Some New Methods and Results in Tree Enumeration

A. M. Odlyzko

Bell Laboratories
Murray Hill, New Jersey 07974

1. Introduction

A wide variety of new results in combinatorial enumeration have been obtained recently. Many of these results were prompted by new problems in computer science, while others answered old questions in combinatorics and other fields. The purpose of this note is to survey a subset of these new results, namely those dealing with tree enumeration. A selection of tree enumeration results is presented, together with a discussion of the methods that led to their derivation. No complete proofs are given; instead, the concentration is on heuristics and on the extent to which these methods can be generalized. The selection of techniques and results is made from those the author has worked on, and several other survey papers on related results could also be written without exhausting the subject.

The main unifying feature of the new methods discussed here is that they rely on intensive study of nonlinear analytic iterations involving generating functions. This is in marked contrast to classical tree enumeration, where the generating functions typically have only algebraic singularities. To oversimplify a little, it can be said that at least asymptotic enumeration of unlabeled trees of various kinds is very well understood. The basic method used there was developed by Pólya [32] and perfected by Otter [31]. Their method is presented also in [16], and it is so well understood that a few years ago a paper was written with the title ‘‘Twenty-step algorithm for determining the asymptotic number of trees of various species’’ [17]. The word ‘‘algorithm’’ in the title of that paper should not be interpreted literally; what those authors present is a sequence of steps, illustrated by examples, which are to be followed in enumerating unlabeled trees. If S_n denotes the number of rooted trees of a particular kind with n vertices, and s_n the number of unrooted trees with n vertices, the generating functions are defined by

$$S(z) = \sum_{n=1}^{\infty} S_n z^n ,$$

$$s(z) = \sum_{n=1}^{\infty} s_n z^n .$$

It is usually the case that these power series for $S(z)$ and $s(z)$ converge in some disk $|z| < \rho$, and have algebraic singularities at $z = \rho$ of the kind

$$S(z) = c_1 - c_2(\rho - z)^{1/2} + O(|\rho - z|) , \quad (1.1)$$

$$s(z) = c_3 + c_4(\rho - z)^{3/2} + O(|\rho - z|^2) , \quad (1.2)$$

(where the c_i are constants depending on the trees being considered), and no other singularities on $|z| = \rho$. It can then be deduced using the Darboux method [10] that for some constants b_1 and b_2,

$$S_n \sim b_1 n^{-3/2} \rho^{-n} \quad \text{as} \quad n \to \infty ,$$

$$s_n \sim b_2 n^{-5/2} \rho^{-n} \quad \text{as} \quad n \to \infty .$$

The above sketch should not be taken to imply that the Pólya method is trivial. Many of the steps leading to (1.1) and (1.2), which involve obtaining relations involving the generating functions $S(z)$ and $s(z)$, can be quite complicated. Moreover, very often it is impossible to determine constants such as the all-important radius of convergence ρ in closed form, and so although these constants can be calculated efficiently to great accuracy, we do not know whether they are algebraic or transcendental. However, the main point of this paragraph is to indicate that at least the basics of the Pólya enumeration theory are fairly well understood.

Beautiful and useful as the Pólya theory is, it only applies to a restricted class of tree enumeration problems, and there is a wide literature (*cf.* [4,16,26]) on other methods. Here we will be concerned with some recent developments that involve nonlinear iteration. As a simple example, we consider the problem of height distribution among binary trees. Binary trees are rooted trees in which every node has either 0 or 2 sons, and left and right sons are distinguished. The size of a binary tree is the number of its internal binary nodes; i.e., the number of nodes with two sons. We let B_n denote the number of binary trees of size *n*, so that $B_0 = B_1 = 1$. Since any binary tree *T* of size > 0 is obtained by making the roots of two other binary trees be the sons of the root of *T*, and since left and right successors are distinguished we obtain the recurrence

$$B_n = \sum_{n_1 + n_2 = n-1} B_{n_1} B_{n_2} \ , \quad n \geq 1 \ . \tag{1.3}$$

Hence, if we define the generating function

$$B(z) = \sum_{n=0}^{\infty} B_n z^n \ , \tag{1.4}$$

then (1.3) yields the functional equation

$$B(z) = 1 + zB(z)^2 \ , \tag{1.5}$$

valid in the ring of formal power series in *z*. Equation (1.5) can be solved explicitly; we find that

$$B(z) = \frac{1-\sqrt{1-4z}}{2z} . \tag{1.6}$$

We immediately see that $B(z)$ is analytic in the entire complex domain with the exception of an algebraic singularity (a square root one) at $z = 1/4$. Hence an application of the Darboux method [10] easily yields the result

$$B_n \sim \pi^{-1/2} n^{-3/2} 4^n \quad \text{as} \quad n \to \infty . \tag{1.7}$$

In this particular case there is no need to use the Darboux method. If we expand $(1-4z)^{1/2}$ by the binomial theorem, we find that

$$B_n = \frac{1}{n+1} \binom{2n}{n} , \tag{1.8}$$

so that B_n is one of the famous Catalan numbers. Furthermore, Eq. (1.8) can be obtained by more combinatorial methods, without any use of generating functions. The point behind the presentation above is to exhibit a case where the classical enumerative techniques apply in a very straightforward way. In the next two sections we will discuss other problems arising in the enumeration of binary trees by height which appear to require more sophisticated methods.

2. Enumeration of trees of a given height

In many situations it is important to study heights of trees, and not just the total number of trees of a given size (*cf.* [2,21,22]). In this section we present some results on the enumeration of trees by size when the height is fixed. For a binary trees, we assign height 0 to the empty tree, height 1 to the tree consisting of just the root and two leaves, and in general height h to the tree in which the longest path from the root to a leaf has length h. We let $B_{h,n}$ denote the number of binary trees of size (number of internal nodes) n and height $\leq h$. We wish to study $B_{h,n}$ as a function of n for h fixed (and large).

No formula for $B_{h,n}$ is known that even remotely approaches the formula (1.8) for B_n in simplicity and utility. However, it is possible to show that if $0 < \lambda < 1$, then as $h, n \to \infty$ in such a way that $n2^{-h} \to \lambda$,

$$f(\lambda) = \lim \; 2^{-h} \log B_{h,n} \tag{2.1}$$

exists and is a certain easily computable function which will be described below, and a graph of which is shown in Fig. 1. The key to the proof lies in the existence of relations among the generating functions of the $B_{h,n}$. If we let

$$B_h(z) = \sum_{n=0}^{\infty} B_{h,n} z^n \ , \tag{2.2}$$

which is a polynomial with $B_{h,n} = 0$ for $n \geq 2^h$, then we have $B_0(z) = 1$ and

$$B_{h+1}(z) = 1 + zB_h(z)^2 \quad \text{for} \quad h \geq 0 \ , \tag{2.3}$$

since a nonempty tree of height $\leq h+1$ consists of a root with two trees of heights $\leq h$ as its descendants. The recurrence (2.3) clearly gives a method of computing all of the $B_{h,n}$ rather quickly, at least if h is small. For large h, though, no method was known until recently for analyzing the behavior of the $B_{h,n}$. An extensive study of the analytic properties of the recurrence (2.3) has given very precise estimates of the $B_{h,n}$ [15], of which (2.1) is just a special case.

The results of [15] cover much more than the enumeration of binary trees of a given height. They give estimates of coefficients $y_{h,n}$ of sequences $y_0(z)$, $y_1(z)$,... of polynomials

$$y_h(z) = \sum_{n=0}^{\infty} y_{h,n} z^n \tag{2.4}$$

which are called PNI-sequences (for positive nonlinear iteration). PNI-sequences are defined by some initial $y_0(z) \neq 0$ which has nonnegative coefficients and a recurrence

$$y_{h+1}(z) = P(z\,,\, y_h(z))\ ,\quad h \geq 0\ , \tag{2.5}$$

where $P(z,y)$ is a polynomial with non-negative coefficients,

$$P(z,y) = \sum_{0 \leq k \leq d} p_k(z) y^k \quad \text{with} \quad p_d(z) \neq 0\ ,\ \ d > 1\ . \tag{2.6}$$

Define

$$\mu = \lim_{h \to \infty} d^{-h} \deg y_h(z)\ ,$$

$$\rho = \inf \{ x : x \in \mathbb{R}^+\ ,\ \ y_h(x) \to \infty \ \text{ as } \ h \to \infty \}\ .$$

Clearly μ and ρ exist and are finite for every PNI-sequence $\{ y_h(z) \}$ that contains non-constant polynomials. As is shown in [15], it is sufficient to consider PNI-sequences in which $P(z,y)$ and $y_0(z)$ satisfy the following conditions:

(A) $P(z,y)$ is not a monomial (i.e., $P(z,y) \neq bz^a y^d$).

(B) At least one of the y_h, $0 \leq h \leq 2$ has the property that $|\, y_h(z)\, | = y_h(1)$ and $|\, z\, | = 1 \Rightarrow z = 1$.

The main result proved in [15] is the following one.

Theorem 2.1. *Suppose that $\{ y_h(z) \}$ is a PNI-sequence that satisfies conditions (A) and (B), and let λ_1 and λ_2 be any real numbers that satisfy*

$$0 < \lambda_1 < \lambda_2 < \mu\ .$$

Then for any integers n and h with

$$\lambda_1 \le nd^{-h} \le \lambda_2$$

we have, uniformly in n and h,

$$y_{h,n} = \frac{rp_d(r)^{-1/(d-1)} \exp(d^h(\beta(r) - r\beta'(r) \log r))}{d^{n/2}\sqrt{2\pi(r^2\beta''(r) + r\beta'(r))}} (1 + O(d^{-h/2})) , \quad (2.7)$$

where r is the unique solution in (ρ , ∞) *of*

$$r\beta'(r) = nd^{-h} ,$$

and $\beta(z)$ *is a function which is defined on* (ρ , ∞) *by*

$$\beta(z) = \log y_0(z) + \frac{1}{d-1} \log p_d(z) + \sum_{j=0}^{\infty} d^{-j-1} \log \left\{ \frac{y_{j+1}(z)}{p_d(z) y_j(z)^d} \right\} , \quad (2.8)$$

and is analytic there.

Some further results are proved in [15]. In particular, it is shown there that if $\rho < 1$, then the $y_{h,n}$ as functions of n are asymptotically gaussian near the peak.

Theorem 2.1 gives information about the $y_{h,n}$, the coefficients of the $y_h(z)$, in terms of the function $\beta(z)$, which is defined by (2.8) in terms of the $y_h(z)$. The circularity here is only apparent, since the series (2.8) converges very rapidly, and the first few terms suffice to compute it to great accuracy.

In the case of binary trees, with $y_h(z) = B_h(z)$, we have $\rho = 0$, $\mu = 1$, $d = 2$, and

$$\beta(z) = \log z + \sum_{j=1}^{\infty} 2^{-j} \log \frac{B_h(z)}{B_h(z) - 1} .$$

Theorem 2.1 implies that for fixed h, $B_{h,n}$ is maximized for

$$n \sim 2^h \ 0.6289... \quad \text{as} \quad h \to \infty \ , \tag{2.9}$$

and its maximum value is asymptotic to

$$2^{-h/2} \exp(2^h \ 0.4072...) \ 0.6855... \quad \text{as} \quad h \to \infty \ . \tag{2.10}$$

For $h = 9$, $B_{9,n}$ is maximized for $n = 322$, as predicted by (2.9), and the value of $B_{9,322}$ differs from that predicted by (2.10) by less than 0.05%.

The proof of Theorem 2.1 presented in [15] depends on a careful study of the analytic behavior of the polynomials $y_h(z)$ as $h \to \infty$ for $z \in C$. It is shown there that on any circle $| z | = r$, $r > \rho$, $y_h(z)$ is large only in a very small neighborhood of the real axis, where

$$y_h(z) = g(z) \alpha(z)^{d^h} (1 + o(1)) \quad \text{as} \quad h \to \infty \ , \tag{2.11}$$

where $g(z)$ and $\alpha(z) = \exp(\beta(z))$ are certain analytic functions. (This result generalizes some earlier work on numerical recurrences by Aho and Sloane [3] and Reingold [33].) The coefficients $y_{h,n}$ are studied by means of the Cauchy formula

$$y_{h,n} = \frac{1}{2\pi i} \int_{|z|=r} y_h(z) z^{-n-1} dz \ .$$

The radius r is chosen to make the value of the integral at $z = r$ (where the maximum over the circle $| z | = r$ is located) as small as possible, and the theorem is obtained by an application of the saddle-point method [8]. The details are quite complicated, but the principal ideas of the method are quite straightforward.

As is discussed in [15], many of the hypotheses of Theorem 2.1 can be weakened considerably. What is crucial is not the nonnegativity of the various coefficients so much as the doubly exponential growth (2.11) of the $y_h(z)$, which follows from that nonnegativity. Further, one does not have to restrict attention to polynomials $P(z,y)$ in the recurrence (2.5). However, when $P(z,y)$ is not a polynomial, one can often obtain asymptotic estimates by using more classical estimates [18,19,29].

3. Average heights of trees

In the preceding section we presented some results on the enumeration of trees of a given height, so that we held the height fixed and varied the size of the trees. In many cases what is needed, though, is information about the distribution of heights among trees of a given size, so that we need to hold the size fixed and vary the height. The first results of this kind which involved nonlinear analytic iteration were obtained by Rényi and Szekeres [34] in a study of rooted nonplanar labeled trees. By means of an extensive study (which relied heavily on [36]) of the sequence of functions $G_0(z)$, $G_1(z), \ldots,$ where $G_0(z) = z$ and

$$G_{h+1}(z) = z \exp(G_h(z)) , \quad h \geq 0 , \tag{3.1}$$

they showed that the average height of rooted nonplanar labeled trees with a nodes is asymptotic to $(2\pi n)^{1/2}$ as $n \to \infty$. Furthermore, Rényi and Szekeres obtained the distribution of heights among such trees, and showed that it was the ubiquitous theta distribution; if H_n denotes the height of a random rooted nonplanar labeled tree with n nodes, then for any $x > 0$,

$$\lim_{n \to \infty} Pr(H_n (2n)^{-1/2} < x) = 4x^{-3} \pi^{5/2} \sum_{k=1}^{\infty} k^2 \exp(-\pi^2 k^2 / x^2) . \tag{3.2}$$

The same result was proved later by a different method by Stepanov [35]. (Similar results for diameters were later obtained in a very interesting paper of Szekeres [37].)

Considerably more general, but very similar, results were obtained later in [13,14]. Those papers are concerned largely with heights in what Meir and Moon [24] call simple families of trees. These families consists of planar trees with labels attached to nodes. All labels are taken from a fixed set $L = L_0 \cup L_1 \cup L_2 \cup \ldots$, where the L_i are disjoint finite sets, and L_r is the set of labels that may be attached to a node of degree r. If we let y_n be the number of trees with a total of exactly n nodes (not just internal nodes), then the generating function

$$y(z) = \sum_{n=1}^{\infty} y_n z^n \tag{3.3}$$

satisfies an equation of the form

$$y(z) = z\phi(y(z)) , \tag{3.4}$$

where

$$\phi(y) = \sum_{r=0}^{\infty} | L_r | y^r . \tag{3.5}$$

We will assume that the expansion (3.5) converges in some neighborhood of the origin. Furthermore, if $y_{h,n}$ is the number of trees in the family of size n and height $\leq h$, and

$$y_h(z) = \sum_{n=1}^{\infty} y_{h,n} z^n , \tag{3.6}$$

then $y_0(z) = 0$ and

$$y_{h+1}(z) = z\phi(y_h(z)) , \quad h \geq 0 . \tag{3.7}$$

Many of the important families of trees fall into this category. For example, binary trees have $|L_r| = 1$ for $r = 0$ and $r = 2$, and $|L_r| = 0$ for other r. (For binary trees, the old and the new notation are related by the equation $y(z) = zB(z^2)$, since size in the new notation refers to the total number of nodes.) Furthermore, the results of [14] apply to many other families of trees, such as the rooted nonplanar labeled trees studied by Rényi and Szekeres, which satisfy relations of the type (3.7), with $\phi(y)$ having a Taylor series with nonnegative coefficients which converges in some neighborhood of the origin.

For simple families of trees, the asymptotic behavior of y_n can be determined from the Pólya theory. Meir and Moon [24] also succeeded in determining many other quantities for these trees, such as the average profiles. The distribution of heights is not amenable to this kind of analysis, and requires new method. In [14] the following result is proved.

Theorem 3.1. *If $n \to \infty$ though values of n such that $n \equiv 1$ (*mod *d), where $d = GCD\{r : |L_r| \neq 0\}$, the average height of trees of size n in a simple family corresponding to $y = z\phi(y)$ is asymptotic to $\lambda n^{1/2}$, where*

$$\lambda = \phi'(\tau)(2\pi)^{1/2}(\phi(\tau)\phi''(\tau))^{-1/2} , \tag{3.8}$$

*and τ is the smallest positive root of the equation $\phi(\tau) = \tau\phi'(\tau)$. (If $n \not\equiv 1$ (*mod *d), there are no trees of size n at all.)*

It is also shown in [14] that a distribution result like (3.2) holds for simple families of trees, but with the scaling factor $(2n)^{-1/2}$ replaced by $(\xi n)^{-1/2}$, where $\xi = \lambda^2/\pi$.

The analysis of [14] relies on a careful study of the iteration (3.7) near the principal singularity. Here we will briefly sketch the analysis in the case $d = 1$. It is easy to show (*cf.* [14,24]) that $y(z)$ is analytic in $|z| < \rho$, where $\rho = \tau/\phi(\tau)$, and has an algebraic singularity at $z = \rho$;

$$y(z) = \tau - (2\phi(\tau)/\phi''(\tau))^{1/2}(1-z/\rho)^{1/2} + O(|1-z/\rho|) . \tag{3.9}$$

This is one step in the application of the Darboux-Pólya theory, and it leads to an asymptotic estimate for y_n. To estimate the average height, it suffices to study the sum H_n of the heights of all trees of size n;

$$H_n = \sum_{h=1}^{\infty} h(y_{h,n} - y_{h-1,n}) = \sum_{h=1}^{\infty} (y_n - y_{h,n}) . \tag{3.10}$$

The generating function of H_n is

$$H(z) = \sum_{n=1}^{\infty} H_n z^n = \sum_{h=1}^{\infty} (y(z) - y_h(z)) . \tag{3.11}$$

Information about H_n is obtained from information about the behavior of $H(z)$ near the singularity $z = \rho$. Since y_n satisfies

$$y_n \sim c\rho^{-n} n^{-3/2} \quad \text{as} \quad n \to \infty$$

for some constant $c > 0$, and the average height was expected to be $\sim c' n^{1/2}$, it was expected that H_n ought to be $\sim cc'\rho^{-1} n^{-1}$, so that $H(z)$ was expected to have a logarithmic singularity at ρ. The proof that this is indeed the case used the expansion (3.11) and a very careful study of the differences $y(z) - y_h(z)$ near $z = \rho$. If we let

$$e_h(z) = y(z) - y_h(z) ,$$

$$\varepsilon(z) = 1 - z\phi'(y(z)) ,$$

then the relations

$$y(z) = z\phi(y(z))$$

and

$$y_{h+1}(z) = z\phi(y_h(z))$$

can be used to show that in a suitable region near $z = \rho$,

$$e_{h+1}(z) = e_h(z)(1-\varepsilon(z))(1-\phi''(\tau)e_h(z)/(2\phi'(\tau)) + g_h(z)) ,$$

where

$$g_h(z) = O(|e_h(z)|^2 + |e_h(z)(y(z)-\tau)|) ,$$

and from this it can be deduced that near $z = \rho$,

$$e_h(z) \sim \frac{c''\varepsilon(z)(1-\varepsilon(z))^h}{1 - (1-\varepsilon(z))^h} \tag{3.12}$$

for an appropriate constant c''. The behavior of $H(z)$ near $z = \rho$ can then for deduced from (3.12) (or, to be exact, from the more precise result proved in [14]), and the estimate for average heights follows.

4. Enumeration of balanced trees

Several difficult problems arise in the enumeration of trees which have balance conditions imposed on them. As an example, we consider 2,3-trees [27]. These are rooted planar trees, each of whose nonleaf nodes has either two or three successors, and (this is the crucial condition) all of whose root-to-leaf paths have the same length. The size of a 2,3-tree is the number of its leaves (since that is the quantity of interest in considering these trees as data structures), with the tree consisting of the root alone being regarded as having size 1 and height 0. Let a_n denote the number of 2,3-trees of size n, and $a_{h,n}$ the number of size n and height exactly h. Also let

$$f(z) = \sum_{n=1}^{\infty} a_n z^n , \tag{4.1}$$

$$Q_h(z) = \sum_{n=1}^{\infty} a_{h,n} z^n , \tag{4.2}$$

with $Q_0(z) = z$. These generating functions turned out to be crucial in the asymptotic estimation of the a_n [27]. Relatively elementary methods had shown [25] that

$$c_1 n^{-1} \phi^n < a_n < c_2 n^{-1} \phi^n , \tag{4.3}$$

where c_1 and c_2 are some positive constants and $\phi = (1 + \sqrt{5})/2$ is the ''golden ratio,'' but the question was left open as to whether $a_n \sim c_3 n^{-1} \phi^n$ as $n \to \infty$ for some constant c_3. It turns out that such a relation does not hold, and the asymptotic behavior of the a_n displays periodic oscillations. (Periodic oscillations occur rather frequently in combinatorial enumeration, as is shown in [7,11].)

Enumeration of unbalanced 2,3-trees (i.e., those in which the leafs are not all required to be at the same distance from the root) is easy; such a tree either consists of just the root or else of a root and either two or three subtrees, each of which is an unbalanced 2,3-tree. Therefore the enumerator $F(z)$ of unbalanced 2,3-trees satisfies

$$F(z) = z + F(z)^2 + F(z)^3 , \tag{4.4}$$

and this gives an explicit representation of $F(z)$ as an algebraic function. It is easy to obtain asymptotic estimates for the number of unbalanced 2,3-trees from (4.4).

The enumeration of standard (i.e., balanced) 2,3-trees is considerably more complicated. Since a tree of height $h+1$ ($h \geq 0$) consists of a root which has either two or three descendants, all of height h, we obtain

$$Q_{h+1}(z) = Q_h(z)^2 + Q_h(z)^3 , \quad h \geq 0 . \tag{4.5}$$

This relation, and the results described in Section 2, give very good estimates for the distribution by size of 2,3-trees of a given height. Another way to obtain relations among the $Q_h(z)$ is to note that each tree of height $h+1$ is obtained from a unique tree of height h by appending either two or three successors to each leaf, and so

$$Q_{h+1}(z) = Q_h(z^2 + z^3) , \quad h \geq 0 . \tag{4.6}$$

The recurrence (4.6) turns out to be more useful than (4.5) in estimating the total number of 2,3-trees of a given size. Since

$$f(z) = \sum_{h=0}^{\infty} Q_h(z) , \tag{4.7}$$

we find that

$$f(z) = z + f(z^2 + z^3) . \tag{4.8}$$

This relation, like the preceding ones, is valid in the ring of formal power series in z.

The functional equation (4.8) can be used to obtain to the asymptotic enumeration of 2,3-trees. Essentially the same method also yields the following more general result [27].

Theorem 4.2. *Let $P(z)$ and $Q(z)$ be nonzero polynomials with real, nonnegative coefficients. Assume that $P(0) = Q(0) = Q'(0) = 0$. Write*

$$Q(z) = \sum_{k=0}^{K} q_k z^{e_k} , \quad 2 \leq e_0 < e_1 < \ldots < e_K , \tag{4.9}$$

where $q_k > 0$ for $0 \leq k \leq K$, and assume that $K > 0$ and $GCD(e_1 - e_0 , e_2 - e_0 , \ldots, e_K - e_0) = 1$. Let α be the unique positive root of

$Q(z) = z$, and set $\beta = Q'(\alpha)$. If

$$f(z) = \sum_{n=1}^{\infty} a_n z^n$$

is a formal power series in z which satisfies the functional equation

$$f(z) = P(z) + f(Q(z)) \ , \tag{4.10}$$

then

$$a_n \sim n^{-1} \alpha^{-n} u(\log n) \quad \text{as} \quad n \to \infty \ , \tag{4.11}$$

where $u(x)$ is a nonconstant positive continuous function which is periodic with period $\log \beta$. *The average value of $u(x)$ is $P(\alpha)(\log \beta)^{-1}$.*

In the case of 2,3-trees, $P(z) = z$ and $Q(z) = z^2 + z^3$, so $\alpha = \phi^{-1}$, $\beta = 4 - \phi$. The results of [27] do not provide a good expansion of $u(x)$, although they do show that it's a very smooth function. Numerically, it appears that in the case of 2,3-trees, $u(x)$ oscillates between 0.682... and 0.806... .

The proof of Theorem 4.1 relies on a study of the functional equation (4.10). If we define

$$Q_0(z) = z \ , \quad Q_{h+1}(z) = Q(Q_h(z)) \quad \text{for} \quad h \geq 0 \ , \tag{4.12}$$

then iteration of the relation (4.10) shows that

$$f(z) = \sum_{h=0}^{\infty} P(Q_h(z)) \ , \tag{4.13}$$

at least in the ring of formal power series in z. It is easy to show that $f(z)$ is analytic in $|z| < \alpha$, and that (4.13) holds there as a relation among analytic functions.

Furthermore, $f(z)$ has a singularity at $z = \alpha$, so the radius of convergence of the Taylor series of $f(z)$ is exactly α. An important part in the analysis in [27] is played by the fact that the sum on the right side of (4.13) converges in a region larger than the disk $|z| < \alpha$, and so provides an analytic continuation of $f(z)$ outside its circle of convergence. The sum on the right side of (4.13) converges at z precisely when $Q_h(z) \to 0$ as $h \to \infty$, and when that happens, that sum is analytic at z.

In the results presented in Section 2, the main interest was in the behavior of $Q_h(z)$ for those z for which $|Q_h(z)| \to \infty$ as $h \to \infty$, and it was necessary to obtain delicate estimates for the rate of divergence. The proof of Theorem 4.1, on the other hand, deals only with those z for which $Q_h(z) \to 0$ as $h \to \infty$. The literature on the general subject of iteration of rational maps is immense (*cf.* [1,5,9,12,20,23,36]) and constantly growing, especially because of the new impetus given to that subject by work in mathematical physics. Some of that work is closely related to the investigations that were undertaken in connection with tree enumeration. From the classical results of Fatou [12] and Julia [20] it can be concluded, for example, that in the case of 2,3-trees, the boundary of the region for which $Q_h(z) \to 0$ as $h \to \infty$ is a nowhere differentiable continuous Jordan curve, and that $|Q_h(z)| \to \infty$ as $h \to \infty$ for z on the outside of that curve. (A picture of the curve is shown in Fig. 2.) The proof of Theorem 4.1 does not utilize that fact, however, but relies on the fact, which does not seem to follow from previously known results, that $Q_h(z) \to 0$ as $h \to \infty$ in the region

$$\{z : |z| < \alpha + \delta\} \cap \{z : |\mathrm{Arg}(z - \alpha)| > \frac{\pi}{2} - \delta\} \tag{4.14}$$

for some $\delta > 0$. In that region it is shown that

$$f(z) = c \log(1-\alpha^{-1} z) + w(\log(\alpha - z)) + O(|z-\alpha|) \tag{4.15}$$

for $c = P(\alpha)(\log \beta)^{-1}$, where $w(z)$ is a function analytic in $|\mathrm{Im}(z)| < \pi/2 + \delta$ and is periodic with period $\log \beta$. The derivation of Theorem 4.1 from the expansion (4.15) is then relatively straightforward.

The method sketched very briefly above can also be adapted to other, more complicated problems. For example, in the enumeration of AVL-trees (height-balanced binary trees), one encounters the problem of estimating the coefficients a_n of

$$f(z) = \sum_{n=1}^{\infty} a_n z^n ,$$

where $f(z)$ is given by

$$f(z) = \sum_{h=0}^{\infty} Q_h(z) ,$$

and

$$Q_0(z) = z , \quad Q_1(z) = z^2 , \tag{4.16}$$

$$Q_{h+2}(z) = Q_{h+1}(z)(Q_{h+1}(z) + 2\, Q_h(z)) , \quad h \geq 0 .$$

In this case $f(z)$ appears not to satisfy any simple functional equation, but by an intensive study of the iteration (4.16) it can be shown [28] that $f(z)$ has an expansion of the form (4.15), and from that it can be concluded that

$$a_n \sim n^{-1} \alpha^{-n} u(\log n) \quad \text{as} \quad n \to \infty ,$$

where $\alpha = 0.5219024...$ is a certain constant, which is not known in closed form, but can be computed rapidly, and $u(x)$ is a continuous function that is periodic with period $\log \beta$, $\beta = (2 + \sqrt{10})/3$.

5. Final remarks

The preceding sections presented a selection of tree enumeration results. At first glance they might seem to have very little in common with each other. In some cases sequences of polynomials are studied in regions where they diverge to infinity, in other cases they are investigated in regions where they converge to zero. In still other cases that could be cited, such as the work on probabilities of very large or very small heights that was done in [38] and is further developed in [30], the corresponding sequences of generating functions are studied practically right on the boundary between convergence and divergence. What this seemingly great variety of different approaches conceals is the basic similarity of the techniques used, all of which depend on very intensive and precise studies of nonlinear iterations involving generating functions. These studies yield results about the analytic behavior of the appropriate generating functions that are sufficient to obtain asymptotic estimates of the coefficients of these functions. Although the methodology used on the problems presented in this survey is not systematic enough to permit the writing of a paper with a title like that of [17], it is hoped that the variety of results and methods that have been discussed will be a convincing demonstration of the utility of nonlinear iteration methods and will lead to further research in that field.

REFERENCES

[1] J. Aczél, *Lectures on Functional Equations and Their Applications,* Academic Press, New York, 1966.

[2] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *The Design and Analysis of Computer Algorithms,* Addison-Wesley, 1974.

[3] A. V. Aho and N. J. A. Sloane, Some doubly exponential sequences, *Fibonacci Quart.* **11** (1973), 429-437.

[4] E. A. Bender, Asymptotic methods in enumeration, *SIAM Review*, **16** (1974), 485-515.

[5] H. Brolin, Invariant sets under iteration of rational functions, *Arkiv för Mat.* **6** (1965), 103-144.

[6] G. B. Brown and B. O. Shubert, On random binary trees, to be published.

[7] N. G. de Bruijn, On Mahler's partition problem, *Nederl. Akad. Wetensch. Proc.* **51** (1948), 659-669 and *Indag. Math.* **10** (1948), 210-220.

[8] N. G. de Bruijn, *Asymptotic Methods in Analysis,* North-Holland, Amsterdam, 1958.

[9] P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems,* Birkhäuser, 1980.

[10] G. Darboux, Mémoire sur l'approximation des fonctions de trés grands nombres, et sur une classe étendu de développements en série, *J. Math. Pures et Appliq.* (3)

4 (1878), 5-56.

[11] P. Erdös and B. Richmond, Concerning periodicity in the asymptotic behavior of partition functions, *J. Austral. Math. Soc., Ser. A,* **21** (1976), 447-456.

[12] P. Fatou, Memoire sur les équations fonctionnelles, *Bull. Soc. Math. France* **47** (1919), 161-271; **48** (1920), 33-94 and 208-314.

[13] P. Flajolet and A. M. Odlyzko, Exploring binary trees and other simple trees, in *Proceedings of 21st IEEE Found. Computer Sci. Symposium,* New York (1980), pp. 207-216.

[14] P. Flajolet and A. M. Odlyzko, The average height of binary trees and other simple trees, *J. Computer Syst. Sci.,* **25** (1982), 171-213.

[15] P. Flajolet and A. M. Odlyzko, Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumerations, *Math. Proc. Cambridge Phil. Soc.*, to appear.

[16] F. Harary and E. M. Palmer, *Graphical Enumeration,* Academic Press, New York and London, 1973.

[17] F. Harary, R. W. Robinson, and A. J. Schwenk, Twenty-step algorithm for determining the asymptotic number of trees of various species, *J. Austr. Math. Soc., Ser. A.,* **20** (1975), 483-503.

[18] B. Harris and L. Schoenfeld, Asymptotic expansions for the coefficients of analytic functions, *Illinois J. Math.,* **12** (1968), 264-277.

[19] W. K. Hayman, A generalization of Stirling's formula, *J. reine angew. Math.*, **196** (1956) 67-95.

[20] G. Julia, Mémoire sur l'iteration des fonctions rationnelles, *J. Math. Pures Appl.* (8) **1** (1918), 47-245.

[21] D. E. Knuth, *The Art of Computer Programming: Fundamental Algorithms*, Addison Wesley, Reading (1968).

[22] D. E. Knuth, *The Art of Computer Programming: Sorting and Searching,* Addison Wesley, Reading (1973).

[23] M. Kuczma, *Functional Equations in a Single Variable*, Polish Scientific Publishers, Warsaw, 1968.

[24] A. Meir and J. W. Moon, On the altitude of nodes in random trees, *Canad. J. Math*, **30** (1978), 997-1015.

[25] R. E. Miller, N. Pippenger, A. L. Rosenberg, and L. Snyder, Optimal 2,3-trees, *SIAM J. Comp.* **8** (1979), 42-59.

[26] J. W. Moon, *Counting Labelled Trees*, Canadian Mathematical Congress, 1970.

[27] A. M. Odlyzko, Periodic oscillations of coefficients of power series that satisfy functional equations, *Adv. Math.* **44** (1982), 180-205.

[28] A. M. Odlyzko, Enumeration of AVL-trees, manuscript in preparation.

[29] A. M. Odlyzko and L. B. Richmond, Asymptotic expansions for the coefficients of analytic generating functions, *Aequationes math.*, to appear.

[30] A. M. Odlyzko and L. B. Richmond, manuscript in preparation.

[31] R. Otter, The number of trees, *Ann. Math.* **49** (1948), 583-599.

[32] G. Pólya, Kombinatorische Anzahlbestimmungen für Graphen, Gruppen und Chemische Verbindungen, *Acta Mathematica* **68** (1937), 145-254.

[33] E. M. Reingold, A note on 3-2 trees, *Fibonacci Quart.* **17** (1979), 151-157.

[34] A. Rényi and G. Szekeres, On the height of trees, *Australian J. of Math.*, **7** (1967), 497-507.

[35] V. E. Stepanov, On the distribution of the number of vertices in strata of a random tree, *Theory Prob. Appl.*, **14** (1969), 65-78.

[36] G. Szekeres, Regular iteration of real and complex functions, *Acta Math.*, **100** (1958), 203-258.

[37] G. Szekeres, Distribution of labelled trees by diameter, pp. 392-397 in *Combinatorial Mathematics X*, Proc. 10th Australian Conference on Combinatorial Mathematics (1982), Springer Lecture Notes in Mathematics.

[38] R. A. Wright, B. Richmond, A. Odlyzko, and B. D. McKay, Constant time generation of free trees, to be published.

FIGURE CAPTIONS

Fig. 1. The function $f(\lambda)$, which equals $\lim 2^{-h} \log B_{h,n}$ as $h,n \to \infty$ with $n2^{-h} \to \lambda$, where $B_{h,n}$ is the number of binary trees of height $\leq h$ and size n.

Fig. 2. Boundary of the region of analyticity of the generating function $f(z)$ for 2,3-trees (dotted curve) and the disk of converge of the Taylor series of $f(z)$ (solid circle).

Some New Methods and Results in Tree Enumeration

A. M. Odlyzko

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

A variety of tree enumeration results have been obtained recently by new techniques. These new methods were developed to deal with situations that arise in many contexts, but especially frequently in computer science, where no nice expressions are available for the generating functions, and only recursive relations are known. A typical example is the sequence of polynomials

$$B_{h+1}(x) = 1 + xB_h(x)^2 \ , \ \ h \geq 0 \ , \ \ B_0(x) = 1 \ ,$$

where $B_h(x)$ enumerates binary trees of height $\leq h$. The new methods yield estimates for many of the interesting quantities, such as distribution of heights, even in situations as complicated as this one. A brief survey of these methods and the results they lead to is given.