Discrete Mathematics 65 (1987) 149–156
North-Holland

LEVEL NUMBER SEQUENCES FOR TREES

Philippe FLAJOLET
INRIA, Rocquencourt, 78153-Le Chesnay, France

Helmut PRODINGER
Technische Universität Wien, A-1040 Wien, Austria

We give explicit asymptotic expressions for the number of "level number sequences" (l.n.s.) associated to binary trees. The level number sequences describe the number of nodes present at each level of a tree.

1. Introduction

This paper concerns some statistical properties of a parameter related to the *profiles* of binary trees.

Define the *level* of a node v in a rooted tree t as the number of nodes on the branch connecting v to the root of t (counting both end nodes). The *level number sequence* of a tree t is the infinite sequence of integers $(n_1, n_2, \ldots)$ such that n_j is the number of nodes at level j in tree t. With our previous definitions, a level number sequence starts with a 1 and consists of eventually null integers.

Figure 1 displays a tree whose level number sequence is (1, 1, 3, 9, 25, 0, 0, . . .), each node being labelled with its depth. (We are indebted to Ms C. Cabart for providing this computer generated diagram which is built after a natural botanical growth model for coffee and tree shrubs.)

Let t be a binary tree as defined for instance in Knuth's book [7]: such trees are: (i) rooted, i.e., a certain node is distinguished as the root of the tree; (ii) binary, that is to say, each node has either 0 or 2 descendents; (iii) planar, i.e., subtrees hanging from a binary node are distinguished as left or right subtrees. (Note planarity is not essential here).

In an unpublished paper [2], Clowes, Mitrani and Wilson address the problem of determining the number of distinct level number sequences associated to all binary trees formed with n binary nodes. Without loss of generality, we shall define level number sequences for binary trees by taking here n_j to be the number of *binary nodes* at depth j in the tree. (In computer science terms, this amount to considering nullary nodes as null pointers).

Let $\boldsymbol{H}_n$ be the set of all such sequences, and let $H_n = \mathrm{card}(\boldsymbol{H}_n)$. It is clear that an element of $\boldsymbol{H}_n$ is formed from a sequence $(n_1, n_2, \ldots, n_k)$ of non-zero integers followed by an infinite sequence of trailing zeroes: $n_j = 0$ for $j > k$. The n_j with

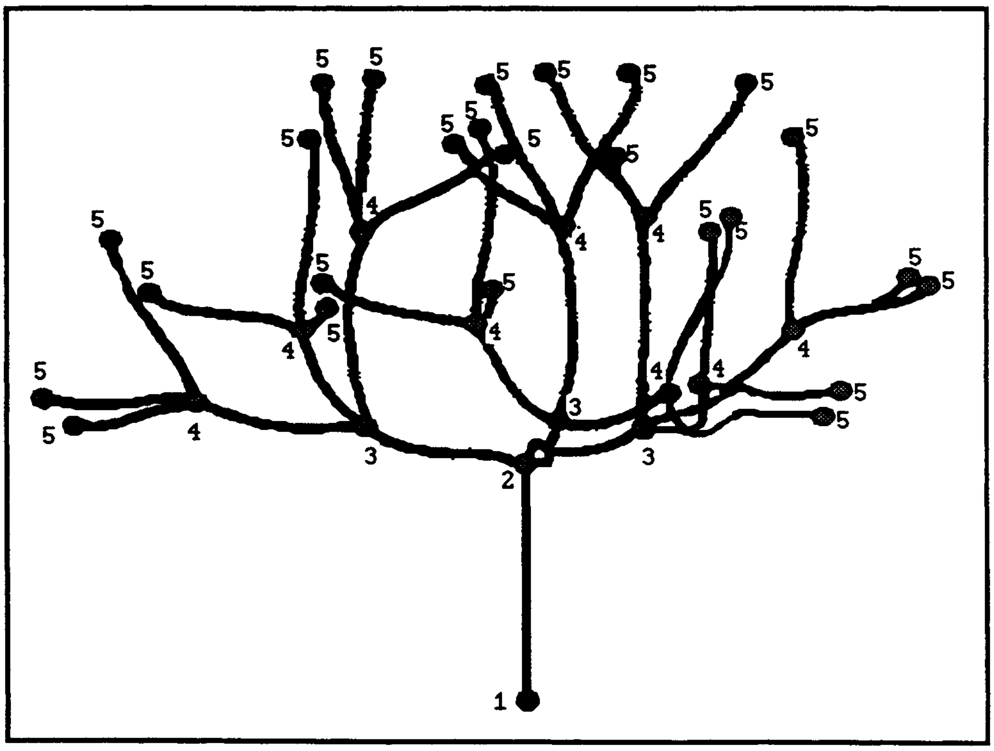

Fig. 1. A tree with level number sequence: $(1, 1, 3, 9, 25, 0, 0, \ldots)$.

$j \leqslant k$ satisfy the following characteristic conditions:

C1. $n_1 = 1$;
C2. For all j such that $1 < j \leqslant k : 1 \leqslant n_j \leqslant 2n_{j-1}$;
C3. $n_1 + n_2 + \cdots + n_k = n$.

The parameter k in the above definition is called the *height* of the level number sequence since it corresponds to the (usual) height of any associated tree.

In classical combinatorial analysis terms [1], [5], $\boldsymbol{H}_n$ is isomorphic to the set of all *compositions* of integer n such that the first summand is equal to 1 and such that each summand is at most twice the previous summand. Thus our result can be interpreted as a counting result for restricted compositions. (See [1], [5]).

The first few values of H_n for $n = 1, \ldots, 10$ are readily found to be:

$$1, 1, 2, 3, 5, 9, 16, 28, 50, 89.$$

Clowes, Mitrani and Wilson first observe the inequalities (with F_n the nth Fibonacci number):

$$F_n \leqslant H_n \leqslant 2^{n-1}. \tag{1}$$

The upper bound results from the fact that 2^{n-1} is the number of (unrestricted) compositions of n, while the lower bound counts the subset of $\boldsymbol{H}_n$ formed with summands only equal to 1 or 2.

The authors of [2] then provide a succession of refinements of this simple combinatorial argument until their best bound which is of the form:

$$C_1 1.755^n < H_n < C_2 1.802^n,$$

and from numerical evidence, they conjecture that H_n grows roughly like 1.794^n. We derive here a precise asymptotic estimate of H_n in the form of

Theorem 1. *The number of level number sequences H_n satisfies the asymptotic estimate*:

$$H_n \sim K \cdot \nu^n, \tag{2}$$

where $K = 0.254\,505\,523\,565\,319$ *and* $\nu = 1.794\,147\,187\,541\,685$ *is the inverse of the smallest positive root* ρ *of the transcendental equation*:

$$\sum_{j\geq 1} (-1)^{j+1} \frac{\rho^{2^{j+1}-2-j}}{(1-\rho)(1-\rho^3)(1-\rho^7)\cdots(1-\rho^{2^j-1})} = 1.$$

Values given by formula (2) are fairly accurate; for instance, $H_{10} = 89$ while the integer truncation, H_{10}^*, of approximation (2) is 88; for $n = 15$, corresponding values are $H_{15} = 1639$ and $H_{15}^* = 1635$. For $n = 100$, one finds:

$$H_{100} = 6\,187\,341\,363\,780\,618\,339\,584\,784, \tag{3a}$$

$$H_{100}^* = 6\,187\,341\,363\,780\,614\,360\,373\,016, \tag{3b}$$

respectively, so that there:

$$1 < \frac{H_{100}}{H_{100}^*} < 1 + 7 \times 10^{-16}.$$

Our approach in this note consists in setting up a difference equation for a series closely related to the *generating function* $H(q) \triangleq \sum_{n\geq 0} H_n q^n$. The equation is then solved and the equation involves a non-standard form of so-called q-series (See [1]). As a by-product of our analysis, we obtain

Theorem 2. *The generating function of the quantities H_n is expressible as:*

$$H(q) = \frac{a(q)}{1 - b(q)}, \tag{4}$$

where

$$a(q) = \sum_{j\geq 1} (-1)^{j+1} \frac{q^{2^{j+1}-2-j}}{(1-q)(1-q^3)(1-q^7)\cdots(1-q^{2^{j-1}-1})}, \tag{5}$$

$$b(q) = \sum_{j\geq 1} (-1)^{j+1} \frac{q^{2^{j+1}-2-j}}{(1-q)(1-q^3)(1-q^7)\cdots(1-q^{2^j-1})}. \tag{6}$$

The proof of Theorem 1 proceeds from there by noticing that $H(q)$ has a meromorphic continuation from which H_n can be recovered using Cauchy's integral formula in combination with a suitable contour of integration.

We finally mention that somewhat related results on the profiles of trees have been obtained by Meir and Moon [8] ("thickness of layers"), Flajolet and Odlyzko [4] ("height"), Odlyzko and Wilf [9] ("width"). However the statistical models are usually different. Notice for instance that, under the uniform statistics, the expected width of a binary tree of size n (this corresponds to $\max n_j$) has not been exactly determined although it is known to be of order $O(\sqrt{n})$.

2. Generating function equations

We let $\boldsymbol{H}_{n,j}^{[k]}$ denote the subset of $\boldsymbol{H}_n$ formed with level number sequences $(l.n.s)$ of height k, whose last non-zero component has value j; we also let $H_{n,j}^{[k]} = \operatorname{card} \boldsymbol{H}_{n,j}^{[k]}$. We introduce the corresponding bivariate generating functions:

$$H^{[k]}(q, u) = \sum_{n,j\geq 1} H_{n,j}^{[k]} q^n u^j, \tag{7a}$$

$$H(q, u) = \sum_{k\geq 1} H^{[k]}(q, u). \tag{7b}$$

Thus $H(q, 1)$ is the generating function of the $H_n: H(q, 1) = H(q)$.

Lemma 1. *The bivariate generating function $H(q, u)$ satisfies the functional equation:*

$$H(q, u) = qu + \frac{uq}{1 - uq}[H(q, 1) - H(q, q^2u^2)]. \tag{8}$$

Proof. From the definition, we have:

$$H^{[0]}(q, u) = qu,$$

corresponding to the unique l.n.s. $(1, 0, 0, \ldots)$ of height 1 whose last non-zero component is 1.

A recurrence relating $H^{[k]}$ to $H^{[k+1]}$ is easily obtained by the technique of "adding a new slice": consider the set $\boldsymbol{H}_{n,j}^{[k]}$; when adding a new non-zero component n_{k+1} to it, it will give rise to l.n.s. of height $k+1$ with last element equal to j^*, where j^* can be any of the integers $1, 2, \ldots, 2j$, the weight (i.e., the size of any representing tree) becoming $n + j^*$.

In terms of generating functions, this means that the process of going from $H^{[k-1]}$ to $H^{[k]}$ is achieved by the substitution

$$u^j \to uq + (uq)^2 + \cdots + (uq)^{2j} = \frac{uq}{1 - uq}(1 - (uq)^{2j}).$$

Whence the recurrence

$$H^{[k+1]}(q, u) = \frac{uq}{1-uq}[H^{[k]}(q, 1) - H^{[k]}(q, q^2u^2)],$$

and summing over all values of k yields the functional equation in the statement of the lemma. □

We can now easily finish the proof of Theorem 2. A functional equation of the form (Φ the unknown function)

$$\Phi(u) = \lambda(u) + \mu(u)\Phi(\sigma(u)) \tag{9}$$

admits, by iteration of (9), the formal solution

$$\Phi(u) = \sum_{k\geq 0}\left[\prod_{j=0}^{k-1} \mu(\sigma^{(j)}(u))\right]\lambda(\sigma^{(k)}(u)), \tag{10}$$

where $\sigma^{(k)}(u)$ denotes the kth iterate of $\sigma(u)$.

Solution (10) applies to eq. (8) with

$$\sigma(u) = q^2u^2, \qquad \lambda(u) = uq + \frac{uq}{1-uq}H(q, 1), \qquad \mu(u) = \frac{-uq}{1-uq},$$

where the iterates of $\sigma(\cdot)$ are given by $\sigma^{(k)}(u) = q^{2^{k+1}-2}u^{2^k}$. In this fashion, one obtains for $H(q, u)$ a 'solution' of the form

$$H(q, u) = A(q, u) + B(q, u)H(q, 1). \tag{11}$$

Setting $u = 1$ in (11), and solving the resulting linear equation for $H(q, 1) \equiv H(q)$, we get:

$$H(q) = \frac{A(q, 1)}{1 - B(q, 1)}. \tag{12}$$

A simple computation from the form of iterates of $\sigma(\cdot)$ and from the scheme (10) shows that $A(q, 1) \equiv a(q)$ and $B(q, 1) \equiv b(q)$ where $a(q)$ and $b(q)$ are given by (5), (6). This concludes the proof of Theorem 2. □

3. Asymptotics

In order to recover H_n from the expression of $H(q)$ provided by Theorem 2 (eq. (4)), we use Cauchy's integral formula for coefficients of analytic functions. Since, by the bound (1), $H(q)$ has a radius of convergence ρ larger than $1/\varphi = 0.61803$, we have:

$$H_n = \frac{1}{2i\pi}\int_{|q|=\frac{1}{2}} H(q)\frac{dq}{q^{n+1}}. \tag{13}$$

Classically (see e.g. [3]), the asymptotic form of H_n is easy to predict, and is simply determined by the dominant singularities of $H(q)$.

We first observe

Lemma 2. *Function $H(q)$ is meromorphic for $|q| \leq \frac{7}{10}$ with a unique simple pole at $q = \rho$, where ρ is defined in the statement of Theorem* 2.

Proof. Since $H(q)$ is the quotient of two functions analytic for $|q| < 1$, it is *meromorphic* for $|q| < 1$. Since $H(q)$ represents a series with positive coefficients, it has at least a real positive singularity on its circle of convergence. Therefore its radius of convergence is equal to the smallest root of the equation $1 - b(q) = 0$ if that root does not cancel $a(q)$. Let ρ denote the smallest positive root of $1 - b(q) = 0$. We find numerically that $\rho = 0.57367\ldots$, and at that point $a(\rho) = 0.34373\ldots$ and $b'(\rho) = 2.42320\ldots$. Thus ρ is a simple pole of $H(q)$.

We can check, by numerical analysis, that the equation $1 - b(q) = 0$ has no other zero satisfying $|q| \leq \frac{7}{10}$ (This checking could if necessary be transformed into an unpleasingly formal proof). To that purpose, we use the principle of the argument [6]:

> The number of solutions to the equation $f(q) = 0$ that lie inside a simple closed curve Γ, with $f(q)$ analytic inside and on Γ is equal to the variation of the argument of $f(q)$ along Γ, a quantity also equal to the winding number of the transformed curve $f(\Gamma)$ around the origin.

Figure 2 shows the shape of the curve $f(\Gamma)$ when Γ is the circle $|q| = \frac{7}{10}$ and $f(q) = 1 - b(q)$. Its winding number is clearly equal to 1, so that $f(q)$ has only $q = \rho$ as a zero when $|q| \leq \frac{7}{10}$. □

We can now conclude the proof of Theorem 1. To that purpose we consider the integral:

$$I_n = \frac{1}{2\mathrm{i}\pi} \int_{|q| = \frac{7}{10}} H(q) \frac{\mathrm{d}q}{q^{n+1}}. \tag{14}$$

By the residue theorem, the quantity $H_n - I_n$ which represents the integral of $H(q)/q^{n+1}$ along two concentric circles is equal to the sum of the residues of the integrand taken with a minus sign.

Since one has, when $q \to \rho$:

$$H(q) \sim -\frac{a(\rho)}{b'(\rho)} \frac{1}{q - \rho},$$

one gets:

$$H_n - I_n = \frac{a(\rho)}{b'(\rho)} \rho^{-n-1}. \tag{15}$$

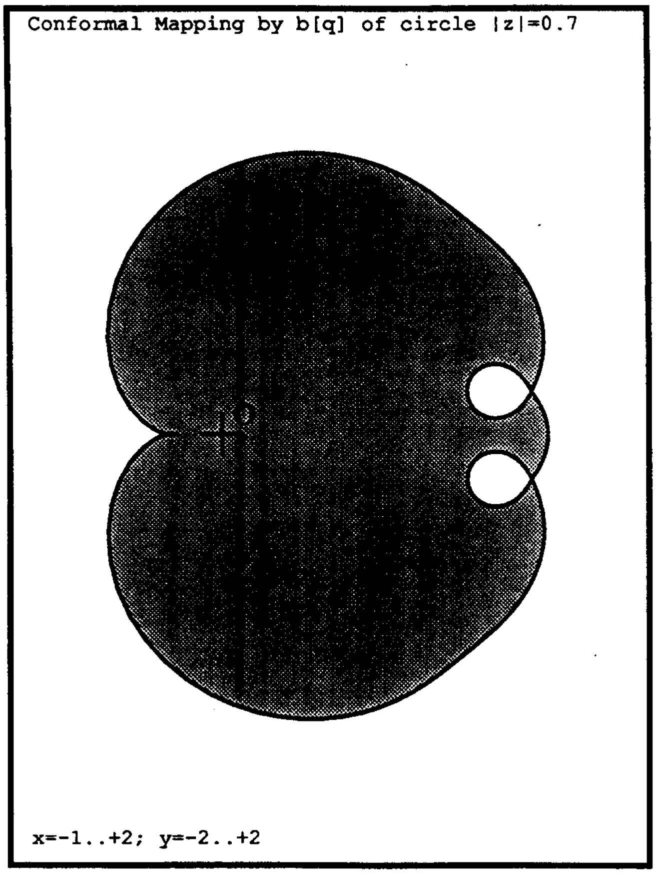

Fig. 2. The transform of the circle $\Gamma = \{q \mid |q| = \frac{7}{10}\}$ by function $1 - b(q)$. (Function $b(q)$ has been estimated from the first 50 terms of its Taylor expansion at $q = 0$; the curve is obtained by transforming 100 regularly spaced points on Γ).

We finally notice that since the integrand is analytic for $|q| = \frac{7}{10}$, quantity I_n is $O((\frac{10}{7})^n)$, whence finally for H_n an expression of the form:

$$H_n = K\nu^n + O((\tfrac{10}{7})^n)$$

where K and ν are obtained from (15), and $\nu = 1/\rho = 1.794\ldots$.

This therefore completes the proof of Theorem 1. Notice that the error term in (15) is *exponentially smaller* than the dominant asymptotic equivalent: our proof shows that it is here at most a fraction $O(0.8^n)$ of the main term. This fact common to the asymptotic behaviour of coefficients of meromorphic functions accounts for the excellent numerical accuracy of the approximation (2), as is exemplified by (3a), (3b).

Notice that if t-ary trees were considered (a node in such a tree may have out-degree 0 or t only), then one could prove similarly that the associated l.n.s. satisfy:

$$H_n(t) = K(t) \cdot \nu^n(t),$$

and we should expect $\nu(t)$ to tend to 2 as $t \to \infty$, since as t increases, a larger fraction of integer compositions become level number sequences.

Note added in proof

After this paper was written, it appeared that the sequence H_n is of interest in other branches of mathematics.

Neil Sloane pointed out that our sequence appears in his *Handbook of Integer Sequences*. It is not difficult to see that, in effect, H_n counts the number of ways of expressing 1 as a sum of n elements of the set $\{2^{-k}\}_{k\geqslant 0}$, with repetitions allowed, the order of summands not being taken into account. An example is $1 \equiv 2^{-4}+2^{-4}+2^{-3}+2^{-2}+2^{-2}+2^{-2}$; to such a partition, a canonical tree (uniquely determined by its l.n.s.) is associated by representing the way terms can be grouped together to form 1, recurrently starting from the smallest ones (on the example that l.n.s. is $(1, 2, 1, 1, 0, 0, \ldots)$).

Jean Lannes arrived at the same problem of partition counting, from algebraic topology. It turns out, rather unexpectedly, that $H(q)$ is the Poincaré series of the module on Steenrod's algebra. That module was considered by Carlsson in *Topology* 22 (1983) 83–103 and noted by him X_1, then further investigated by Lannes et al. in *Ann. Sc. Ec. Norm. Sup.* 19, 303–333 and noted there K. The series is defined as

$$P(q) = \sum_n \dim[K^n(1)] \cdot q^n$$

and it coincides with our $H(q)$. It seems further to be the case that our dominant pole ρ appears in connection with other Poincaré series in algebraic topology.

The authors are extremely grateful to N. Sloane and J. Lannes for communicating these observations.

References

[1] G. Andrews: The Theory of Partitions, Encyclopedia of Mathematics and its Applications, Vol 2 (Addison Wesley, Reading, 1976).
[2] J. Clowes, I. Mitrani and L. Wilson: Level number sequences for binary trees, University of Newcastle Computing La. Res. Rep. 112 (1977), 25p.
[3] P. Flajolet: Mathematical methods in the analysis of algorithms and data structures, INRIA Res. Rep. 400 (May 1985), 72p.; to appear in A Graduate Course on Computation Theory, Computer Science Press.
[4] P. Flajolet and A. Odlyzko: The average height of binary trees and other simple trees, J. Comput. System Sci. 25 (1982) 171–213.
[5] D. Jackson and I. Goulden: Combinatorial Enumerations (Wiley, New York, 1983).
[6] P. Henrici: Applied and Computational Complex Analysis (Wiley, New York, 1974–1977).
[7] D.E. Knuth: The Art of Computer Programming: Fundamental Algorithms (Addison Wesley, Reading, 1968).
[8] A. Meir and J. Moon: On the Altitude of Nodes in Random Trees, Canad J. Math. 30 (1978) 997–1015.
[9] A. Odlyzko and H. Wilf: Bandwidths and profiles of trees, Preprint (1984).