = \sqrt{1 - w/w_0(y)}$ is as above. Hence, since all functions of interest (D_0, D_1, B_0, B_1) can be analytically continued to a Δ-region, the asymptotic relation for d_k follows immediately. Since $w_0(1)$ is the dominant singularity, we have $q = 1/w_0(1)$.

The next theorem provides the degree distribution in connected (and all) series-parallel graphs.

Theorem 9.29. *For every $k \ge 1$ the limiting probability d_k that a vertex of a connected (or general) labelled series-parallel graph has degree k exists and we have*

$$p(w) = \sum_{k\ge 1} d_k w^k = \rho \cdot \frac{\partial}{\partial x}\, e^{B^\bullet(x,1,w)} \Big|_{x=\rho C'(\rho)},$$

where $B^\bullet$ is given by Lemma 9.26 and equations (9.62) and (9.63).

Moreover $p(1) = 1$, so that the d_k are indeed a probability distribution. We have asymptotically, as $k \to \infty$,

$$d_k \sim c \cdot k^{-3/2} q^k,$$

where $c \approx 3.5952391$ is a computable constant and

$$q = \left((1 + 1/(\rho C'(\rho)E(\rho C'(\rho), 1)))\, e^{-1/\tau E(\rho C'(\rho),1)} - 1 \right)^{-1} \approx 0.7504161.$$

Proof. The proof of the first statement is exactly the same as for Theorem 9.17. Again, we know that $\rho C'(\rho) = \tau \approx 0.127$ is larger than the radius of convergence $\rho \approx 0.110$ of $C(x)$, so that Lemma 2.26 applies. The proof that $p(1) = 1$ is also the same.

Recall that $\tau < R(1)$. Hence the dominant singularity $x = R(1)$ of the mapping $x \mapsto B^\bullet(x, 1, w)$ will have no influence to the analysis of $p(w)$. Nevertheless, since

$$\frac{\partial}{\partial x} e^{B^\bullet(x,1,w)} = e^{B^\bullet(x,1,w)} \frac{\partial B^\bullet(x,1,w)}{\partial x},$$

we have to get some information on $D(x,1,w)$ and its derivative $\partial D(x,1,w)/\partial x$ with $x = \tau$.

Let us start with the analysis of the mapping $w \mapsto D(\tau,1,w)$. Since $D(x,y,w)$ satisfies Equation (9.63), it follows that $D(\tau,1,w)$ satisfies

$$D(\tau,1,w) = (1+w)\exp\left(\frac{\tau E(\tau,1)D(\tau,1,w)}{1+\tau E(\tau,1)}\right) - 1.$$

Hence there exists a dominant singularity w_1 and a singular expansion of the form

$$D(\tau,1,w) = \widetilde{D}_0 + \widetilde{D}_1\widetilde{W} + \widetilde{D}_2\widetilde{W}^2 + \cdots,$$

where $\widetilde{W} = \sqrt{1 - w/w_1}$. Furthermore, if we set

$$\varXi(w,z) = (1+w)\exp\left(\frac{\tau E(\tau,1)z}{1+\tau E(\tau,1)}\right) - 1$$

then w_1 and $\widetilde{D}_0$ satisfy the equations

$$\varXi(w_1,\widetilde{D}_0) = \widetilde{D}_0 \quad \text{and} \quad \varXi_z(w_1,\widetilde{D}_0) = 0.$$

Consequently,

$$\widetilde{D}_0 = \frac{1}{\tau E(\tau,1)} \quad \text{and} \quad w_1 = \left(1 + \frac{1}{\tau E(\tau,1)}\right)\exp\left(-\frac{1}{1+\tau E(\tau,1)}\right) - 1.$$

Next, by taking derivatives with respect to x in (9.63), we obtain the relation

$$\frac{\partial D(x,1,w)}{\partial x} = \frac{(1+D(x,1,w))D(x,1,w)(E(x,1)+xE_x(x,1))}{(xE(x,1)D(x,1,w)-1)(1+xE(x,1))}.$$

Thus if we set $x = \tau$ and insert the singular representation of $D(\tau,1,w)$ it follows that $\frac{\partial D(x,1,w)}{\partial x}\big|_{x=\tau}$ has a corresponding singular representation, too. By Lemma 9.26 we get the same property for $\frac{\partial B^\bullet(x,1,w)}{\partial x}\big|_{x=\tau}$ and finally for

$$\rho \cdot \frac{\partial}{\partial x} e^{B^\bullet(x,1,w)}\big|_{x=\tau} = \widetilde{C}_0 + \widetilde{C}_1\widetilde{W} + \widetilde{C}_2\widetilde{W}^2 + \cdots.$$

This implies the asymptotic relation for d_k with $q = 1/w_1$.

In this case, we obtain an expression for $p(w)$ in terms of the functions $E(x,1)$ and $D(x,1,w)$ and their derivatives. The derivatives can be computed using Equations (9.62) and (9.63) as in the previous proof.

9.4.2 Vertices of Given Degree in Series-Parallel Networks

Our next purpose is to study the distribution of the number of vertices of given degree in a random series-parallel network. Recall that the generating function $D(x, y)$ and $S(x, y)$ for all series-parallel networks and series-parallel networks that have a series decomposition satisfy the system of equations (9.13)–(9.14):

$$D(x, y) = (1 + y)e^{S(x,y)} - 1, \tag{9.65}$$

$$S(x, y) = (D(x, y) - S(x, y))xD(x, y). \tag{9.66}$$

In particular, it follows that $D(x, y)$ satisfies the equation

$$\log\left(\frac{1 + D(x, y)}{1 + y}\right) = \frac{xD(x, y)^2}{1 + xD(x, y)}. \tag{9.67}$$

We recall that the first equation (9.65) expresses the fact that a series-parallel network is a parallel composition of series networks (this is the exponential term), to which we may add or not the edge connecting the two poles. The second equation (9.66) means that a series network is formed by taking first a non-series network (this is the term $D - S$), and concatenating to it an arbitrary network. Since two of the poles are identified, a new internal vertex is created, hence the factor x.

As in the case of outer-planar graphs we will extend these relations to generating functions where we take into account the vertex degrees. We fix some $k \ge 2$ and define by

$$d_{i,j;m,n;n_1,n_2,\ldots,n_k,n_\infty}$$

the number of series-parallel networks with $2 + n = 2 + n_1 + n_2 + \cdots + n_k + n_\infty \ge 3$ vertices and m edges such that the poles have degrees i and $j \in \{1, 2, \ldots, k, \infty\}$[6] and that for $1 \le \ell \le k$ there are exactly n_ℓ internal vertices of degree ℓ, and there are n_∞ internal vertices with degree $> k$. The corresponding generating functions are then defined by

$$\begin{aligned} &D_{i,j}(x, y, z_1, z_2, \ldots, z_k, z_\infty) \\ &= \sum_{m,n,n_1,\ldots,n_k,n_\infty} d_{i,j;m,n;n_1,n_2,\ldots,n_k,n_\infty} y^m \frac{x^n z_1^{n_1} \cdots z_k^{n_k} z_\infty^{n_\infty}}{n!}. \end{aligned} \tag{9.68}$$

Similarly we define

$$\begin{aligned} &S_{i,j}(x, y, z_1, z_2, \ldots, z_k, z_\infty) \\ &= \sum_{m,n,n_1,\ldots,n_k,n_\infty} s_{i,j;m,n;n_1,n_2,\ldots,n_k,n_\infty} y^m \frac{x^n z_1^{n_1} \cdots z_k^{n_k} z_\infty^{n_\infty}}{n!} \end{aligned} \tag{9.69}$$

---

[6] Again infinite degree means degree greater than k.

where we count series-parallel networks that have a series decomposition into at least two series-parallel networks.

The next lemma provides a system of equations for $D_{i,j}$ and $S_{i,j}$. Again, in order to state the results in a more compact form we use the convention that ∞ means $> k$ and $\infty - 1$ means $> k-1$, in particular we set $\ell + \infty = \infty$ for all positive integers ℓ.

Lemma 9.30. *The generating functions $D_{ij} = D_{i,j}(x, y, z_1, z_2, \ldots, z_k, z_\infty)$ and $S_{ij} = S_{i,j}(x, y, z_1, z_2, \ldots, z_k, z_\infty)$, $i, j \in \{1, \ldots, k, \infty\}$, satisfy the following system of equations:*

$$D_{i,j} = \sum_{r\ge 1} \sum_{i_1+\cdots+i_r=i} \sum_{j_1+\cdots+j_r=j} \frac{1}{r!} \prod_{\ell=1}^{r} S_{i_\ell,j_\ell} \tag{9.70}$$

$$+ y \sum_{r\ge 1} \sum_{i_1+\cdots+i_r=i-1} \sum_{j_1+\cdots+j_r=j-1} \frac{1}{r!} \prod_{\ell=1}^{r} S_{i_\ell,j_\ell}$$

$$S_{i,j} = x \sum_{\ell_1+\ell_2\le k} (D_{i,\ell_1} - S_{i,\ell_1}) z_{\ell_1+\ell_2} D_{\ell_2,j} + x z_\infty \sum_{\ell_1+\ell_2>k} (D_{i,\ell_1} - S_{i,\ell_1}) D_{\ell_2,j}. \tag{9.71}$$

Proof. This is a refinement of the counting procedure that leads to the system (9.65) and (9.66). The first equation means that a series-parallel network with degrees i and j at the poles is obtained by parallel composition of series networks whose degrees at the left and right pole sum up to i and j, respectively; one has to distinguish according to whether the edge between the poles is added or not.

The second equation reflects the series composition. In this case only the degrees of the right pole in the first network and of the left pole in the second network have to be added.

Remark 9.31 *The system provided in Lemma 9.30 is not a positive system since the equation for $S_{i,j}$ contains negative terms. However, we can replace the term D_{i,ℓ_1} by the right hand side of (9.70). Further, note that S_{i,ℓ_1} appears in this sum so that we really achieve a positive system.*

Finally, it is easy to see that this (new) system is strongly connected.

9.4.3 Vertices of Given Degree in 2-Connected Series-Parallel Graphs

The next theorem concerns 2-connected series-parallel graphs.

Theorem 9.32. *For $k \ge 2$, let $X_n^{(k)}$ denote the number of vertices of degree k in a random 2-connected labelled series-parallel graph with n vertices. Then $X_n^{(k)}$ satisfies a central limit theorem with linear expected value and variance.*

Let $b_{n,m}$ be the number of 2-connected vertex labelled series-parallel graphs with n vertices and m edges and let

$$B(x,y) = \sum_{n,m} b_{n,m} \frac{x^n}{n!} y^m$$

be the corresponding exponential generating function. By the definition of series-parallel networks we have

$$\frac{\partial B(x,y)}{\partial y} = \frac{x^2}{2} e^{S(x,y)}. \tag{9.72}$$

Note that $2\frac{\partial B(x,y)}{\partial y}$ can be interpreted as the generating function of 2-connected series-parallel graphs with a rooted and directed edge. We now also take into account the degrees of the vertices. In particular let $B_{i,j} = B_{i,j}(x, y, z_1, \ldots, z_k, z_\infty)$, $i, j \in \{2, \ldots, k, \infty\}$, denote the exponential generating function of 2-connected series-parallel graphs with a directed rooted edge, where the two root vertices have degrees i and j.[7] The directed root edge connects the root vertex of degree i with the other root vertex of degree j.

Furthermore, let $B_i = B_i(x, y, z_1, \ldots, z_k, z_\infty)$, $i \in \{2, \ldots, k, \infty\}$, be the generating function of 2-connected series-parallel graphs where we just root at one vertex that has degree i (and where no edge is rooted). Finally, let $B = B(x, y, z_1, \ldots, z_k, z_\infty)$ be the generating function of all 2-connected series-parallel graphs.

The next lemma quantifies the relation between series-parallel networks and 2-connected series-parallel graphs. As above we use the convention that ∞ means $> k$ and $\infty - 1$ means $\geq k - 1$, in particular we set $\ell + \infty = \infty$ for all positive integers ℓ.

Lemma 9.33. *We have*

$$B_{i,j} = x^2 z_i z_j y \sum_{r \geq 1} \sum_{i_1 + \cdots + i_r = i-1} \sum_{j_1 + \cdots + j_r = j-1} \frac{1}{r!} \prod_{\ell=1}^{r} S_{i_\ell, j_\ell},$$

$$B_i = \frac{1}{i} \sum_{j=2}^{k} B_{i,j} + \frac{1}{i} B_{i,\infty} \qquad (i \in \{2, \ldots, k\})$$

$$B_\infty = x \frac{\partial B}{\partial x} - \sum_{i=2}^{k} B_i$$

$$2y \frac{\partial B}{\partial y} = \sum_{i,j \in \{2,\ldots,k,\infty\}} B_{i,j},$$

$$x \frac{\partial B}{\partial x} = \sum_{i \in \{2,\ldots,k,\infty\}} B_i.$$

[7] As above we interpret ∞ as $> k$.

Proof. The first equation is essentially a refinement of the relation (9.72). It means that 2-connected series-parallel graphs are formed by taking parallel compositions of series-parallel networks and adding the edge between the poles. The sum of the degrees of the poles in the networks must be one less than the degree of the resulting vertex in the series-parallel graph.

The second equation reflects the fact that a series-parallel graph rooted at a vertex of degree i comes from a series-parallel graph rooted at an edge whose first vertex has degree i, but then each of them has been counted i times. The remaining equations are clear if we recall that $x\partial B/\partial x$ corresponds to graphs rooted at a vertex and $y\partial B/\partial y$ corresponds to graphs rooted at an edge (in this last case the factor 2 appears because in the definition of $B_{i,j}$ the root edge is directed).

Let

$$B_{d=k}(x,u) = \sum_{n,\nu} b_{n,\nu}^{(k)} \frac{x^n}{n!} u^\nu$$

denote the exponential generating function for the numbers $b_{n,\nu}^{(k)}$ that count the number of 2-connected vertex labelled series-parallel graphs with n vertices, where ν vertices have degree k. Then we have

$$B_{d=k}(x,u) = B(x,1,1,\ldots,1,u,1).$$

We are now ready to prove Theorem 9.32. We first observe that the function D_{ij} and S_{ij} satisfy a positive and strongly connected system of equations (Lemma 9.30), so that by Theorem 2.19 all these functions have a common square-root singularity of the kind

$$g(x,y,z_1,\ldots,z_k,z_\infty) - h(x,y,z_1,\ldots,z_k,z_\infty)\sqrt{1-\frac{x}{\rho(y,z_1,\ldots,z_k,z_\infty)}}. \tag{9.73}$$

Hence, by Lemma 2.26 the same is true for all functions B_{ij} and consequently for

$$2y\frac{\partial B}{\partial y} = \sum_{i,j\in\{2,\ldots,k,\infty\}} B_{ij}.$$

From Theorem 2.30 it follows that B has a singularity of the kind

$$\overline{g}(x,y,z_1,\ldots,z_k,z_\infty) + \overline{h}(x,y,z_1,\ldots,z_k,z_\infty)\left(1-\frac{x}{\rho(z_1,\ldots,z_k,z_\infty)}\right)^{3/2}. \tag{9.74}$$

As a consequence, by applying Theorem 2.25 we derive Theorem 9.32 for the 2-connected case.

Note that $B_1,\ldots B_k$ also have a square-root singularity of the kind (9.73), since they can be expressed with the help of the function B_{ij}. Interestingly, B_∞ has the same kind of singularity. This follows from the formula

$$B_\infty = x\frac{\partial B}{\partial x} - \sum_{i=1}^{k} B_i,$$

since $\frac{\partial B}{\partial x}$ also has a square-root singularity of the kind (9.73).

9.4.4 Vertices of Given Degree in Connected Series-Parallel Graphs

Finally we also have a central limit theorem for the number of vertices of given degree in connected (and all) series-parallel graphs.

Theorem 9.34. *For $k \ge 2$, let $X_n^{(k)}$ denote the number of vertices of degree k in a random connected (or general) labelled series-parallel graph with n vertices. Then $X_n^{(k)}$ satisfies a central limit theorem with linear expected value and variance.*

For the proof we introduce the generating function

$$C_j(x, y, z_1, z_2, \ldots, z_k, z_\infty) = \sum_{m,n,n_1,\ldots,n_k,n_\infty} c_{j;m,n;n_1,\ldots,n_k,n_\infty}\, y^m\, \frac{x^n z_1^{n_1}\cdots z_k^{n_k} z_\infty^{n_\infty}}{n!},$$

$j \in \{1, 2, \ldots, k, \infty\}$, where $c_{j;m,n,n_1,\ldots,n_k,n_\infty}$ is the number of labelled series parallel graphs with $n = n_1 + \cdots + n_k + n_\infty$ vertices and m edges, where one vertex of degree j is marked and where n_ℓ of these n vertices have degree ℓ, $1 \le \ell \le k$, and n_∞ of these vertices have degree greater than k. Further set

$$B_j^\bullet(x, y, z_1, z_2, \ldots, z_k, z_\infty) = \frac{1}{xz_j} B_j(x, y, z_1, z_2, \ldots, z_k, z_\infty)$$

and

$$C_j^\bullet(x, y, z_1, z_2, \ldots, z_k, z_\infty) = \frac{1}{xz_j} C_j(x, y, z_1, z_2, \ldots, z_k, z_\infty).$$

Then $B_j^\bullet$ and $C_j^\bullet$ have the same interpretation as in the case of outerplanar graphs. Hence we get the same relations as stated in Lemma 9.23. The only difference is that we also take the number of edges into account, that is, we have an additional variable y.

Lemma 9.35. *Let $W_j = W_j(z_1, \ldots, z_k, z_\infty, C_1^\bullet, \ldots, C_k^\bullet, C_\infty^\bullet)$, $j \in \{1, 2, \ldots, k, \infty\}$, be defined by*

$$W_j = \sum_{i=0}^{k-j} z_{i+j} C_i^\bullet(x, y, z_1, \ldots, z_k, z_\infty) + z_\infty \left(\sum_{i=k-j+1}^{k} C_i^\bullet(x, y, z_1, \ldots, z_k, z_\infty) + C_\infty^\bullet(x, y, z_1, \ldots, z_k, z_\infty) \right),$$

$$(1 \le j \le k),$$

$$W_\infty = z_\infty \left(\sum_{i=0}^{k} C_i^\bullet(x,y,z_1,\ldots,z_k,z_\infty) + C_\infty^\bullet(x,y,z_1,\ldots,z_k,z_\infty) \right).$$

Then the functions $C_1^\bullet, \ldots, C_k^\bullet, C_\infty^\bullet$ *satisfy the system of equations*

$$\begin{aligned} C_j^\bullet(x,y,z_1,\ldots,&z_k,z_\infty) \\ &= \sum_{\ell_1+2\ell_2+3\ell_3+\cdots+j\ell_j=j} \prod_{r=1}^{j} \frac{B_r^\bullet(x,y,W_1,\ldots,W_k,W_\infty)^{\ell_r}}{\ell_r!} \\ &\qquad (1 \le j \le k), \\ C_\infty^\bullet(x,y,z_1,\ldots,&z_k,z_\infty) = \exp\left(\sum_{j=1}^{k} B_j^\bullet(x,y,W_1,\ldots,W_k,W_\infty) \right. \\ &\left. + B_\infty^\bullet(x,y,W_1,\ldots,W_k,W_\infty) \right) - 1 \\ &- \sum_{1\le \ell_1+2\ell_2+3\ell_3+\cdots+k\ell_k \le k} \prod_{r=1}^{k} \frac{B_r^\bullet(x,y,W_1,\ldots,W_k,W_\infty)^{\ell_r}}{\ell_r!}. \end{aligned}$$

Consequently the generating function $C_{d=k}(x,u)$ (of the numbers $c_{n,\nu}^{(k)}$ that count the number of connected vertex labelled series-parallel graphs with n vertices, where ν vertices have degree k) is given by

$$\begin{aligned} \frac{\partial C_{d=k}(x,u)}{\partial x} &= \sum_{j=1}^{k-1} C_j^\bullet(x,1,1,\ldots,1,u,1) + u C_k^\bullet(x,1,1,\ldots,1,u,1) \\ &+ C_\infty^\bullet(x,1,1,\ldots,1,u,1) \end{aligned}$$

and the generation function $G_{d=k}(x,u)$ for all series-parallel graphs by

$$G_{d=k}(x,u) = e^{C_{d=k}(x,u)}.$$

The proof of Theorem 9.34 is exactly the same as in the outerplanar case. The essential point is to observe that the singularity of $B_i^\bullet$ does not interfere with the singularity of $C_j^\bullet$. Thus, we can safely apply Theorem 2.35.

As in the outerplanar case it follows that $C_{d=k}(x,u)$ has a singularity of the kind (9.58). Hence, the same follows for $G_{d=k}(x,u)$ and, thus, we can apply Theorem 2.25.

9.5 All Planar Graphs

The goal of this section is to characterise the degree distribution of labelled planar graphs. For this purpose we have to find the generating function of

3-connected planar graphs according to the degree of the root. We will also discuss the distribution of the number of vertices of degree $k = 1$ and $k = 2$. The case $k \geq 3$ cannot be covered with the present methodology since the counting procedure for 3-connected planar graphs that is used here cannot be extended in order to take care of all vertices of given degree.

From now on (and for the rest of Chapter 9) all generating functions are associated to planar graphs.

9.5.1 The Degree of a Rooted Vertex

We will now repeat (more or less) the counting procedure for labelled planar graphs, where we also take care of the degree of a rooted vertex.

Theorem 9.36. *Let $b^\bullet_{n,m,k}$ denote the number of vertex-rooted 2-connected labelled planar graphs with n vertices, m edges and root-degree k, and $c^\bullet_{n,m,k}$ and $g^\bullet_{n,m,k}$ the corresponding numbers for connected and all vertex-rooted planar graphs. Then the corresponding generating functions*

$$B^\bullet(x,y,w) = \sum_{n,m,k} b^\bullet_{n,m,k} \frac{x^n}{n!} y^m z^k,$$

$$C^\bullet(x,y,w) = \sum_{n,m,k} c^\bullet_{n,m,k} \frac{x^n}{n!} y^m z^k,$$

$$G^\bullet(x,y,w) = \sum_{n,m,k} g^\bullet_{n,m,k} \frac{x^n}{n!} y^m z^k$$

are determined by the following system of equations:

$$G^\bullet(x,y,w) = \exp\left(C(x,y,1)\right) C^\bullet(x,y,w), \tag{9.75}$$

$$C^\bullet(x,y,w) = \exp\left(B^\bullet\left(xC^\bullet(x,y,1),y,w\right)\right), \tag{9.76}$$

$$\frac{\partial B^\bullet(x,y,w)}{\partial w} = xy\frac{1+D(x,y,w)}{1+yw} \tag{9.77}$$

$$\frac{1}{x^2 D(x,y,w)} T^\bullet\left(x, D(x,y,1), \frac{D(x,y,w)}{D(x,y,1)}\right) = \log\left(\frac{1+D(x,y,w)}{1+yw}\right) - \frac{xD(x,y,w)D(x,y,1)}{1+xD(x,y,1)}, \tag{9.78}$$

$$T^\bullet(x,y,w) = \frac{x^2y^2w^2}{2}\left(\frac{1}{1+wy} + \frac{1}{1+xy} - 1 - \frac{(U+1)^2\left(-w_1(U,V,w) + (U-w+1)\sqrt{w_2(U,V,w)}\right)}{2w(Vw+U^2+2U+1)(1+U+V)^3}\right), \tag{9.79}$$

$$U(x,y) = xy(1+V(x,y))^2, \quad V(x,y) = y(1+U(x,y))^2 \tag{9.80}$$

with polynomials $w_1 = w_1(U,V,w)$ and $w_2 = w_2(U,V,w)$ given by

$$w_1 = -UVw^2 + w(1 + 4V + 3UV^2 + 5V^2 + U^2 + 2U + 2V^3 + 3U^2V + 7UV) + (U+1)^2(U + 2V + 1 + V^2),$$
$$w_2 = U^2V^2w^2 - 2wUV(2U^2V + 6UV + 2V^3 + 3UV^2 + 5V^2 + U^2 + 2U + 4V + 1) + (U+1)^2(U + 2V + 1 + V^2)^2.$$

Remark 9.37 *Observe that the above system looks very similar to the system of Theorem 9.12. However, the third equation (9.77) contains a partial derivative with respect to w instead of a partial derivative with respect y.*

Proof. The first two equations (9.75) and (9.76) follow the general principles between 2-connected, connected and all graphs. The remaining equations concern 3-connected planar graphs and the relation between 2- and 3-connected ones.

Let $T^\bullet(x, z, w)$ denote the generating function of edge rooted 3-connected planar graphs, where the rooted edge is directed, and where (the exponent of) x counts vertices, z counts edges, and w counts the degree of the tail of the root edge. Now we relate $T^\bullet$ to the generating function $Q(X, Y, W)$ of simple quadrangulations that were defined in (9.23). By the bijection between simple quadrangulations and 3-connected planar maps, and using Euler's relation, the generating function $xwQ(xz, z, w)$ counts rooted 3-connected planar maps, where z marks edges (we have added an extra term w to correct the 'minus one' in the definition of Q).

By Whitney's theorem 3-connected planar graphs have a unique embedding on the sphere. There are two ways of rooting an embedding of a directed edge-rooted graph in order to get a rooted map, since there are two ways of choosing the root face adjacent to the root edge. Thus, we have

$$T^\bullet(x, z, w) = \frac{xw}{2} Q(xz, z, w) \tag{9.81}$$

and consequently

$$T^\bullet(x, y, w) = \frac{x^2y^2w^2}{2}\left(\frac{1}{1+wy} + \frac{1}{1+xy} - 1 - \frac{(U+1)^2\left(-w_1(U, V, w) + (U - w + 1)\sqrt{w_2(U, V, w)}\right)}{2w(Vw + U^2 + 2U + 1)(1 + U + V)^3}\right). \tag{9.82}$$

We only have to combine equation (9.81), together with (9.31), (9.32), and (9.35).

Next we denote by $D(x, y, w)$ and $S(x, y, w)$, respectively, the generating functions of (planar) networks and series networks, with the same meaning for the variables x, y and w (as in Section 9.4.1). Here we obtain

$$\frac{1}{x^2D(x, y, w)} T^\bullet\left(x, E(x, y), \frac{D(x, y, w)}{E(x, y)}\right) = \log\left(\frac{1 + D(x, y, w)}{1 + yw}\right) - S(x, y, w) \tag{9.83}$$

and

$$S(x,y,w) = xE(x,y)\left(D(x,y,w) - S(x,y,w)\right), \tag{9.84}$$

where $E(x,y) = D(x,y,1)$ is the generating function for planar networks (without marking the degree of the root). Of course, these two equations simplify to (9.78). The relations (9.83)–(9.84) follow from a slight extension of the proof of the equations (9.43)–(9.47), taking into account the degree of the first pole in a network. The main point is the substitution of variables in $T^\bullet$: An edge is substituted by an ordinary planar network (this accounts for the term $E(x,y)$), except if it is incident with the first pole, in which case it is substituted by a planar network marking the degree, hence the term $D(x,y,w)$ (it is divided by $E(x,y)$ in order to avoid over-counting of ordinary edges).

As in Lemma 9.25, and for the same reason, we have

$$\begin{aligned} w\frac{\partial B^\bullet(x,y,w)}{\partial w} &= \sum_{k\ge 1} kB_k^\bullet(x,y)w^k \\ &= xyw\exp\Bigg(S(x,y,w) \\ &\qquad + \frac{1}{x^2D(x,y,w)}T^\bullet\left(x, E(x,y), \frac{D(x,y,w)}{E(x,y)}\right)\Bigg) \\ &= xyw\frac{1+D(x,y,w)}{1+yw}. \end{aligned}$$

This proves (9.77) and completes the proof of Theorem 9.36.

In principal, the above relations are sufficient for the proof of the existence of a degree distribution. However, if we want to obtain it explicitly we have to get rid of some partial derivatives.

Lemma 9.38. *The generating function of rooted 2-connected planar graphs is equal to*

$$\begin{aligned} B^\bullet(x,y,w) = x\left(D - \frac{xED}{1+xE}\left(1+\frac{D}{2}\right)\right) - \frac{1+D}{xD}T^\bullet(x,E,D/E) \\ + \frac{1}{x}\int_0^D \frac{T^\bullet(x,E,t/E)}{t}\,dt, \end{aligned} \tag{9.85}$$

where $D = D(x,y,w)$ *and* $E = E(x,y)$ *and*

$$\int_0^w \frac{T^\bullet(x,z,t)}{t}\,dt = -\frac{x^2(z^3xw^2 - 2wz - 2xz^2w + (2+2xz)\log(1+wz))}{4(1+xz)}$$

$$
\begin{aligned}
&- \frac{UVx}{2(1+U+V)^3} \left(\frac{w(2U^3 + (6V+6)U^2 + (6V^2 - Vw + 14V + 6)U)}{4V(V+1)^2} \right. \\
&+ \frac{4V^3 + 10V^2 + 8V + 2}{4V(V+1)^2} + \frac{(1+U)(1+U+2V+V^2)}{4UV^2(V+1)^2} \\
&\quad \times (2U^3 + (4V+5)U^2 + (3V^2 + 8V + 4)U + 2V^3 + 5V^2 + 4V + 1) \\
&- \frac{\sqrt{Q}(2U^3 + (4V+5)U^2 + (3V^2 - Vw + 8V + 4)U + 5V^2 + 2V^3 + 4V + 1)}{4UV^2(V+1)^2} \\
&+ \frac{(1+U)^2(1+U+V)^3 \log(Q_1)}{2V^2(1+V)^2} \\
&\left. + \frac{(U^3 + 2U^2 + U - 2V^3 - 4V^2 - 2V)(1+U+V)^3 \log(Q_2)}{2V^2(1+V)^2 U} \right),
\end{aligned}
$$

with expressions Q, Q_1 *and* Q_2 *that are given by*

$$
\begin{aligned}
Q &= U^2V^2w^2 - 2UVw(U^2(2V+1) + U(3V^2 + 6V + 2) + 2V^3 + 5V^2 \\
&\quad + 4V + 1) + (1+U)^2(U + (V+1)^2)^2, \\
Q_1 &= \frac{1}{2(wV + (U+1)^2)^2(V+1)(U^2 + U(V+2) + (V+1)^2)} \\
&\quad \times \big(-UVw(U^2 + U(V+2) + 2V^2 + 3V + 1) + (U+1)(U+V+1)\sqrt{Q} \\
&\quad\quad + (U+1)^2(2U^2(V+1) + U(V^2 + 3V + 2) + V^3 + 3V^2 + 3V + 1\big), \\
Q_2 &= \frac{-wUV + U^2(2V-1) + U(3V^2 + 6V + 2) + 2V^3 + 5V^2 + 4V + 1 - \sqrt{Q}}{2V(U^2 + U(V+2) + (V+1)^2)}.
\end{aligned}
$$

Proof. We start as in the proof of Lemma 9.26.

$$
\int \frac{1+D}{1+yw}\, dw = y^{-1}\log(1+yw) + y^{-1}\log(1+yw)D - \int y^{-1}\log(1+yw)\frac{\partial D}{\partial w}\, dw.
$$

For the last integral we change variables $t = D(x,y,w)$ and use the fact that

$$
\log(1+yw) = \log(1+D) - \frac{xED}{(1+xE)} - \frac{1}{x^2 D} T^\bullet(x, E, D/E).
$$

We obtain

$$
\begin{aligned}
\int \log(1+yw)\frac{\partial D}{\partial w}\, dw = \int_0^D \log(1+t)\, dt - \frac{xE}{1+xE}\int_0^D t\, dt \\
+ \frac{1}{x^2}\int_0^D \frac{T^\bullet(x, E, t/E)}{t} dt.
\end{aligned}
$$

On the right-hand side, all the integrals except the last one are elementary. Now we use

$$
B^\bullet(x,y,w) = xy \int \frac{1+D}{1+yw}\, dw
$$

and after a simple manipulation (9.85) follows.

Next we use Equation (9.82) to integrate $T^\bullet(x,z,w)/w$. Notice that neither U nor V have any dependence on w.

A key point in the previous derivation is, that, by expressing $w(X,Y,W)$ (see Equation (9.34)) in terms of R,S instead of X,Y (see Equation (9.35)), we obtain a quadratic polynomial $w_2(R,S,W)$ in terms of W inside the square root of $T^\bullet(x,z,w)$ in Equation (9.82). Otherwise, we would have obtained a cubic polynomial inside the square root, and the integration would have been much harder.

With Lemma 9.38 we can produce an explicit (although quite long) expression for $B^\bullet(x,y,w)$ in terms of $D(x,y,w)$, $E(x,y)$, and the algebraic functions $U(x,y), V(x,y)$. This is needed in the next section for computing the singular expansion of $B^\bullet(x,y,w)$ at its dominant singularity.

9.5.2 Singular Expansions

In this section we find singular expansions of $T^\bullet(x,z,w), D(x,y,w)$ and $B^\bullet(x,y,w)$ at their dominant singularities. As we show here, these singularities do not depend on w and were found in [9] and [91]. However the coefficients of the singular expansions do depend on w, and our task is to compute them exactly in each case.

What is needed in the next section is the singular expansion for $B^\bullet$, but to compute it we first need the singular expansions of u, v, $T^\bullet$ and D (for u and v compare also with [11] and [9]).

Lemma 9.39. *Let $U = U(x,z)$ and $V = V(x,z)$ be the solutions of the system of equations $U = xz(1+V)^2$ and $V = z(1+U)^2$. Let $r(z)$ be explicitly given by*

$$r(z) = \frac{\widetilde{u}_0(z)}{z(1+z(1+\widetilde{u}_0(z))^2)^2}, \tag{9.86}$$

where

$$\widetilde{u}_0(z) = -\frac{1}{3} + \sqrt{\frac{4}{9} + \frac{1}{3z}}.$$

Furthermore, let $\tau(x)$ be the inverse function of $r(z)$ and let $u_0(x) = \widetilde{u}_0(\tau(x))$ which is also the solution of the equation

$$x = \frac{(1+u)(3u-1)^3}{16u}.$$

Then, for x sufficiently close to the positive real axis, the function $U(x,z)$ and $V(x,z)$ have a dominant singularity at $z = \tau(x)$ and have local expansions of the form

$$U(x,z) = u_0(x) + u_1(x)Z + u_2(x)Z^2 + u_3(x)Z^3 + O(Z^4),$$
$$V(x,z) = v_0(x) + v_1(x)Z + v_2(x)Z^2 + v_3(x)Z^3 + O(Z^4),$$

where $Z = \sqrt{1 - z/\tau(x)}$*. The functions* $u_j(x)$ *and* $v_j(x)$ *are also analytic and can be explicitly given in terms of* $u = u_0(x)$*. In particular we have*

$$
\begin{aligned}
u_0(x) &= u & v_0(x) &= \frac{1+u}{3u-1},\\
u_1(x) &= -\sqrt{2u(u+1)} & v_1(x) &= -\frac{2\sqrt{2u(u+1}}{3u-1},\\
u_2(x) &= \frac{(1+u)(7u+1)}{2(1+3u)} & v_2(x) &= \frac{2u(3+5u)}{(3u-1)(1+3u)},\\
u_3(x) &= -\frac{(1+u)(67u^2+50u+11)u}{4(1+3u)^2\sqrt{2u^2+2u}} & v_3(x) &= -\frac{\sqrt{2}u(1+u)(79u^2+42u+7)}{4(1+3u)^2(3u-1)\sqrt{u(1+u)}}.
\end{aligned}
$$

Similarly, for z *sufficiently close to the real axis the functions* $U = U(x,z)$ *and* $V = V(x,z)$ *have a dominant singularity* $x = r(z)$ *and there is also a local expansion of the form*

$$
\begin{aligned}
U(x,z) &= \widetilde{u}_0(z) + \widetilde{u}_1(z)X + \widetilde{u}_2(z)X^2 + \widetilde{u}_3(z)X^3 + O(X^4),\\
V(x,z) &= \widetilde{v}_0(z) + \widetilde{v}_1(z)X + \widetilde{v}_2(z)X^2 + \widetilde{v}_3(z)X^3 + O(X^4),
\end{aligned}
$$

where $X = \sqrt{1 - x/r(z)}$*. The functions* $\widetilde{u}_j(z)$ *and* $\widetilde{v}_j(z)$ *are analytic and can be explicitly given in terms of* $\widetilde{u} = \widetilde{u}_0(z)$*. For example, we have*

$$
\begin{aligned}
\widetilde{u}_0(z) &= \widetilde{u} & \widetilde{v}_0(z) &= \frac{1+\widetilde{u}}{3\widetilde{u}-1}\\
\widetilde{u}_1(z) &= -\frac{2\widetilde{u}\sqrt{1+\widetilde{u}}}{\sqrt{1+3\widetilde{u}}} & \widetilde{v}_1(z) &= -\frac{4\widetilde{u}\sqrt{1+\widetilde{u}}}{(3\widetilde{u}-1)\sqrt{1+3\widetilde{u}}}\\
\widetilde{u}_2(z) &= \frac{2(1+\widetilde{u})\widetilde{u}(2\widetilde{u}+1)}{(1+3\widetilde{u})^2} & \widetilde{v}_2(z) &= \frac{4\widetilde{u}(5\widetilde{u}^2+4\widetilde{u}+1)}{(3\widetilde{u}-1)(1+3\widetilde{u})^2}\\
\widetilde{u}_3(z) &= -\frac{2\widetilde{u}(10\widetilde{u}^3+11\widetilde{u}^2+5\widetilde{u}+1)}{(1+3\widetilde{u})^{7/2}(1+\widetilde{u})^{-1/2}} & \widetilde{v}_3(z) &- \frac{4\widetilde{u}(2\widetilde{u}+1)(11\widetilde{u}^2+5\widetilde{u}+1)}{(3\widetilde{u}-1)(1+3\widetilde{u})^{7/2}(1+\widetilde{u})^{-1/2}}
\end{aligned}
$$

Proof. Since $U(x,z)$ satisfies the functional equation $U = xz(1+z(1+U)^2)^2$, it follows that for any fixed real and positive z the function $x \mapsto U(x,z)$ has a square-root singularity at $r(z)$ that satisfies the equations

$$\Phi(r(z), z, u) = u \quad \text{and} \quad \Phi_u(r(z), z, u) = 0,$$

where $\Phi(x,z,u) = xz(1+z(1+u)^2)^2$. Now a short calculation gives the explicit formula (9.86) for $r(z)$. By continuity we obtain the same kind of representation if z is complex but sufficiently close to the positive real axis.

We proceed in the same way if x is fixed and z is considered as the variable. Then $\tau(x)$, the functional inverse of $r(z)$, is the singularity of the mapping $z \mapsto U(x,z)$. Furthermore, the coefficients $u_1(x)$ etc. can be calculated. The derivations for $V(x,z)$ are completely of the same kind.

Lemma 9.40. *Suppose that x and w are sufficiently close to the positive real axis and that $|w| \leq 1$. Then the dominant singularity $z = \tau(x)$ of $T^\bullet(x, z, w)$ (that is given in Lemma 9.39) does not depend on w. The singular expansion at $\tau(x)$ is*

$$T^\bullet(x, z, w) = T_0(x, w) + T_2(x, w)Z^2 + T_3(x, w)Z^3 + O(Z^4), \qquad (9.87)$$

where $Z = \sqrt{1 - z/\tau(x)}$, and the T_i are given in the appendix.

Proof. Suppose for a moment that all variables x, z, w are non-negative real numbers and let us look at the expression (9.82) for $T^\bullet$. The algebraic functions $U(x, z)$ and $V(x, z)$ are always non-negative and, since the factor w in the denominator cancels with a corresponding factor in the numerator, the only possible sources of singularities are: a) those coming from U and V, or b) the vanishing of $w_2(u, v, w)$ inside the square root.

We can discard source b) as follows. For fixed $u, v > 0$, let $w_2(w) = w_2(u, v, w)$. We can check that

$$w_2(1) = (1 + 2u + u^2 + 2v + v^2 + uv - uv^2)^2,$$
$$w_2'(w) = -2uv\left((6 - w)uv + 1 + 2u + u^2 + 4v + 5v^2 + 2v^3 + 3uv^2 + 2u^2v\right).$$

In particular $w_2(1) > 0$ and $w_2'(w) < 0$ for $w \in [0, 1]$. Thus it follows that $w_2(w) > 0$ in $w \in [0, 1]$. Hence the singularities come from source a) and do not depend on w.

Following Lemma 9.39 (see also [11] and [9]), we have that $z = \tau(x)$ is the radius of convergence of $U(x, z)$, as a function of z. Now by using the expansions of $U(x, z)$ and $V(x, z)$ from Lemma 9.39 we obtain (9.87) and also the explicit representations for T_i.

Finally, by continuity all properties are also valid if x and w are sufficiently close to the real axis, thus completing the proof.

Similarly we get an alternate expansion expanding in the variable x.

Lemma 9.41. *Suppose that z and w are sufficiently close to the positive real axis and that $|w| \leq 1$. Then the dominant singularity $x = r(z)$ of $T^\bullet(x, z, w)$ does not depend on w. The singular expansion at $r(z)$ is*

$$T^\bullet(x, z, w) = \widetilde{T}_0(z, w) + \widetilde{T}_2(z, w)X^2 + \widetilde{T}_3(z, w)X^3 + O(X^4), \qquad (9.88)$$

where $X = \sqrt{1 - x/r(z)}$. Furthermore we have

$$\widetilde{T}_0(z, w) = T_0(r(z), w),$$
$$\widetilde{T}_2(z, w) = T_2(r(z), w)H(r(z), z) - T_{0,x}(r(z), x)r(z),$$
$$\widetilde{T}_3(z, w) = T_3(r(z), w)H(r(z), z)^{3/2},$$

where $H(x, z)$ is a non-zero analytic function with $Z^2 = H(x, z)X^2$.

Proof. We could repeat the proof of Lemma 9.40. However, we present an alternate approach that uses the results of Lemma 9.40 and a kind of singularity transfer.

By applying the Weierstrass preparation theorem it follows that there is a non-zero analytic function with $Z^2 = H(x,z)X^2$. Furthermore, by using the representation $x = r(z)(1-X^2)$ and Taylor expansion we have

$$H(x,z) = H(r(z),z) - H_x(r(z),z)r(z)X^2 + O(X^4),$$
$$T_j(x,w) = T_j(r(z),w) - T_{j,x}(r(z),w)r(z)X^2 + O(X^4).$$

Hence, Lemma (9.40) proves the result. In fact, H can be computed explicitly and is equal to $H(x,\tau(x)) = (1+3u)/2u$.

Lemma 9.42. *Suppose that y and w are sufficiently close to the positive real axis and that $|w| \le 1$. Then the dominant singularity $x = R(y)$ of $D(x,y,w)$ does not depend on w. The singular expansion at $R(y)$ is*

$$D(x,y,w) = D_0(y,w) + D_2(y,w)X^2 + D_3(y,w)X^3 + O(X^4), \qquad (9.89)$$

where $X = \sqrt{1 - x/R(y)}$, and the D_i are given by in the Appendix.

Proof. By (9.48) we already know that $E(x,y) = D(x,y,1)$ has a singular expansion of the form

$$E(x,y) = E_0(y) + E_2(y)X^2 + E_3(y)X^3 + O(X^4). \qquad (9.90)$$

Next we rewrite (9.78) to

$$D(x,y,w) = (1+yw)\exp\left(\frac{xE(x,y)D(x,y,w)}{1+xE(x,y)} + \frac{1}{x^2D(x,y,w)}T^\bullet\left(x, E(x,y), \frac{D(x,y,w)}{E(x,y)}\right)\right) - 1. \qquad (9.91)$$

Now an application of Lemma 9.41 and Theorem 2.31 yields the result. Note that the singularity does not depend on w.

Lemma 9.43. *Suppose that y and w are sufficiently close to the positive real axis and that $|w| \le 1$. Then the dominant singularity $x = R(y)$ of $B^\bullet(x,y,w)$ does not depend on w, and is the same as for $D(x,y,w)$. The singular expansion at $R(y)$ is*

$$B^\bullet(x,y,w) = B_0(y,w) + B_2(y,w)X^2 + B_3(y,w)X^3 + O(X^4), \qquad (9.92)$$

where $X = \sqrt{1 - x/R(y)}$, and the B_i are given in the Appendix.

Proof. We just have to use the representation of $B^\bullet(x,y,w)$ that is given in Lemma 9.38 and the singular expansion of $D(x,y,w)$ from Lemma 9.42.

9.5.3 Degree Distribution for Planar Graphs

We start with the degree distribution in 3-connected graphs, both for edge-rooted and vertex-rooted graphs.

Theorem 9.44. *For every $k \geq 3$ the limiting probability d_k that a vertex of a three-connected planar graph has degree k and the limiting probability e_k that the tail root vertex of an edge-rooted (where the edge is oriented) three-connected planar graph has degree k exist. We have*

$$\sum_{k \geq 3} e_k w^k = \frac{T_3(r,w)}{T_3(r,1)},$$

with $T_3(x,w)$ from above and $r = r(1) = (7\sqrt{7} - 17)/32$ is explicitly given by (9.86). Obviously the e_k are indeed a probability distribution. We have asymptotically, as $k \to \infty$,

$$e_k \sim c \cdot k^{1/2} q^k,$$

where $c \approx 0.9313492$ is a computable constant and $q = 1/(u_0 + 1) = \sqrt{7} - 2$, and where $u_0 = u(r) = (\sqrt{7} - 1)/3$.

Moreover, we have

$$d_k = \alpha \frac{e_k}{k} \sim c\alpha \cdot k^{-1/2} q^k,$$

where

$$\alpha = \frac{(3u_0 - 1)(3u_0 + 1)(u_0 + 1)}{u_0} = \frac{\sqrt{7} + 7}{2}$$

is the asymptotic value of the expected average degree in 3-connected planar graphs.

We remark that the degree distribution in 3-connected planar maps counted according to the number of edges was obtained in [8]. The asymptotic estimates have the same shape as our d_k, but the corresponding value of q is equal to $1/2$.

Proof. First the proof uses the singular expansion (9.88). The representation

$$\sum_{k \geq 3} e_k w^k = T_3(r,w)/T_3(r,1)$$

follows in completely the same way as the proof of Theorem 9.28, using now Lemma 9.40, with the difference that now the dominant term is the coefficient of Z^3.

In order to characterise the dominant singularity of $T_3(1,w)$ and to determine the singular behaviour we observe that the explicit representation of T_3 contains in the denominator a (dominating) singular term of the form

$(-w+U+1)^{3/2}$. Hence, it follows that u_0+1 is the dominant singularity and we also get the proposed asymptotic relation for e_k.

For the proof of the second part of the statement, let $t_{n,k}$ be the number of vertex-rooted graphs with n vertices and with degree of the root equal to k, and let $s_{n,k}$ be the analogous quantity for edge-rooted graphs. Let also $t_n = \sum_k t_{n,k}$ and $s_n = \sum_k s_{n,k}$. Since a vertex-rooted graph with a root of degree k is counted k times as an oriented-edge-rooted graph, we have $s_{n,k} = kt_{n,k}$ (a similar argument is used in [137]). Notice that $e_k = \lim s_{n,k}/s_n$ and $d_k = \lim t_{n,k}/t_n$.

Using Theorem 2.22 it follows that the expected number of edges μ_n in 3-connected planar graphs is asymptotically $\mu_n \sim \kappa n$, where $\kappa = -\tau'(1)/\tau(1)$, and $\tau(x)$ is as in Lemma 9.39. Clearly $s_n = 2\mu_n t_n/n$.

Finally $2\mu_n/n$ is asymptotic to the expected average degree $\alpha = 2\kappa$. Summing up, we obtain

$$kd_k = \alpha\, e_k.$$

A simple calculation gives the value of α as claimed.

Theorem 9.45. *For every $k \ge 2$ the limiting probability d_k that a vertex of a two-connected planar graph has degree k exists and we have*

$$p(w) = \sum_{k\ge 2} d_k w^k = \frac{B_3(1,w)}{B_3(1,1)},$$

with $B_3(y,w)$ given in Lemma 9.43.

Obviously, $p(1) = 1$, so that the d_k are indeed a probability distribution and we have asymptotically, as $k \to \infty$,

$$d_k \sim ck^{-1/2} q^k,$$

where $c \approx 3.0826285$ is a computable constant and

$$q = \left(\frac{1}{1-t_0} \exp\left(\frac{(t_0-1)(t_0+6)}{6t_0^2+20t_0+6} \right) - 1 \right)^{-1} \approx 0.6734506,$$

and $t_0 = t(1) \approx 0.6263717$ is a computable constant given by the equation (9.99).

Proof. The representation of $p(w)$ follows from (9.92).

Now, in order to characterise the dominant singularity of $B_3(1,w)$ and to determine the singular behaviour we first observe that the right hand side of the equation for D_0 contains a singular term of the form $(D_0(t-1)+t)^{3/2}$ that dominates the right hand side. Hence, by applying Theorem 2.31 it follows that $D_0(1,w)$ has the dominant singularity w_3, where $D_0(1,w_3) = t/(1-t)$ and we have a local singular representation of the form

$$D_0(1,w) = \widetilde{D}_{00} + \widetilde{D}_{02}\widetilde{W}^2 + \widetilde{D}_{03}\widetilde{W}^3 + \cdots,$$

where $\widetilde{W} = \sqrt{1 - w/w_3}$ and $\widetilde{D}_{00} = t/(1-t)$. The fact that the coefficient of $\widetilde{W}$ vanishes is due to the shape of the equation (9.98) satisfied by $D_0(1,w)$.

We now insert this expansion into the representation for D_2. Observe that S has an expansion of the form

$$S = S_2\widetilde{W}^2 + S_3\widetilde{W}^3 + \cdots$$

with $S_2 \neq 0$. Thus we have $\sqrt{S} = \sqrt{S_2}\widetilde{W} + O(\widetilde{W}^2)$. Furthermore, we get expansions for $S_{2,1}$, $S_{2,2}$, $S_{2,3}$, and $S_{2,4}$. However, we observe that $S_{2,3}(1,w_3) = 0$, whereas $S_{2,1}(1,w_3) \neq 0$, $S_{2,2}(1,w_3) \neq 0$, and $S_{2,4}(1,w_3) \neq 0$. Consequently we can represent $D_2(1,w)$ as

$$D_2(1,w) = \widetilde{D}_{2,-1}\frac{1}{\widetilde{W}} + \widetilde{D}_{2,0} + \widetilde{D}_{2,1}\widetilde{W} + +\widetilde{D}_{22}\widetilde{W}^2 + \cdots,$$

where $\widetilde{D}_{1,-1} \neq 0$.

In completely the same way it follows that $D_3(1,w)$ has a local expansion of the form

$$D_3(1,w) = \widetilde{D}_{3,-3}\frac{1}{\widetilde{W}^3} + \tilde{+}\widetilde{D}_{3,-1}\frac{1}{\widetilde{W}} + \widetilde{D}_{3,0} + \widetilde{D}_{3,1}\widetilde{W} + \cdots,$$

where $\widetilde{D}_{3,-3} \neq 0$ and the coefficient of $\widetilde{W}^{-2}$ vanish identically.

These types of singular expansions lead to $B_3(1,w)$, and we get

$$B_3(1,w) = \widetilde{B}_{3,-1}\frac{1}{\widetilde{W}} + \widetilde{B}_{3,0} + \widetilde{B}_{3,1}\widetilde{W} + \cdots, \tag{9.93}$$

where $\widetilde{B}_{3,-1} \neq 0$. The fact that the coefficients of $\widetilde{W}^{-3}$ and $\widetilde{W}^{-2}$ vanish is a consequence of non trivial cancellations. Hence, we obtain the proposed asymptotic relation for d_k.

Theorem 9.46. *For every $k \geq 1$ the limiting probability d_k that a vertex of a connected planar graph has degree k exists and we have*

$$p(w) = \sum_{k\geq 1} d_k w^k = -e^{B_0(1,w)-B_0(1,1)}B_2(1,w)$$
$$+ e^{B_0(1,w)-B_0(1,1)}\frac{1+B_2(1,1)}{B_3(1,1)}B_3(1,w),$$

where $B_j(y,w)$, $j = 0,2,3$ are given above.

Moreover, $p(1) = 1$, so that the d_k is indeed a probability distribution and we have asymptotically, as $k \to \infty$,

$$d_k \sim ck^{-1/2}q^k,$$

where $c \approx 3.0175067$ is a computable constant and q is as in Theorem 9.45.

Proof. The degree distribution is encoded in the function

$$C^\bullet(x,w) = \sum_{k\ge 1} C_k(x,1)w^k = e^{B^\bullet(xC'(x),1,w)},$$

where the generating function $xC'(x)$ of connected rooted planar graphs satisfies the equation

$$xC'(x) = xe^{B^\bullet(xC'(x),1,1)}.$$

From Lemma 9.43 we get the local expansions

$$e^{B^\bullet(x,1,w)} = e^{B_0(1,w)}\left(1 + B_2(1,w)X^2 + B_3(1,w)X^3 + O(X^4)\right),$$

where $X = \sqrt{1-x/R}$. Thus, we first get an expansion for $xC'(x)$

$$xC'(x) = R - \frac{R}{1+B_2(1,1)}\widetilde{X}^2 + \frac{RB_3(1,1)}{(1+B_2(1,1))^{5/2}}\widetilde{X}^3 + O(\widetilde{X}^4),$$

where $\widetilde{X} = \sqrt{1-x/\rho}$ and ρ is the radius of convergence of $C'(x)$ (compare with [91]). Note also that $R = \rho e^{B_0(\rho,1)}$. Thus, we can apply Lemma 2.32 with $H(x,z,w) = xe^{B^\bullet(z,1,w)}$ and $f(x) = xC'(x)$. We have

$$f_0 = R,\ f_2 = -\frac{R}{1+B_2(1,1)},\ f_3 = \frac{RB_3(1,1)}{(1+B_2(1,1))^{5/2}}$$

and

$$h_0(\rho,w) = \rho e^{B_0(1,w)},$$
$$h_2(\rho,w) = \rho e^{B_0(1,w)}B_2(1,w),$$
$$h_3(\rho,w) = \rho e^{B_0(1,w)}B_3(1,w).$$

We can express the probability generating function $p(w)$ as

$$\lim_{n\to\infty} \frac{[x^n]xC^\bullet(x,w)}{[x^n]xC'(x)}.$$

Consequently, we have

$$\begin{aligned} p(w) &= -\frac{h_2(\rho,w)}{f_0} + \frac{h_3(\rho,w)}{f_3}\left(-\frac{f_2}{f_0}\right)^{3/2} \\ &= -e^{B_0(1,w)-B_0(1,1)}B_2(1,w) \\ &\quad + e^{B_0(1,w)-B_0(1,1)}\frac{1+B_2(1,1)}{B_3(1,1)}B_3(1,w). \end{aligned}$$

The singular expansion for $B_2(1,w)$ turns out to be of the form

$$B_2(1,w) = \widetilde{B}_{2,0} + \widetilde{B}_{2,1}\widetilde{W} + \cdots.$$

Hence the expansion (9.93) for $B_3(1,w)$ gives the leading part in the asymptotic expansion for $p(w)$. It follows that $p(w)$ has the same dominant singularity as for 2-connected graphs and we obtain the asymptotic estimate for the d_k as claimed, with a different multiplicative constant. This concludes the proof of the main result.

9.5.4 Vertices of Degree 1 or 2 in Planar Graphs

In this final section we discuss vertices of degree ≤ 2 in planar graphs.

Theorem 9.47. *For $k \in \{1,2\}$, let $X_n^{(k)}$ denote the number of vertices of degree k in random planar graphs with n vertices. Then $X_n^{(k)}$ satisfies a central limit theorem with linear expected value and variance.*

There are no vertices of degree 1 in 2-connected graphs. Hence, we can apply the general relation between connected and 2-connected graphs and obtain

$$\frac{\partial C_{d=1}(x,u)}{\partial x} = \exp\left(B'\left(x\frac{\partial C_{d=1}(x,u)}{\partial x} + x(u-1)\right)\right), \tag{9.94}$$

where $B(x) = B(x,1)$ is the generating function of 2-connected planar graphs and $C_{d=1}(x,u)$ the counting function of vertices of degree 1 in connected planar graphs. Hence, by an application of Theorem 2.31 it follows that we have a local representation of the form

$$x\frac{\partial C_{d=1}(x,u)}{\partial x} = g(x,u) + h(x,u)\left(1 - \frac{x}{\rho(u)}\right)^{\frac{3}{2}},$$

which implies

$$C_{d=1}(x,u) = g_2(x,u) + h_2(x,u)\left(1 - \frac{x}{\rho(u)}\right)^{\frac{5}{2}}$$

and a corresponding singular representation for the generating function of all graphs:

$$G_{d=1}(x,u) = e^{C_{d=1}(x,u)} = g_3(x,u) + h_3(x,u)\left(1 - \frac{x}{\rho(u)}\right)^{\frac{5}{2}}.$$

Of course, by Theorem 2.25 a central limit theorem follows.

Remark 9.48 *Due to the simplicity of the equation (9.94) we also find $\mu = \sigma^2 = \rho$ and, thus,*

$$\mathbb{E}\, X_n^{(1)} = \rho n + O(1) \quad \textit{and} \quad \mathbb{V}\, X_n^{(1)} = \rho n + O(1).$$

Vertices of degree $k = 2$ are more difficult to handle. Our first goal is to characterise the generating function $B(x, y, z_1, z_2, z_\infty)$ of 2-connected planar graphs where x marks vertices, y edges and z_j vertices of degree j, $j \in \{1, 2, \infty\}$.

Our method is based on a result of Walsh [207] on 2-connected graphs without vertices of degree two. The original result is stated for arbitrary labelled graphs, but it applies also to planar graphs; the reason is that a graph is planar, if and only if it remains planar after removing the vertices of degree 2. Note that series-parallel networks and graphs appear here. In order to avoid confusion, we write, for example, $B_{\mathrm{SP}}(x, y)$ to denote the generating function of 2-connected series-parallel graphs (described in Theorem 9.4).

Lemma 9.49. *Let $B(x, y)$ be the generating function for 2-connected planar graphs, and let $H(x, y)$ be the generating function for 2-connected planar graphs without vertices of degree 2, where x marks vertices and y marks edges. Also, let $D_{\mathrm{SP}}(x, y)$ be the generating function of series-parallel networks, and $B_{\mathrm{SP}}(x, y)$ the generating function of 2-connected series-parallel graphs. Then*

$$H(x, D_{\mathrm{SP}}(x, y)) = B(x, y) - B_{\mathrm{SP}}(x, y). \tag{9.95}$$

Proof. Given a 2-connected planar graph, we perform the following operation repeatedly: we remove a vertex of degree two, if there is any, and we remove parallel edges created, if any. In this way we get either a graph with minimum degree three, or a single edge in case the graph was series-parallel. This gives

$$B(x, y) = H(x, D_{\mathrm{SP}}(x, y)) + B_{\mathrm{SP}}(x, y).$$

Corollary 9.50 *The generating function $H(x, y)$ is given by*

$$H(x, y) = B(x, \phi(x, y)) - B_{\mathrm{SP}}(x, \phi(x, y)),$$

where

$$\phi(x, y) = (1 + y) \exp\left(\frac{-xy^2}{1 + xy}\right) - 1.$$

Proof. Since $D_{\mathrm{SP}}(x, y)$ satisfies the equation (9.17), we can express y by

$$y = \phi(x, D_{\mathrm{SP}}) = (1 + D_{\mathrm{SP}}) \exp\left(\frac{-xD_{\mathrm{SP}}^2}{1 + xD_{\mathrm{SP}}}\right) - 1$$

and obtain

$$H(x, D_{\mathrm{SP}}) = B(x, \phi(x, D_{\mathrm{SP}})) - B_{\mathrm{SP}}(x, \phi(x, D_{\mathrm{SP}}))$$

where we can now interpret D_{SP} as an independent variable.

With help of theses preliminaries we can obtain an explicit expression for $B(x, y, z_1, z_2, z_\infty)$.

Lemma 9.51. *Let $k = 2$ and let $B_{\mathrm{SP}}(x, y, z_1, z_2, z_\infty)$ be the generating function for 2-connected series parallel graphs and*

$$D_{\mathrm{SP}}(x, y, z_1, z_2, z_\infty) = \sum_{i,j \in \{1,2,\infty\}} D_{SP;ij}(x, y, z_1, z_2, z_\infty)$$

the corresponding generating function of series-parallel networks (compare with Lemma 9.30 and 9.33). Then we have

$$B(x, y, z_1, z_2, z_\infty) = B_{\mathrm{SP}}(x, y, z_1, z_2, z_\infty) + H(x, D_{\mathrm{SP}}(x, y, z_1, z_2, z_\infty)). \quad (9.96)$$

Proof. We just add the counting of vertices of degrees one and two that come from the series-parallel networks and from series-parallel graphs.

Set

$$B_j^\bullet(x, y, z_1, z_2, z_\infty) = \frac{1}{x} \frac{\partial B(x, y, z_1, z_2, z_\infty)}{\partial z_j} \qquad (j \in \{1, 2, \infty\}).$$

The division by x is necessary because in the definition of $B^\bullet$ the root bears no label.

Further, let $C_j^\bullet(x, y, z_1, z_2, z_\infty)$, $j \in \{1, 2, \infty\}$, denote the corresponding generating functions for connected planar graphs. Then we have (as in Lemma 9.23) the system of equations

$$\begin{aligned}
C_1^\bullet &= B_1^\bullet(x, y, W_1, W_2, W_\infty), \\
C_2^\bullet &= \frac{1}{2!}(B_1^\bullet(x, y, W_1, W_2, W_\infty))^2 + B_2^\bullet(x, y, W_1, W_2, W_\infty), \\
C_\infty^\bullet &= e^{B_1^\bullet(x,y,W_1,W_2,W_\infty) + B_2^\bullet(x,y,W_1,W_2,W_\infty) + B_\infty^\bullet(x,y,W_1,W_2,W_\infty)} \\
&\quad - 1 - B_1^\bullet(x, y, W_1, W_2, W_\infty) - B_2^\bullet(x, y, W_1, W_2, W_\infty) \\
&\quad - \frac{1}{2!}(B_1^\bullet(x, y, W_1, W_2, W_\infty))^2,
\end{aligned}$$

where the W_j, $j \in \{1, 2, \infty\}$ are

$$\begin{aligned}
W_1 &= z_1 + z_2 C_1^\bullet + z_\infty(C_2^\bullet + C_\infty^\bullet), \\
W_2 &= z_2 + z_\infty(C_1^\bullet + C_2^\bullet + C_\infty^\bullet), \\
W_\infty &= z_\infty(1 + C_1^\bullet + C_2^\bullet + C_\infty^\bullet).
\end{aligned}$$

From this system of equations we get a single equation for

$$C^\bullet = 1 + C_1^\bullet + C_2^\bullet + C_\infty^\bullet$$

if we set $z_1 = z_\infty = 1$, and $y = 1$.

Lemma 9.52. *The function $C^\bullet(x, y, 1, z_2, 1)$ satisfies a functional equation of the form*

$$C^\bullet = F(x, y, z_2, C^\bullet).$$

Proof. Since $B_1^\bullet = xz_1$, the equation for $C_1^\bullet$ is

$$C_1^\bullet = xW_1 = x\left(1 + z_2C_1^\bullet + C_2^\bullet + C_\infty^\bullet\right) = xC^\bullet + x(z_2-1)C_1^\bullet,$$

which gives

$$C_1^\bullet = \frac{xC^\bullet}{1-x(z_2-1)}.$$

Consequently, we have

$$W_1 = C^\bullet + (z_2-1)C_1^\bullet = \frac{C^\bullet}{1-x(z_2-1)}.$$

Since $W_2 = z_2 - 1 + C^\bullet$ and $W_\infty = C^\bullet$, we can sum the three equations for $C_1^\bullet, C_2^\bullet, C_\infty^\bullet$ and obtain

$$C^\bullet = e^{B_1^\bullet(x,y,W_1,W_2,W_\infty)+B_2^\bullet(x,y,W_1,W_2,W_\infty)+B_\infty^\bullet(x,y,W_1,W_2,W_\infty)}$$

which is now a single equation for $C^\bullet$.

Finally the generating function $C_{d=2}(x,u)$ for connected planar graphs is determined by

$$\frac{\partial C_{d=2}(x,u)}{\partial x} = C_1^\bullet(x,1,1,u,1) + uC_2^\bullet(x,1,1,u,1) + C_\infty^\bullet(x,1,1,u,1).$$

We recall that

$$\begin{aligned} B(x,y,z_1,z_2,z_\infty) &= B_{\mathrm{SP}}(x,y,z_1,z_2,z_\infty) + H(x, D_{\mathrm{SP}}(x,y,z_1,z_2,z_\infty)) \\ &= B(x,\phi(x,D_{\mathrm{SP}}(x,y,z_1,z_2,z_\infty))) \\ &\quad + B_{\mathrm{SP}}(x,y,z_1,z_2,z_\infty) - B_{\mathrm{SP}}(x,\phi(x,D_{\mathrm{SP}}(x,y,z_1,z_2,z_\infty))). \end{aligned}$$

We already know that $B(x,y)$ has a singular expansion of the form

$$B(x,y) = B_0(y) + B_2(y)X^2 + B_4(y)X^4 + B_5(y)X^5 + B_6(y)X^6 + \cdots,$$

where $X = \sqrt{1-x/R(y)}$ and $R(1) = 0.03819\ldots$ Alternatively this can be rewritten as

$$B(x,y) = g(x,y) + h(x,y)\left(1 - \frac{x}{R(y)}\right)^{\frac{5}{2}}.$$

Further note that $B_{\mathrm{SP}}(x,y,z_1,z_2,z_\infty)$, $B_{\mathrm{SP}}(x,y) = B_{\mathrm{SP}}(x,y,1,\ldots,1)$, and $\phi(x, D_{\mathrm{SP}}(x,y,z_1,z_2,z_\infty))$ are analytic around $x = R(1)$, $y = 1$ and $z_j = 1$, $j \in \{1,\ldots,k,\infty\}$. Hence, $B(x,y,z_1,z_2,z_\infty)$ can be represented as

$$B(x,y,z_1,z_2,z_\infty) = g_2(x,y,z_1,z_2,z_\infty) + h_2(x,y,z_1,z_2,z_\infty)\left(1 - \frac{x}{\overline{R}(y,z_1,z_2,z_\infty)}\right)^{5/2}, \tag{9.97}$$

where g_2, h_2, and $\overline{R}$ are analytic functions with

$$
\begin{aligned}
g_2(x, y, 1, 1, 1) &= g(x, y), \\
h_2(x, y, 1, 1, 1) &= h(x, y), \\
\overline{R}(y, 1, 1, 1) &= R(y).
\end{aligned}
$$

We can apply Theorem 2.25 with $\alpha = \frac{5}{2}$ and obtain the central limit theorem in the 2-connected case.

From (9.97) it follows that $B_j^\bullet(x, y, z_1, z_2, z_\infty)$ has a singularity of the kind (9.74). Hence, the function $F(x, y, u)$ from Lemma 9.52 has also a singularity of that kind. Note that $F(x, y, 1) = e^{B^\bullet(x, xy)}$. Furthermore, for $u = 1$ it is known that the singularity of $C^\bullet$ interferes with the singularity of $B^\bullet$.

Hence all assumptions of Proposition 2.31 are satisfied and it follows that $C^\bullet(x, 1, 1, u, 1)$ has a singular expansion of the form (9.74). Of course, with help of Theorem 2.25 we get a central limit theorem in the connected case. Finally, since $G_{d=2}(x, u) = e^{C_{d=2}(x, u)}$ the result follows for general planar graphs, too.

Appendix

We collect explicit expressions of coefficients that appear in singular expansion of generating functions in the formulations of Lemmas 9.40, 9.42 and 9.43.

We start with $T_j(x,w)$ from Lemma 9.40. Here the abbreviation $u = u(x)$ stands for $u(x,\tau(x))$, which is the solution of $x = (1+u)(3u-1)^3/(16u)$ (as in Lemma 9.39).

$$\begin{aligned} T_0(x,w) \ &= -\frac{(w-u-1)w(3u-1)^6\sqrt{P}}{27648(2u-1+3u^2+w)(1+u)u^4} \\ &\quad -\frac{(3u-1)^6 w P_0}{27648(9u+1)(2u-1+3u^2+w)(1+u)u^4}, \end{aligned}$$

$$\begin{aligned} T_2(x,w) \ &= \frac{(3u-1)^6 P_{2,0} w}{82944u^5(2u-1+3u^2+w)^2(u+1)^2(9u+1)^2} \\ &\quad -\frac{P_{2,1} w(3u-1)^6}{82944u^5(2u-1+3u^2+w)^2(u+1)^2\sqrt{P}}, \end{aligned}$$

$$\begin{aligned} T_3(x,w) \ &= -\frac{\sqrt{2u(1+u)}w(3u-1)^6((9u^2-6u+1)w-9u^2-10u-1)}{373248u^6(u+1)^3} \\ &\quad \times(3u+1) + \frac{\sqrt{2u(1+u)}w(3u-1)^6 P_3(3u+1)}{373248u^6(u+1)^3\sqrt{P}(-w+u+1)} \end{aligned}$$

and

$$\begin{aligned} P &= (-w+u+1)((-9u^2+6u-1)w+81u^3+99u^2+19u+1), \\ P_0 &= (27u^2+6u+1)w^2+(-126u^3-150u^2-26u-2)w \\ &\quad +81u^4+180u^3+118u^2+20u+1 \\ P_{2,0} &= (1458u^5+3807u^4+900u^3+114u^2-6u-1)w^3 \\ &\quad +(6561u^7+20898u^6+8532u^5-7281u^4-1635u^3-132u^2+30u+3)w^2 \end{aligned}$$

$$+ (-3645u^8 - 30942u^7 - 46494u^6 - 13230u^5 + 7536u^4 + 1590u^3 - 18u^2$$
$$- 42u - 3)w + 13122u^9 + 47385u^8 + 61560u^7 + 30708u^6 - 228u^5$$
$$- 4530u^4 - 872u^3 + 36u^2 + 18u + 1$$

$$P_{2,1} = (-54u^4 - 45u^3 + 57u^2 - 15u + 1)w^4 +$$
$$(-243u^6 + 27u^5 + 1278u^4 + 858u^3 - 111u^2 + 35u - 4)w^3 +$$
$$(1944u^7 + 6507u^6 + 5553u^5 - 576u^4 - 1530u^3 + 15u^2 - 15u + 6)w^2$$
$$+ (-1215u^8 - 6561u^7 - 11439u^6 - 7005u^5 + 231u^4 + 1229u^3$$
$$+ 75u^2 - 15u - 4)w + 1458u^9 + 6561u^8 + 11376u^7 + 8988u^6 + 2388u^5$$
$$- 794u^4 - 512u^3 - 36u^2 + 10u + 1$$

$$P_3 = (-27u^3 + 27u^2 - 9u + 1)w^3 + (162u^4 + 135u^3 - 27u^2 - 3u - 3)w^2$$
$$+ (81u^5 + 243u^4 + 270u^3 + 138u^2 + 33u + 3)w$$
$$- 81u^5 - 261u^4 - 298u^3 - 138u^2 - 21u - 1.$$

Next we consider $D_j(y, w)$ from Lemma 9.42:

$$1 + D_0 = (1 + yw) \exp\left(\frac{\sqrt{S}(D_0(t-1) + t)}{4(3t+1)(D_0+1)} - \frac{D_0^2(t^4 - 12t^2 + 20t - 9)}{4(t+3)(D_0+1)(3t+1)} \right.$$
$$\left. - \frac{D_0(2t^4 + 6t^3 - 6t^2 + 10t - 12) + t^4 + 6t^3 + 9t^2}{4(t+3)(D_0+1)(3t+1)} \right), \qquad (9.98)$$

where S abbreviates

$$S = (D_0 t - D_0 + t)(D_0(t-1)^3 + t(t+3)^2),$$

and $t = t(y)$ stands for the unique solution in $(0, 1)$ of

$$y = \frac{(1-2t)}{(1+3t)(1-t)} \exp\left(-\frac{t^2(1-t)(18 + 36t + 5t^2)}{2(3+t)(1+2t)(1+3t)^2} \right) - 1, \qquad (9.99)$$

by

$$D_2 = \frac{4(D_0+1)^2(t-1)(S_{2,1} + S_{2,2}\sqrt{S})}{(17t^5 + 237t^4 + 1155t^3 + 2527t^2 + 1808t + 400)(S_{2,3} + S_{2,4}\sqrt{S})},$$

where

$$S_{2,1} = -D_0^2(t-1)^4(t+3)(11t^5 + 102t^4 + 411t^3 + 588t^2 + 352t + 72)$$
$$- D_0 t(t-1)(t+3)$$
$$\times (22t^7 + 231t^6 + 1059t^5 + 2277t^4 + 2995t^3 + 3272t^2 + 2000t + 432)$$
$$- t^2(t+3)^3(11t^5 + 85t^4 + 252t^3 + 108t^2 - 48t - 24)$$

$$S_{2,2} = D_0(t-1)(11t^7 + 124t^6 + 582t^5 + 968t^4 - 977t^3 - 4828t^2 - 4112t$$
$$- 984) + t(t+3)^2(11t^5 + 85t^4 + 252t^3 + 108t^2 - 48t - 24)$$

$$S_{2,3} = -(t+3)(D_0 t - D_0 + t)(D_0^2(t-1)^4 + 2D_0(t-1)(t^3 - t^2 + 5t - 1) \\ + t(t^3 - 3t - 14)$$

$$S_{2,4} = D_0^2(t^2 + 2t - 9)(t-1)^2 + D_0(2t^4 - 12t^2 + 80t - 6) + t(t^3 - 3t + 50),$$

and by

$$D_3 = \frac{24(t+3)(D_0+1)^2(t-1)t^2(t+1)^2 S_{3,1}^{3/2}\big(S_{3,2} - S_{3,3}(D_0 t - D_0 + t)\sqrt{S}\big)}{(\beta)^{5/2}(D_0 t - D_0 + t)\big(S_{3,4}\sqrt{S} - (t+3)(D_0 t - D_0 + t)S_{3,5}\big)},$$

where

$$\beta = 3t(1+t)(17t^5 + 237t^4 + 1155t^3 + 2527t^2 + 1808t + 400),$$

$$S_{3,1} = -5t^5 + 6t^4 + 135t^3 + 664t^2 + 592t + 144,$$

$$\begin{aligned} S_{3,2} = {} & D_0^3(81t^{11} + 135t^{10} - 828t^9 - 180t^8 + 1982t^7 + 1090t^6 - 5196t^5 \\ & \quad + 2108t^4 + 2425t^3 - 1617t^2 - 256t + 256), \\ & + D_0^2(243t^{11} + 1313t^{10} + 1681t^9 - 51t^8 - 5269t^7 - 7325t^6 + 2571t^5 \\ & \quad + 10271t^4 + 1846t^3 - 3888t^2 - 1392t) \\ & + D_0(243t^{11} + 2221t^{10} + 8135t^9 + 15609t^8 + 12953t^7 - 3929t^6 - 12627t^5 \\ & \quad - 13293t^4 - 7680t^3 - 1632t^2) \\ & + 81t^{11} + 1043t^{10} + 5626t^9 + 16806t^8 + 30165t^7 + 30663t^6 + 13344t^5 \\ & \quad + 1008t^4 - 432t^3, \end{aligned}$$

$$\begin{aligned} S_{3,3} = {} & D_0(81t^8 + 378t^7 + 63t^6 - 1044t^5 + 1087t^4 - 646t^3 - 687t^2 + 512t + 256) \\ & + 81t^8 + 800t^7 + 3226t^6 + 7128t^5 + 8781t^4 + 4320t^3 + 384t^2 - 144t, \end{aligned}$$

$$S_{3,4} = D_0^2(t^4 - 12t^2 + 20t - 9) + D_0(2t^4 - 12t^2 + 80t - 6) + t^4 - 3t^2 + 50t$$

$$\begin{aligned} S_{3,5} = {} & D_0^2(t^4 - 4t^3 + 6t^2 - 4t + 1) + D_0(2t^4 - 4t^3 + 12t^2 - 12t + 2) \\ & + t^4 - 3t^2 - 14t. \end{aligned}$$

Finally the coefficients $B_j(y,w)$ from Lemma 9.43 are given by the following formulas:

$$\begin{aligned} B_0 = \frac{1}{128t^3}\Big(& -8\log(2)(3t^4 + 6t^2 - 1) - 8\log(t+1)(3t-1)(t+1)^3 \\ & - 4\log(t)(3t-1)(t+1)^3 + 2\log(A)(t-1)^3(3t+1) \\ & + 2\log(B)(3t^4 + 24t^3 + 6t^2 - 1) \\ & + \sqrt{S}(t-1)(D_0(t^3 - 3t^2 + 3t - 1) + t^3 - 8t^2 + t - 2) \\ & - \frac{D_0}{t+3}\big(D_0(t-1)^5(t^2 + 2t - 9) + 2(t-1)^3\big)(t^4 + 60t + 3) \\ & + (t+3)^2(t-1)(t^3 - 8t^2 + t - 2)t\big)\Big), \end{aligned}$$

$$B_2 = \frac{RD_0(D_0(R^2E_0^2 + RE_2) - 2(1 + RE_0))}{2(1 + RE_0)^2} + \frac{1}{R}(I_{0,0} + I_{0,2} + I_{2,2})$$
$$+ \left(\log(1 + D_0) - \log(1 + yw) - \frac{R^2E_0D_0}{1 + RE_0}\right)(1 + D_0 - D2)R,$$
$$B_3 = \frac{R^2D_0\left(2D_3E_0^2R + 2D_3E_0 + E_3D_0\right)}{2(E_0R + 1)^2} + RD_3(\log(1 + yw) - \log(D_0 + 1))$$
$$+ \frac{1}{R}(I_{0,3} + I_{2,3} + I_{3,3}),$$

with

$$I_{0,0} = \frac{(3t + 1)(t - 1)^3}{2048t^6}\Big(4(3t - 1)(t + 1)^3\log(t) \;+ 8(3t - 1)(t + 1)^3\log(t + 1)$$
$$+ 8(3t^4 + 6t^2 - 1)\log(2) \;- 2(t - 1)^3(3t + 1)\log(A)$$
$$- 2(3t^4 + 6t^2 - 1)\log(B)$$
$$+ (t - 1)(D_0(t^3 - 3t^2 + 3t - 1) + t^3 + 4t^2 + t + 2)\sqrt{S}$$
$$- \frac{t - 1}{t + 3}\big(D_0^2(t^6 - 2t^5 + t^4 - 4t^3 + 11t^2 - 10t + 3)$$
$$+ D_0(2t^6 + 8t^5 - 10t^4 - 32t^3 + 46t^2 - 8t - 6)$$
$$+ t^6 + 10t^5 + 34t^4 + 44t^3 + 21t^2 + 18t)\Big),$$

$$I_{0,2} = \frac{-(3t + 1)(t - 1)^3}{512t^6}\Bigg(4(3t^4 - 4t^3 + 6t^2 - 1)\log(2) + 2(3t^4 + 6t^2 - 1)\log(t)$$
$$+ 4(3t^4 + 6t^2 - 1)\log(t + 1) + (-3t^4 + 8t^3 - 6t^2 + 1)\log(3A)$$
$$+ (-3t^4 - 8t^3 - 6t^2 + 1)\log(B)$$
$$+ \frac{D_2(t - 1)^5(R_{0,0}\sqrt{S} + R_{0,1}) - \frac{(t-1)^2}{\beta(t+3)}(R_{0,4}\sqrt{S} + R_{0,5})}{(t + 3)\left(R_{0,6}\sqrt{S} + R_{0,7}\right)}\Bigg),$$

$$I_{0,3} = -\frac{(3t + 1)t^2(t - 1)^8(R_{0,0}\sqrt{S} + R_{0,1})(D_3\beta^{5/2} + D_0\alpha^{3/2}R_{0,2})}{512(t + 3)t^8\beta^{5/2}\left((D_0(t^2 - 2t + 1) + t^2 + t + 2)\sqrt{S} + R_{0,3}\right)},$$

$$I_{2,2} = \frac{(t - 1)^6R_{2,0}\left(R_{2,1}\sqrt{S} + R_{2,2}\right)}{3072\beta t^6(t + 1)(D_0 + 1)},$$

$$I_{2,3} = \frac{(t + 1)^2(1 + 2t)^2\alpha^{3/2}(3t + 1)(t - 1)^7\left((t + 3)^2R_{2,3}\sqrt{S} - R_{2,4}\right)}{16t^5\beta^{5/2}(1 + D_0)},$$

$$I_{3,3} = \frac{\sqrt{3}(t - 1)^6R_{2,0}^{3/2}\left(R_{3,0}\sqrt{S} - R_{3,1}(D_0t - D_0 + t)\right)}{2304\sqrt{3t + 1}\,t^5\beta^{3/2}(t + 1)^{3/2}(D_0t - D_0 + t)},$$

where the expressions A, B, and polynomials $R_{i,j}$ are given by

$$
\begin{aligned}
A &= D_0(5t^3 - 3t^2 - t - 1) + 5t^3 + 6t^2 + 5t + (3t+1)\sqrt{S},\\
B &= D_0(t^3 - 3t^2 + 3t - 1) + t^3 + 2t^2 + 5t + (t-1)\sqrt{S},\\
R_{0,0} &= 3D_0^2(t-1)^2 - D_0(7t-3) - t(t+3),\\
R_{0,1} &= 3D_0^3(t-1)^4 - D_0^2(t-1)(3t^3 - t^2 + 25t - 3)\\
&\quad + D_0 t(t^3 + 8t^2 + 21t - 14) + (t+3)^2 t^2,\\
R_{0,2} &= 128t(3t+1)(t-1)(1+2t)^2(t+3)^2(t+1)^2,\\
R_{0,3} &= D_0^2(t^4 - 4t^3 + 6t^2 - 4t + 1) + D_0(2t^4 + 4t^2 - 8t + 2)\\
&\quad + t^4 + 4t^3 + 7t^2 + 2t + 2,\\
R_{0,4} &= 3D_0^3(t-1)^5(51t^8 + 1081t^7 + 8422t^6 + 31914t^5 + 59639t^4 + 42461t^3\\
&\qquad\qquad + 7584t^2 - 2832t - 864)\\
&\quad - D_0^2(t-1)^3(153t^{10} + 3204t^9 + 29055t^8 + 146710t^7 + 432951t^6\\
&\qquad\qquad + 717528t^5 + 561457t^4 + 208750t^3 + 47040t^2 + 13248t + 2592)\\
&\quad - D_0(t+3)^2 t(408t^{10} + 6177t^9 + 34003t^8 + 92097t^7 + 122523t^6\\
&\qquad\qquad + 126075t^5 + 145777t^4 + 82707t^3 - 1543t^2 - 15088t - 3312)\\
&\quad + 3t(t-1)(t+3)^2(2t^4 + 3t^3 - 2t^2 + 3t + 2)\\
&\qquad\qquad \times (400 + 1808t + 2527t^2 + 1155t^3 + 237t^4 + 17t^5),\\
R_{0,5} &= 3D_0^4(t-1)^7(51t^8 + 1081t^7 + 8422t^6 + 31914t^5 + 59639t^4 + 42461t^3\\
&\qquad\qquad + 7584t^2 - 2832t - 864)\\
&\quad + 2D_0^3(t-1)^4(249t^{10} + 3333t^9 + 22417t^8 + 105245t^7 + 339675t^6 +\\
&\qquad\qquad 664087t^5 + 513315t^4 + 127943t^3 - 6936t^2 - 1152t + 1296)\\
&\quad - D_0^2 t(t-1)^2(357t^{12} + 7089t^{11} + 58637t^{10} + 273500t^9 + 828314t^8\\
&\qquad\qquad + 1886278t^7 + 3638786t^6 + 5441836t^5 + 4731121t^4 + 1945329t^3\\
&\qquad\qquad + 179665t^2 - 96240t - 20304)\\
&\quad - 2D_0 t(t+3)^2(51t^{12} + 849t^{11} + 6580t^{10} + 33465t^9 + 115887t^8\\
&\qquad\qquad + 253743t^7 + 285517t^6 + 148083t^5 + 130634t^4 + 141380t^3\\
&\qquad\qquad + 59715t^2 + 4944t - 1200)\\
&\quad + 3t^2(t-1)(t+3)^3(2t^4 + 3t^3 - 2t^2 + 3t + 2)\\
&\qquad\qquad \times (400 + 1808t + 2527t^2 + 1155t^3 + 237t^4 + 17t^5),\\
R_{0,6} &= D_0(t-1)^2 + t^2 + t + 2,\\
R_{0,7} &= D_0^2(t-1)^4 + D_0 2(t-1)(t^3 + t^2 + 3t - 1) + t^4 + 4t^3 + 7t^2 + 2t + 2,\\
R_{2,0} &= 3(3t+1)(t+1)(-5t^5 + 6t^4 + 135t^3 + 664t^2 + 592t + 144),\\
R_{2,1} &= D_0^2(3t^3 - 12t^2 + 7t + 2) + D_0(6t^3 - 3t^2 + t) + 3t^3 + 9t^2,
\end{aligned}
$$

$$
\begin{aligned}
R_{2,2} &= D_0^3(3t^7 - 47t^5 - 18t^4 + 21t^3 + 164t^2 - 105t - 18) \\
&\quad + D_0^2(9t^7 + 36t^6 - 19t^5 - 168t^4 - 165t^3 + 292t^2 + 15t) \\
&\quad + D_0(9t^7 + 72t^6 + 190t^5 + 156t^4 - 63t^3 - 108t^2) \\
&\quad + 3t^7 + 36t^6 + 162t^5 + 324t^4 + 243t^3, \\
R_{2,3} &= D_0^2(3t^3 - 12t^2 + 7t + 2) + D_0(6t^3 - 3t^2 + t) + 3t^3 + 9t^2, \\
R_{2,4} &= D_0^3(3t^7 - 47t^5 - 18t^4 + 21t^3 + 164t^2 - 105t - 18) \\
&\quad + D_0^2(9t^7 + 36t^6 - 19t^5 - 168t^4 - 165t^3 + 292t^2 + 15t) \\
&\quad + D_0(9t^7 + 72t^6 + 190t^5 + 156t^4 - 63t^3 - 108t^2) \\
&\quad + 3t^7 + 36t^6 + 162t^5 + 324t^4 + 243t^3, \\
R_{3,0} &= D_0^2(t^4 - 2t^3 + 2t - 1) + D_0(2t^4 + 4t^3 - 2t^2 - 4t) + t^4 + 6t^3 + 9t^2, \\
R_{3,1} &= D_0^2(t^5 - 3t^4 + 2t^3 + 2t^2 - 3t + 1) + D_0(2t^5 + 6t^4 - 2t^3 - 6t^2) \\
&\quad + t^5 + 9t^4 + 27t^3 + 27t^2.
\end{aligned}
$$

References

1. R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.
2. D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.
3. D. Aldous. The continuum random tree. II: An overview. Stochastic analysis, Proc. Symp., Durham/UK 1990, Lond. Math. Soc. Lect. Note Ser. 167, 23-70 (1991)., 1991.
4. D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.
5. D. Aldous. Tree-based models for random distribution of mass. *J. Stat. Phys.*, 73(3-4):625–641, 1993.
6. E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *J. Combin. Theory Ser. A*, 15:91–111, 1973.
7. E. A. Bender. Asymptotic methods in enumeration. *SIAM Review*, 16(4):485–515, 1974.
8. E. A. Bender and E. R. Canfield. Face sizes of 3-polytopes. *J. Combin. Theory Ser. B*, 46(1):58–65, 1989.
9. E. A. Bender, Z. Gao, and N. C. Wormald. 2-connected labelled planar graphs. *Elec. J. Combinatorics*, 9(#43), 2002.
10. E. A. Bender and L. B. Richmond. Central and local limit theorems applied to asymptotic enumeration. II. Multivariate generating functions. *J. Combin. Theory Ser. A*, 34(3):255–265, 1983.
11. E. A. Bender and L. B. Richmond. The asymptotic enumeration of rooted convex polyhedra. *J. Combin. Theory Ser. B*, 36(3):276–283, 1984.
12. F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP '92 (Rennes, 1992)*, volume 581 of *Lecture Notes in Comput. Sci.*, pages 24–48. Springer, Berlin, 1992.
13. N. Bernasconi, K. Panagiotou, and A. Steger. On properties of random dissections and triangulations. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '08)*, pages 132–141.
14. N. Bernasconi, K. Panagiotou, and A. Steger. On the degree sequences of random outerplanar and series-parallel graph. In *Proceedings of the 12th International Workshop on Randomized Techniques in Computation (RANDOM '08), LNCS 5171, Springer Verlag*, pages 303–316, 2008.

15. P. Biane and M. Yor. Valeurs principales associées aux temps locaux browniens. *Bull. Sci. Math. (2)*, 111(1):23–101, 1987.
16. P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 1981.
17. P. Billingsley. *Probability and measure.* John Wiley & Sons Inc., New York, second edition, 1986.
18. P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, second edition, 1999.
19. M. Bodirsky, E. Fusy, M. Kang, and S. Vigerske. Enumeration and asymptotic properties of unlabeled outerplanar graphs. *Electronic J. Combinat.*, 14:R66.
20. M. Bodirsky, O. Giménez, M. Kang., and M. Noy. Enumeration and limit laws for series-parallel graphs. *European J. Combin.*, 28(8):2091–2105, 2007.
21. B. Bollobás. *Random graphs.* Cambridge University Press, Cambridge, second edition, 2001.
22. N. Bonichon, C. Gavoille, N. Hanusse, D. Poulalhon, and G. Schaeffer. Planar graphs, via well-orderly maps and trees. *Graphs and Combinatorics*, 22:185–202, 2006.
23. M. Bousquet-Mélou. Limit laws for embedded trees: applications to the integrated super-Brownian excursion. *Random Structures Algorithms*, 29(4):475–523, 2006.
24. M. Bousquet-Mélou and S. Janson. The density of the ISE and local limit laws for embedded trees. *Ann. Appl. Probab.*, 16(3):1597–1632, 2006.
25. J. Bouttier, P. Di Francesco, and E. Guitter. Geodesic distance in planar graphs. *Nucl. Phys., B*, 663(3):535–567, 2003.
26. J. Bouttier, P. Di Francesco, and E. Guitter. Statistics of planar graphs viewed from a vertex: A study via labeled trees. *Nucl. Phys., B*, 675(3):631–660, 2003.
27. N. Broutin, L. Devroye, E. McLeish, and M. de la Salle. The height of increasing trees. *Random Structures Algorithms*, 32:494–518, 2008.
28. N. Broutin and P. Flajolet. The height of random binary unlabelled trees. In *Fifth Colloquium on Mathematics and Computer Science (2008)*, Discrete Math. Theor. Comput. Sci. Proc., AC.
29. W. G. Brown. Enumeration of quadrangular dissections on the disk. *Can. J. Math.*, 17:302–317, 1965.
30. W. G. Brown and W. T. Tutte. On the enumeration of rooted non-separable planar maps. *Can. J. Math.*, 16:572–577, 1964.
31. E. R. Canfield. Remarks on an asymptotic method in combinatorics. *J. Combin. Theory Ser. A*, 37:348–352, 1984.
32. R. Carr, W. M. Y. Goh, and E. Schmutz. The maximum degree in a random tree an related problems. *Random Structures Algorithms*, 5:13–24, 1994.
33. P. Chassaing and J.-F. Marckert. Parking functions, empirical processes, and the width of rooted labeled trees. *Electron. J. Combin.*, 8(1):Research Paper 14, 19 pp., 2001.
34. P. Chassaing, J. F. Marckert, and M. Yor. The height and width of simple trees. In *Mathematics and computer science (Versailles, 2000)*, Trends Math., pages 17–30. Birkhäuser, Basel, 2000.
35. P. Chassaing and G. Schaeffer. Random planar lattices and integrated super-Brownian excursion. *Probab. Theory Relat. Fields*, 128(2):161–212, 2004.
36. B. Chauvin and M. Drmota. The random multisection problem, travelling waves and the distribution of the height of m-ary search trees. *Algorithmica*, 46(3-4):299–327, 2006.

37. B. Chauvin, M. Drmota, and J. Jabbour-Hattab. The profile of binary search trees. *Ann. Appl. Probab.*, 11(4):1042–1062, 2001.
38. B. Chauvin, T. Klein, J.-F. Marckert, and A. Rouault. Martingales and profile of binary search trees. *Electron. J. Probab.*, 10:no. 12, 420–435, 2005.
39. B. Chauvin and A. Rouault. Connecting yule process, bisection and binary search trees via martingales. *J. Iranian Statist. Soc.*, 3(2):89–116, 2004.
40. F. Chyzak, M. Drmota, T. Klausner, and G. Kok. The distribution of patterns in random trees. *Combin. Probab. Comput.*, 17:21–59, 2008.
41. J. Clément, P. Flajolet, and B. Vallée. Dynamical sources in information theory: A general analysis of trie structures. *Algorithmica*, 29(1–2):307–369, 2001.
42. J. W. Cohen and G. Hooghiemstra. Brownian excursion, the $M/M/1$ queue and their occupation times. *Math. Oper. Res.*, 6(4):608–629, 1981.
43. G. Dall'Aglio. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuoloa Norm. Sup. Pisa (3)*, 10:35–74, 1956.
44. R. de la Briandais. File searching using variable length keys. In *Proceedings of the AFIPS Spring Joint Computer Conference*, pages 295–298. AFIPS Press, Reston, Va., 1959.
45. C. Dellacherie and P.-A. Meyer. *Probabilités et potentiel*. Hermann, Paris, 1975.
46. J.-F. Delmas. Computation of moments for the length of the one dimensional ISE support. *Electron. J. Probab.*, 8(Paper No. 17):15 p., 2003.
47. A. Denise, M. Vasconcellos, and D. J. A. Welsh. The random planar graph. *Congr. Numer.*, 113:61–79, 1996.
48. L. Devroye. A probabilistic analysis of the height of tries and of the complexity of triesort. *Acta Inform.*, 21(3):229–237, 1984.
49. L. Devroye. A note on the height of binary search trees. *J. Assoc. Comput. Mach.*, 33(3):489–498, 1986.
50. L. Devroye. A study of trie-like structures under the density model. *Ann. Appl. Probab.*, 2(2):402–434, 1992.
51. L. Devroye. Universal asymptotics for random tries and PATRICIA trees. *Algorithmica*, 42(1):11–29, 2005.
52. L. Devroye and S. Janson. Distances between pairs of vertices and vertical profile in conditioned Galton–Watson trees. *manuscript*.
53. L. Devroye and J. Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures Algorithms*, 7(1):1–14, 1995.
54. R. P. Dobrow and R. T. Smythe. Poisson approximations for functionals or random trees. *Random Structures Algorithms*, 9(1–2):79–92, 1996.
55. A. A. Dobrynin, R. Entringer, and I. Gutman. Wiener index of trees: theory and applications. *Acta Appl. Math.*, 66(3):211–249, 2001.
56. M. Dondajewski and J. Szymański. On the distribution of vertex-degrees in a strata of a random recursive tree. *Bull. Acad. Polon. Sci. Sér. Sci. Math.*, 30(5-6):205–209, 1982.
57. M. Drmota. The height of increasing trees. *Ann. Combinat.* to appear.
58. M. Drmota. Distribution of the height of leaves of rooted trees. *Diskret. Mat.*, 6(1):67–82, 1994.
59. M. Drmota. Systems of functional equations. *Random Structures Algorithms*, 10(1-2):103–124, 1997.

60. M. Drmota. The saturation level in binary search tree. In *Mathematics and computer science (Versailles, 2000)*, Trends Math., pages 41–51. Birkhäuser, Basel, 2000.
61. M. Drmota. The variance of the height of binary search trees. *Theoret. Comput. Sci.*, 270(1-2):913–919, 2002.
62. M. Drmota. An analytic approach to the height of binary search trees. II. *J. ACM*, 50(3):333–374), 2003.
63. M. Drmota, O. Giménez, and M. Noy. Degree distribution in random planar graph. manuscript.
64. M. Drmota, O. Giménez, and M. Noy. Vertices of given degree in series-paralles graphs. *Random Structures Algorithms*, 2009. to appear.
65. M. Drmota and B. Gittenberger. The shape of unlabeled rooted random trees. manuscript.
66. M. Drmota and B. Gittenberger. On the profile of random trees. *Random Structures Algorithms*, 10(4):421–451, 1997.
67. M. Drmota and B. Gittenberger. The width of Galton-Watson trees conditioned by the size. *Discrete Math. Theor. Comput. Sci.*, 6(2):387–400, 2004.
68. M. Drmota, S. Janson, and R. Neininger. A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.*, 18(1):288–333, 2008.
69. M. Drmota and J.-F. Marckert. Reinforced weak convergence of stochastic processes. *Statist. Probab. Lett.*, 71(3):283–294, 2005.
70. M. Drmota and W. Szpankowsi. (Un)expected behavior of digital search tree profile. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '09).*
71. R. M. Dudley. *Real analysis and probability.* Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.
72. T. Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003.
73. T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281):vi+147 pages, 2002.
74. T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005.
75. S. N. Evans. *Probability and real trees*, volume 1920 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008.
76. S. N. Evans, J. Pitman, and A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields*, 134(1):81–126, 2006.
77. J. A. Fill and S. Janson. Smoothness and decay properties of the limiting Quicksort density function. In *Mathematics and Computer Science (Versailles, 2000)*, Trends Math., pages 53–64. Birkhäuser, Basel, 2000.
78. J. A. Fill and S. Janson. Approximating the limiting Quicksort distribution. *Random Structures Algorithms*, 19(3-4):376–406, 2001.
79. J. A. Fill and S. Janson. Quicksort asymptotics. *J. Algorithms*, 44(1):4–28, 2002.
80. P. Flajolet, Z. Gao, A. Odlyzko, and B. Richmond. The distribution of heights of binary trees and other simple trees. *Combin. Probab. Comput.*, 2(2):145–156, 1993.
81. P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.*, 144(1-2):3–58, 1995.
82. P. Flajolet and A. Odlyzko. The average height of binary trees and other simple trees. *J. Comput. System Sci.*, 25(2):171–213, 1982.

83. P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990.
84. P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge University Press, Cambridge, 2008.
85. P. Flajolet and J.-M Steyaert. A branching process arising in dynamic hashing, trie searching and polynomial factorization. In *Automata, languages and programming (Aarhus, 1982)*, volume 140 of *Lecture Notes in Comput. Sci.*, pages 239–251. Springer, Berlin, 1982.
86. E. Fredkin. Trie memory. *Communications of the ACM*, 3:490–499, 1960.
87. M. Fuchs, H.-K. Hwang, and R. Neininger. Profiles of random trees: limit theorems for random recursive trees and binary search trees. *Algorithmica*, 46(3-4):367–407, 2006.
88. J. Gastwirth and P. Bhattacharya. Two probability models of pyramids or chain letter schemes demo strating that their promotional claims are unreliable. *Operations Research*, 32:527–536, 1984.
89. S. Gerke and C. McDiarmid. On the number of edges in random planar graphs. *Combin. Probab. Comput.*, 13(2):165–183, 2004.
90. R. K. Getoor and M. J. Sharpe. Excursions of Brownian motion and Bessel processes. *Z. Wahrsch. Verw. Gebiete*, 47(1):83–106, 1979.
91. O. Giménez and M. Noy. Asymptotic enumeration and limit laws of planar graphs. *J. Amer. Math. Soc.* to appear, arXiv:math/0501269.
92. B. Gittenberger. On the contour of random trees. *SIAM J. Discrete Math.*, 12(4):434–458, 1999.
93. B. Gittenberger and G. Louchard. The Brownian excursion multi-dimensional local time density. *J. Appl. Probab.*, 36(2):350–373, 1999.
94. W. Goh and E. Schmutz. Limit distribution for the maximum degree of a random recursive tree. *J. Comput. Appl. Math.*, 142(1):61–82, 2002.
95. R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete mathematics.* Addison-Wesley, Reading, MA, 1989.
96. D. H. Greene and D. E. Knuth. *Mathematics for the analysis of algorithms.* Birkhäuser Boston Inc., Boston, MA, third edition, 1990.
97. I. Gutman and O. E. Polanski. *Mathematical Concepts in Organic Chemistry.* Springer, Berlin, 1986.
98. F. Harary and E. M. Palmer. *Graphical enumeration.* Academic Press, New York, 1973.
99. W. K. Hayman. A generalisation of Stirling's formula. *J. Reine Angew. Math.*, 196:67–95, 1956.
100. C. A. R. Hoare. Algorithm 64: Quicksort. *Communic. Assoc. Comput. Masch.*, 4:321, 1961.
101. C. A. R. Hoare. Quicksort. *Computer Journal*, 5:10–15, 1962.
102. M. Hofri. *Analysis of algorithms.* The Clarendon Press Oxford University Press, New York, 1995.
103. G. Hooghiemstra. On the explicit form of the density of Brownian excursion local time. *Proc. Amer. Math. Soc.*, 84(1):127–130, 1982.
104. H.-K. Hwang. *Théoremes limites pour les structures combinatoires et les fonctions arithmetiques.* PhD Thesis, École Polytechnique, New York, 1994.
105. H.-K. Hwang. Large deviations for combinatorial distributions. I. central limit theorems. *Ann. Appl. Probab.*, 6:297–319, 1996.
106. H.-K. Hwang. Profiles of random trees: plane-oriented recursive trees. *Random Structures Algorithms*, 30(3):380–413, 2007.

107. H.-K. Hwang and R. Neininger. Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.*, 31(6):1687–1722, 2002.
108. J. Jabbour-Hattab. Martingales and large deviations for binary search trees. *Random Structures Algorithms*, 19(2):112–127, 2001.
109. B. Jacquard and G. Schaeffer. A bijective census of nonseparable planar maps. *J. Combin. Theory Ser. A*, 83(1):1–20, 1998.
110. P. Jacquet and M. Régnier. Trie partitioning process: limiting distributions. In *CAAP '86 (Nice, 1986)*, volume 214 of *Lecture Notes in Comput. Sci.*, pages 196–210. Springer, Berlin, 1986.
111. P. Jacquet and W. Szpankowski. Analysis of digital tries with Markovian dependency. *IEEE Trans. Information Theory*, 37:1470–1475, 1991.
112. P. Jacquet and W. Szpankowski. Asymptotic behavior of the Lempel-Ziv parsing scheme and [in] digital search trees. *Theoret. Comput. Sci.*, 144(1-2):161–197, 1995.
113. P. Jacquet and W. Szpankowski. Analytical de-Poissonization and its applications. *Theoret. Comput. Sci.*, 201(1-2):1–62, 1998.
114. S. Janson. Random cutting and records in deterministic and random trees. *Random Structures Algorithms*, 29(2):139–179, 2006.
115. S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Structures Algorithms*, 4(3):231–358, 1993.
116. S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience, New York, 2000.
117. S. Janson and J.-F. Marckert. Convergence of discrete snakes. *J. Theor. Probab.*, 18(3):615–645, 2005.
118. A. Joffe, L. Le Cam, and J. Neveu. Sur la loi des grands nombres pour des variables aléatoires de Bernoulli attachées à un arbre dyadique. *C. R. Acad. Sci. Paris Sér. A-B*, 277:A963–A964, 1973.
119. O. Kallenberg. *Foundations of modern probability*. Springer-Verlag, New York, second edition, 2002.
120. I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, New York, 1988.
121. L. Kaup and B. Kaup. *Holomorphic functions of several variables*. Walter de Gruyter & Co., Berlin, 1983.
122. R. Kemp. *Fundamentals of the average case analysis of particular algorithms*. John Wiley & Sons Ltd., Chichester, 1984.
123. D. P. Kennedy. The distribution of the maximum Brownian excursion. *J. Appl. Probab.*, 13:371–376, 1976.
124. G. Kersting. On the height profile of a conditioned Galton-Watson tree. *manuscript*, 1998.
125. C. Knessl and W. Szpankowski. Quicksort algorithm again revisited. *Discrete Math. Theor. Comput. Sci.*, 3(2):43–64, 1999.
126. C. Knessl and W. Szpankowski. On the average profile of symmetric digital search trees. *manuscript*, 2008.
127. F. B. Knight. On the excursion process of Brownian motion. *Trans. Am. Math. Soc.*, 258:77–86, 1980.
128. D. E. Knuth. *The art of computer programming. Vol. 1*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, second edition, 1975.
129. D. E. Knuth. *The art of computer programming. Vol. 2*. Addison-Wesley Publishing Co., Reading, Mass., second edition, 1981.

130. D. E. Knuth. *The art of computer programming. Vol. 3.* Addison-Wesley Publishing Co., Reading, Mass., second edition, 1998.
131. V. F. Kolchin. Branching processes, random trees, and a generalized scheme of arrangements of particles. *Math. Notes*, 21:386–394, 1977.
132. V. F. Kolchin. *Random mappings.* Optimization Software Inc. Publications Division, New York, 1986.
133. V. F. Kolchin. *Random graphs.* Cambridge University Press, Cambridge, 1999.
134. M. Kuba and A. Panholzer. On the degree distribution of the nodes in increasing trees. *J. Combin. Theory Ser. A*, 114:597–618, 2007.
135. J.-F. Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.
136. J.-F. Le Gall. Random real trees. *Ann. Fac. Sci. Toulouse Série 6*, 15:35–62, 2006.
137. V. A. Liskovets. A pattern of asymptotic vertex valency distributions in planar maps. *J. Combin. Theory Ser. B*, 75(1):116–133, 1999.
138. C. L. Liu. *Introduction to combinatorial mathematics.* McGraw-Hill Book Co., New York, 1968.
139. M. Loève. *Probability theory. I.* Springer-Verlag, New York, fourth edition, 1977.
140. G. Louchard. The Brownian excursion area: a numerical analysis. *Comput. Math. Appl.*, 10(6):413–417 (1985), 1984.
141. G. Louchard. Kac's formula, Levy's local time and Brownian excursion. *J. Appl. Probab.*, 21(3):479–499, 1984.
142. G. Louchard. Exact and asymptotic distributions in digital and binary search trees. *RAIRO Inform. Théor. Appl.*, 21(4):479–495, 1987.
143. G. Louchard and W. Szpankowski. Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm. *IEEE Trans. Inf. Th.*, 41:478–488, 1995.
144. G. Louchard and W. Szpankowski. A probabilistic analysis of a string editing problem and its variations. *Combin. Probab. Comput.*, 4(2):143–166, 1995.
145. W. C. Lynch. More combinatorial properties of certain trees. *The Computer Journal*, 7(4):299–302, 1965.
146. H. M. Mahmoud. *Evolution of random search trees.* John Wiley & Sons Inc., New York, 1992.
147. H. M. Mahmoud. *Pólya urn modles.* Text in Statistical Science. Boca Raton, FL: CRC Press., 2009.
148. P. Major. On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.*, 8(4):487–517, 1978.
149. C. McDiarmid and B. Reed. On the maximum degree of a random planar graph. *Combin. Probab. Comput.*, 17(4):591–601, 2008.
150. C. McDiarmid, A. Steger, and D. J. A. Welsh. Random planar graphs. *J. Combin. Theory Ser. B*, 93:187–205, 2005.
151. A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canad. J. Math.*, 30(5):997–1015, 1978.
152. A. Meir and J. W. Moon. On an asymptotic method in enumeration. *J. Combin. Theory Ser. A*, 51(1):77–89, 1989.
153. A. Meir and J. W. Moon. On nodes of large out-degree in random trees. In *Proceedings of the Twenty-second Southeastern Conference on Combinatorics, Graph Theory and Computing (Baton Rouge, LA, 1991), Congr. Numer.*, 82:3–13, 1991.

154. H. Minc. *Nonnegative matrices.* John Wiley & Sons Inc., New York, 1988.
155. J. Moon. The distance between nodes in recursive trees. *London Math . Soc . Lecture Notes,* Cambridge Univ. Press, London, 13:125–132, 1974.
156. R. C. Mullin and P. J. Schellenberg. The enumeration of c-nets via quadrangulations. *J. Combin. Theory*, 4:259–276, 1968.
157. D. Najock and C. Heyde. On the number of terminal vertices in certain random trees with an application to stemma construction in Philology. *J. Applied Probab.*, 19:675–680, 1982.
158. R. Neininger. On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms*, 19(3-4):498–524, 2001.
159. R. Neininger. The Wiener index of random trees. *Combin. Probab. Comput.*, 11(6):587–597, 2002.
160. R. Neininger. *Stochastische Analyse von Algorithmen, Fixpunktgleichungen und ideale Metriken.* Habilitation Thesis, Universität Frankfurt a.M., 2004.
161. R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14(1):378–418, 2004.
162. R. Neininger and L. Rüschendorf. On the contraction method with degenerate limit equation. *Ann. Probab.*, 32(3B):2838–2856, 2004.
163. R. Neininger and L. Rüschendorf. A survey of multivariate aspects of the contraction method. *Discrete Math. Theor. Comput. Sci.*, 8(1):31–56, 2006.
164. J. Neveu. *Martingales à temps discret.* Masson et Cie, éditeurs, Paris, 1972.
165. A. M. Odlyzko. Asymptotic enumeration methods. In *Handbook of combinatorics, Vol. 1, 2*, pages 1063–1229. Elsevier, Amsterdam, 1995.
166. A. M. Odlyzko and H. S. Wilf. Bandwidths and profiles of trees. *J. Combin. Theory Ser. B*, 42(3):348–370, 1987.
167. D. Osthus, H. J. Prömel, and A. Taraz. On random planar graphs, the number of planar graphs and their triangulations,. *J. Combin. Theory Ser. B*, 88:119–134, 2003.
168. R. Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.
169. G. Park, H.-K. Hwang, P. Nicodème, and W. Szpankowski. Profiles of tries. *SIAM J. Computing*, page to appear.
170. J. Pitman. The SDE solved by local times of a brownian excursion or bridge derived from the height profile of a random tree or forest. *Ann. Probab.*, 27:261–283, 1999.
171. J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2006.
172. B. Pittel. On growing random binary trees. *J. Math. Anal. Appl.*, 103(2):461–480, 1984.
173. B. Pittel. Asymptotical growth of a class of random trees. *Ann. Probab.*, 13(2):414–427, 1985.
174. B. Pittel. Paths in a random digital tree: limiting distributions. *Adv. in Appl. Probab.*, 18(1):139–155, 1986.
175. B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994.
176. G. Polya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.*, 68:145–254, 1937.
177. H. Prodinger. Digital search trees and basic hypergeometric functions. *Bulletin of the EATCS*, 56, 1995.

178. S. T. Rachev. *Probability metrics and the stability of stochastic models.* John Wiley & Sons Ltd., Chichester, 1991.
179. S. T. Rachev and L. Rüschendorf. Probability metrics and recursive algorithms. *Adv. in Appl. Probab.*, 27(3):770–799, 1995.
180. S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I.* Springer-Verlag, New York, 1998.
181. B. Reed. The height of a random binary search tree. *J. Assoc. Comput. Mach.*, 50(3):306–332, 2003.
182. M. Régnier. A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.*, 23(3):335–343, 1989.
183. D. Revuz and M. Yor. *Continuous martingales and Brownian motion.* Springer-Verlag, Berlin, third edition, 1999.
184. R. J. Riddell. *Contribution to the theory of condensation.* PhD Thesis, Univ. Michigan, Ann Arbor, 1951.
185. J. M. Robson. On the concentration of the height of binary search trees. *ICALP 97 Proceedings, LNCS 1256*, pages 441–448, 1997.
186. U. Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.
187. U. Rösler. On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, 29(1-2):238–261, 2001.
188. G. Schaeffer. *Conjugaison darbres et cartes combinatoires aléatoires.* PhD Thesis, Université Bordeaux I, 1998.
189. E.-M. Schopp. *Tail Behaviour and Martingale Convergence of Random Recursive Structures and Algorithms.* PhD Thesis, Univ. Freiburg (Germany), 2008.
190. R. Sedgewick. *Quicksort.* Gurland Publishing, New York, 1980.
191. R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms.* Addison Wesley, Reading, MA, 1996.
192. R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Theory Probab. Math. Stat.*, 51:1–27, 1994.
193. V. E. Stepanov. The distribution of the number of vertices in the layers of a random tree. *Teor. Verojatnost. i Primenen.*, 14:64–77, 1969.
194. H. Sulzbach. A functional limit law for the profile of plane-oriented recursive trees. In *Fifth Colloquium on Mathematics and Computer Science (2008)*, Discrete Math. Theor. Comput. Sci. Proc., AC.
195. W. Szpankowski. A characterization of digital search trees from the successful search viewpoint. *Theoret. Comput. Sci.*, 85(1):117–134, 1991.
196. W. Szpankowski. On the height of digital trees and related problems. *Algorithmica*, 6(2):256–277, 1991.
197. W. Szpankowski. *Average case analysis of algorithms on sequences.* Wiley-Interscience, New York, 2001.
198. J. Szymański. On the maximum degree and the height of a random recursive tree. In *Random graphs '87 (Poznań, 1987)*, pages 313–324. Wiley, Chichester, 1990.
199. L. Takács. Conditional limit theorems for branching processes. *J. Appl. Math. Stochastic Anal.*, 4(4):263–292, 1991.
200. L. Takács. The asymptotic distribution of the total heights of random rooted trees. *Acta Sci. Math.*, 57:613–625, 1993.
201. K. H. Tan and P. Hadjicostas. Some properties of a limiting distribution of Quicksort. *Statistics Probab. Letter*, 25:87–94, 1995.

202. B. A. Trakhtenbrot. Towards a theory of non-repeating contract schemes. *Trudi Mat. Inst. Akad. Nauk SSSR*, 51:226–269, 1958.
203. W. T. Tutte. A census of planar triangulations. *Canad. J. Math.*, 14:21–38, 1962.
204. W. T. Tutte. A census of planar maps. *Canad. J. Math.*, 15:249–271, 1963.
205. W. T. Tutte. On the enumeration of planar maps. *Bull. Amer. Math. Soc.*, 74:64–74, 1968.
206. B. Vallée. Dynamical sources in information theory: fundamental intervals and word prefixes. *Algorithmica*, 29(1–2):262–306, 2001.
207. T. R. S. Walsh. Counting unlabelled three-connected and homeomorphically irreducible two-connected graphs. *J. Combin. Theory Ser. B*, 32(1):12–32, 1982.
208. J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory*, 24(5):530–536, 1978.
209. V. M. Zolotarev. Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Teor. Verojatnost. i Primenen.*, 21(4):741–758, 1976.

Index