

On Heights of Monotonically Labelled Binary Trees

A. M. Odlyzko

Bell Laboratories
Murray Hill, New Jersey 07974

1. Introduction

A monotonically labelled binary tree is a binary tree with integer labels attached to the internal nodes in such a way that the labels along any path from the root are nondecreasing. If there is only one label, we are dealing with ordinary binary trees. Prodinger and Urbanek [6] introduced a variety of families of monotonically labelled trees and obtained enumerative results about many of them. This note is concerned largely with monotonically labelled binary trees that have at most two labels, call them 1 and 2. Several examples of such trees are shown in Fig. 1. Prodinger and Urbanek showed that the number $y_{2,n}$ of monotonically 2-labelled binary trees of size n (i.e., with n internal nodes and with $k = 2$ labels) satisfies

$$y_{2,n} \sim 4(6\pi)^{-1/2} (16/3)^n n^{-3/2} \quad \text{as } n \rightarrow \infty . \quad (1.1)$$

This result followed rather easily from the formula

$$y_2(z) = \frac{1 - (-1 + 2(1-4z)^{1/2})^{1/2}}{2z} , \quad (1.2)$$

proved in [6], where $y_2(z)$ is the generating function of the $y_{2,n}$:

$$y_2(z) = \sum_{n=0}^{\infty} y_{2,n} z^n . \quad (1.3)$$

In the case of ordinary binary trees ($k = 1$), it is well-known that the corresponding generating function $y_1(z)$ satisfies

$$y_1(z) = \frac{1 - (1-4z)^{1/2}}{2z} , \quad (1.4)$$

and

$$y_{1,n} = \frac{1}{n+1} \binom{2n}{n} \sim \pi^{-1/2} 4^n n^{-3/2} \quad \text{as } n \rightarrow \infty . \quad (1.5)$$

Kirschenhofer and Prodinger [5] later studied the average heights of monotonically 2-labelled binary trees. (For some other results about average shapes of such trees, see [4].) As is true of ordinary binary trees (*cf.* [2]), there do not appear to be any closed-form formulas like (1.2) for any generating functions of heights. However, in [2] it was shown that if $h_{k,n}$ denotes the average height of monotonically labelled binary trees of sizes n with k labels,

$$h_{k,n} = y_{k,n}^{-1} \sum_T ht(T) , \quad (1.6)$$

where the sum is over all the different monotonic k -labellings of all the binary trees of size n , then

$$h_{1,n} \sim 2(\pi n)^{1/2} \quad \text{as } n \rightarrow \infty . \quad (1.7)$$

The proof relied on a very detailed study of singularities of generating functions. (A different proof of (1.7), which also used deep analytic methods, has been given in [1].)

Kirschenhofer and Prodinger [5] extended the method of [2] and showed that

$$h_{2,n} \sim (8\pi n/3)^{1/2} \quad \text{as } n \rightarrow \infty . \quad (1.8)$$

This note shows that it is possible to obtain the result (1.8) from (1.7) by quite elementary methods, without having to resort to the analytic machinery of Kirschenhofer and Prodinger. Furthermore, the new method leads to a simple proof that

for any fixed $k \geq 1$, there is an easily computable constant c_k such that

$$h_{k,n} \sim c_k n^{1/2} \quad \text{as } n \rightarrow \infty .$$

The proof is not completely elementary in that it relies on the analytic results of [2], but it is short and easy to motivate.

2. Proof of the $k = 2$ result

The basic idea behind our proof is very simple. A typical monotonically 2-labelled binary tree ($k = 2$ labels will be assumed throughout this section) looks like the tree in Fig. 2. If m nodes have the label 2, then the $n - m$ nodes with label 1 form (after the addition of the necessary leaves) a binary tree of size $n - m$. The m nodes with label 2 then can be regarded as forming an ordered $(n - m + 1)$ -rooted forest of binary trees of total size m , with some of the trees in that forest possibly being empty. Figure 3 illustrates such a decomposition. For each pair of a binary tree of size $n - m$ and an ordered $(n - m + 1)$ -rooted binary forest, there is exactly one monotonically 2-labelled binary tree that decomposes into this pair. Since it is easy to enumerate both binary trees and ordered binary forests, this decomposition leads to a quick enumeration of monotonically 2-labelled binary trees by m , the number of nodes labelled with 2. This can then be used to give another proof of (1.2). Furthermore, this enumeration shows that the only significant contribution to the total number of monotonically 2-labelled binary trees comes from the trees for which m satisfies $m \sim n/3$. Now the height of an ordered N -rooted binary forest of size M , where $M \sim N/2$, is very small (Lemma 2), so the average height of a monotonically 2-labelled binary tree of size n is approximately that of an ordinary binary tree of size $\sim 2n/3$, which satisfies (1.8) by (1.7). The rest of

this section fills in the details of the above argument.

Lemma 1. The number $t_{r,n}$ of ordered r -rooted binary forests of size n satisfies

$$t_{r,n} = \frac{r}{2n+r} \binom{2n+r}{n}. \quad (2.1)$$

Proof. This result is old and well-known, although perhaps not in this particular form. Note that $t_{r,n}$ is the coefficient of z^n in $y_1(z)^r$, and so satisfies (2.1) by [7; p. 153], or by Lagrange inversion, or by using Cauchy's formula and shifting the contour of integration. In addition, this lemma can be deduced from [3; p. 125, Problem 2.7.3(a)].

□

Lemma 1, Eq. (1.5), and the decomposition described at the beginning of this section show that the number of monotonically 2-labelled binary trees of size n with m nodes labelled 2 is exactly

$$u_m = u_m(n) = \frac{1}{n+m+1} \binom{2n-2m}{n-m} \binom{n+m+1}{m}. \quad (2.2)$$

Hence for $m \leq n-2$,

$$\frac{u_m}{u_{m+1}} = 4 \frac{(n-m)(m+1)}{(n+m+1)(n-m-1)}, \quad (2.3)$$

so

$$\left[\frac{u_m}{u_{m+1}} - 1 \right] (n+m+1)(n-m-1) = 1+6n-4m-(n-3m)(n-m),$$

and therefore $u_m < u_{m+1}$ for $0 \leq m \leq [n/3] - 20$, and $u_m > u_{m+1}$ for $[n/3] + 20 \leq m \leq n-20$. Let $m^* = [n/3]$. Then Eq. (2.3) shows that uniformly for

$$0 \leq h \leq n^{2/3},$$

$$\begin{aligned} \frac{u_{m^*-h}}{u_{m^*}} &= \prod_{j=0}^{h-1} \frac{4(m^*-j)(n-m^*+1+j)}{(n+m^*-j)(n-m^*+j)} \\ &= \prod_{j=0}^{h-1} \left[1 - \frac{n-3m^*+3j}{n+m^*-j} \right] \left[1 + \frac{1}{n-m^*+j} \right] \quad (2.4) \\ &= \exp(-9h^2(8n)^{-1} + o(1)) \quad \text{as } n \rightarrow \infty. \end{aligned}$$

An analogous argument shows that (2.4) also holds for $-n^{2/3} \leq h \leq 0$. Therefore we conclude that the probability of a random monotonically 2-labelled binary tree having m labels equal to 2, with $0 \leq m \leq n-20$ and $|m-m^*| \geq n^{2/3}$, is $\mathbf{O}(\exp(-n^{1/3}))$. The same bound also applies to the probability of $m \geq n-20$, by a simple estimate of the expression (2.2) for u_m . Since the height of a tree of size n is $\leq n+1$, we conclude that trees with $|m-m^*| \geq n^{2/3}$ contribute $\mathbf{O}(n \exp(-n^{1/3}))$ to the average height.

We now observe that the height of a monotonically 2-labelled binary tree which decomposes into a binary tree T and an ordered binary forest F is bounded below by the height of T and above by the sum of the height of T and the height of F . Since the average height of T is by the preceding discussion asymptotic to the average height of an ordinary binary tree of size $\sim 2n/3$, it only remains to show that the average height of an $(n-m+1)$ -rooted ordered binary forest of size m is $o(n^{1/2})$ for $m \sim n/3$. That result, however, follows easily from the following lemma.

Lemma 2. The average height of an ordered r -rooted binary forest of size n , where $1/10 \leq n/r \leq 10$, is $\mathbf{O}(\log n)$.

Proof. We will actually obtain an estimate for the probability that any one of the r trees is large. Let

$$N(r, n, s) = |\{r\text{-rooted forests of size } n \text{ with some tree of size } s\}| .$$

Then $N(r, n, s) \leq U(r, n, s)$, where

$$U(r, n, s) = ry_{1,s} t_{r-1, n-s} = \frac{r}{s+1} \binom{2s}{s} \frac{r-1}{2n-2s+r-1} \binom{2n-2s+r-1}{n-s} .$$

But for n sufficiently large,

$$\begin{aligned} \frac{U(r, n, s+1)}{U(r, n, s)} &= \frac{2(n-s)(n-s+r-1)(2s+1)}{(s+2)(2n-2s+r-3)(2n-2s+r-2)} \\ &< \frac{4(n-s)(n-s+r-1)}{(2n-2s+r-3)(2n-2s+r-2)} < 1 , \end{aligned}$$

so that $U(r, n, s)$ is monotone decreasing in s . We now consider $s \leq c \log n$ for some $c > 0$. We have

$$\begin{aligned} \frac{U(r, n, s)}{t_{r,n}} &= \mathcal{O} \left[n \binom{2s}{s} \binom{2n-2s+r-1}{n-s} \binom{2n+r}{n}^{-1} \right] \\ &= \mathcal{O} \left[n 4^s \frac{n(n-1)\dots(n-s+1)}{(n+r+1)\dots(n+r+s)} \prod_{j=0}^{n-s-1} \frac{2n+r-2s-1-j}{2n+r-j} \right] \\ &= \mathcal{O} \left[n 4^s \exp(-2ns/(2n+r)) \right] \\ &= \mathcal{O} \left[n \exp(-s/10) \right] . \end{aligned}$$

If $s \geq 100 \log n$, we find that the probability that any tree in the forest has size $\geq s$ is $\mathcal{O}(n^{-8})$. Since the height of the forest is bounded above by the size of the largest tree in it, this proves the lemma.

□

3. Extensions to $k \geq 3$ labels

The result obtained above can be easily extended to estimate the average height of monotonically k -labelled binary trees for any fixed k . Any monotonically k -labelled binary tree of size n and with m labels equal to k can be decomposed uniquely into a monotonically $(k-1)$ -labelled binary tree of size $n-m$ (with internal nodes being precisely the nodes of the original tree with labels $< k$) and an ordered $(n-m+1)$ -rooted binary forest of size m . Hence

$$y_{k,n} = \sum_{m=0}^n y_{k-1,n-m} t_{n-m+1,n} . \quad (3.1)$$

Since Prodinger and Urbanek [6] have shown that

$$y_{g,n} \sim p_g q_g^{-n} r^{-3/2} \quad \text{as } n \rightarrow \infty$$

for some positive real numbers p_g and q_g , it is readily seen that the only significant contribution to the sum in (3.1) comes from the terms with

$$m \sim \frac{q_{k-1}}{1-q_{k-1}} n \quad \text{as } n \rightarrow \infty . \quad (3.2)$$

Therefore the average height of monotonically k -labelled binary trees of size n is asymptotic to the average height of monotonically $(k-1)$ -labelled binary trees of size $[nq_{k-1}/(1-q_{k-1})]$. Induction on k and (1.7) show that this average height is asymptotic to

$$c_k n^{1/2} \quad \text{as } n \rightarrow \infty ,$$

where

$$c_k^2 = 4\pi \prod_{i=1}^{k-1} \frac{q_i}{1-q_i} . \quad (3.3)$$

Since $q_1 = 1/4$ and $q_{k+1} = q_k(1-q_k)$, as is shown in [6], the c_k can be evaluated fast. It can also be shown that

$$\log c_k \sim -\frac{1}{2} k \log k \quad \text{as } k \rightarrow \infty .$$

REFERENCES

1. G. B. Brown and B. O. Schubert, On random binary trees, preprint.
2. P. Flajolet and A. M. Odlyzko, The average height of binary trees and other sample trees, *J. Computer Systems Sci.*, **25** (1982), 171-213.
3. I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, Wiley, 1983.
4. P. Kirschenhofer, On the average shape of monotonically labelled tree structures, *Discrete Appl. Math.* **7** (1984), 161-181.
5. P. Kirschenhofer and H. Prodinger, On the average height of monotonically labelled binary trees, Colloq. Math. Soc. Janos Bolyai, Proc. 1981 Eger Combinatorics Conference, to appear.
6. H. Prodinger and F. J. Urbanek, On monotone functions of tree structures, *Discrete Appl. Math.* **5** (1983), 223-239.
7. J. Riordan, *Combinatorial Identities*, Wiley, 1968.

FIGURE CAPTIONS

- Fig. 1. Half of the six monotonically labelled binary trees of size three. (The other three are mirror images of the ones shown.)
- Fig. 2. General schematic of a monotonically 2-labelled binary tree.
- Fig. 3. Decomposition of a monotonically labelled binary tree with labels 1 and 2 into a binary tree and an ordered binary forest. (\square denotes the empty binary tree.)

On Heights of Monotonically Labelled Binary Trees

A. M. Odlyzko

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

Monotonically k -labelled binary trees are binary trees in which the internal nodes are labelled with integers from $\{1, 2, \dots, k\}$ for some $k \geq 2$, and the labels along any path starting at the root are non-decreasing. A simple proof is given of a result of Kirschenhofer and Prodinger, namely that the average height of monotonically 2-labelled binary trees of size (number of internal nodes) n is

$$\sim \sqrt{\frac{8\pi n}{3}} \quad \text{as } n \rightarrow \infty .$$

It is also shown how to extend this result to the case of $k \geq 3$ labels.