Europ. J. Combinatorics (1984) 5, 351–357

The Average Height of the Second Highest Leaf of a Planted Plane Tree

HELMUT PRODINGER

The 2-height of a planted plane tree is the distance between the root and the second highest leaf (which can be the same as the height). In this note the following theorem is proved:

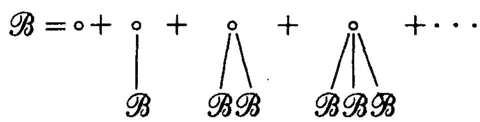
The average 2-height of a planted plane tree with n nodes, considering all such trees to be equally likely, is

$$(\pi n)^{1/2}-\tfrac{7}{6}+O(n^{-1/2+\varepsilon}), \qquad \text{for } \varepsilon>0 \text{ and } n\to\infty.$$

1. INTRODUCTION

Three pioneering papers [8], [2] and [4] exist on the subject of the *height of trees.* In [8] Rényi and Szekeres have studied *labelled trees*; in [4] Flajolet and Odlyzko have studied *binary trees* using a function theoretic approach, and in [2] de Bruijn, Knuth and Rice have studied *planted plane trees* using the so-called *Gamma-function-method*:

The family $\mathcal{B}$ of planted plane trees can be described by the formal equation



(1)

yielding the equation

$$B(z)=\frac{z}{1-B(z)} \tag{2}$$

for the generating function $B(z)=\sum_{n\geq 1} t(n)z^n$ of the numbers $t(n)$ of planted plane trees. This gives

$$B(z)=\frac{1-(1-4z)^{1/2}}{2} \quad \text{and} \quad t(n)=\frac{1}{n}\binom{2n-2}{n-1}, \qquad n\geq 1. \tag{3}$$

If the height of a tree denotes the number of nodes on a maximal simple path starting at the root, the essential result of [2] reads:

The average height of a planted plane tree with n nodes, considering all such trees to be equally likely, is

$$(\pi n)^{1/2}-\tfrac{1}{2}+O(n^{-1/2+\varepsilon}) \quad \text{for } \varepsilon>0 \text{ and } n\to\infty. \tag{4}$$

The present paper gives a related result to (4). The height measures the distance between the root and the highest leaf. We consider as notion of height (called *2-height*) the distance between the root and the *second highest leaf* (the second highest leaf may be as high as the highest leaf) and prove:

The average 2-height of a planted plane tree with n nodes, considering all such trees to be equally likely, is

$$(\pi n)^{1/2}-\tfrac{7}{6}+O(n^{-1/2+\varepsilon}) \quad \text{for } \varepsilon>0 \text{ and } n\to\infty. \tag{5}$$

We remark that is convenient to define the 2-height of a tree to be zero if it has just one leaf.

For the coefficient of z^n in f we write $[z^n]f$.

2. The Average 2-Height

Let $B_h(z)$ be the generating function of trees with n nodes and 2-height $\leqslant h$, $A_h(z)$ the generating function of trees with n nodes and height $\leqslant h$ and $C_h(z) = B_h(z) - A_h(z)$.

LEMMA 1. *With* $z = u/(1+u)^2$,

$$C_{h-1}(z) = \frac{(1-u)^3}{1-u^3} \cdot \frac{u^h}{(1+u^h)^2}. \tag{6}$$

PROOF. Regarding a tree counted by $C_h(z)$ we obtain a recursion by considering the subtree counted by $C_{h-1}(z)$:

$$\begin{aligned} C_h &= z \sum_{i\geqslant 0} A_{h-1}^i \cdot C_{h-1} \cdot \sum_{j\geqslant 0} A_{h-1}^j \\ &= z\frac{1}{(1-A_{h-1})^2} \cdot C_{h-1}. \end{aligned} \tag{7}$$

Since $C_0 = z/(1-z)$ and by [2] $A_h = z/(1-A_{h-1})$ an iteration of (7) yields

$$C_h = z^{-h} \cdot \frac{z}{1-z} \cdot \prod_{i=1}^{h} A_i^2. \tag{8}$$

From [2] we know that

$$A_h(z) = z \cdot \frac{p_h(z)}{p_{h+1}(z)} \quad \text{with} \quad p_i = \frac{(1+u)^{1-i}}{1-u}(1-u^i). \tag{9}$$

Hence

$$C_h = z^{-h} \cdot \frac{z}{1-z} \cdot z^{2h} \prod_{i=1}^{h} \frac{p_i^2}{p_{i+1}^2} = \frac{z^{h+1}}{1-z} \cdot \frac{p_1^2}{p_{h+1}^2}. \tag{10}$$

Since

$$\frac{1}{1-z} = \frac{(1+u)^2}{1+u+u^2} = \frac{(1-u)(1+u)^2}{1-u^3}, \tag{11}$$

the result follows by an elementary computation.

For the average 2-height we have to compute

$$t^{-1}(n) \cdot \sum_{h\geqslant 0} [z^n](B(z) - A_h(z) - C_h(z)). \tag{12}$$

Since

$$t^{-1}(n) \cdot \sum_{h\geqslant 0} [z^n](B(z) - A_h(z)) \tag{13}$$

is the average height, we have to consider

$$\sum_{h\geqslant 0} [z^n] C_h(z). \tag{14}$$

For a given sequence of numbers a_k we write

$$\Delta^3 a_k = a_k - 3a_{k-1} + 3a_{k-2} - a_{k-3} \quad \text{and} \quad \Delta^4 a_k = a_k - 4a_{k-1} + 6a_{k-2} - 4a_{k-3} + a_{k-4}.$$

LEMMA 2.

$$[z^n]C_{h-1}(z)=\sum_{\lambda\geq 1}\lambda\cdot\Delta^4\sum_{s\geq 0}\binom{2n-1}{n-h\lambda-3s} \tag{15}$$

where the difference operator works on $n-h\lambda-3s$.

PROOF.

$$\begin{aligned}[z^n]C_{h-1}(z)&=\frac{1}{2\pi i}\int^{(0_+)}\frac{dz}{z^{n+1}}C_{h-1}(z)\\ &=\frac{1}{2\pi i}\int^{(0_+)}\frac{du}{u^{n+1}}(1+u)^{2n-1}\cdot\frac{(1-u)^4}{1-u^3}\cdot\frac{u^h}{(1-u^h)^2}\\ &=[u^n](1+u)^{2n-1}\frac{(1-u)^4}{1-u^3}\cdot\frac{u^h}{(1-u^h)^2}.\end{aligned} \tag{16}$$

The result follows by expanding the denominator.

Thus we have to consider

$$\xi=\sum_{h,\lambda\geq 1}\lambda\cdot\Delta^4\sum_{s\geq 0}\binom{2n-1}{n-h\lambda-3s}. \tag{17}$$

LEMMA 3. *If $k=O(n^{+1/2+\varepsilon})$,*

$$\frac{\Delta^4\binom{2n-1}{n-k}}{\binom{2n}{n}}=\frac{2}{n^2}e^{-k^2/n}\left[3-\frac{12k^2}{n}+\frac{4k^4}{n^2}+O(n^{-1/2+\varepsilon})\right]. \tag{18}$$

Outside the range $k=O(n^{+1/2+\varepsilon})$, the left hand side of (18) *is exponentially small.*

PROOF. First, an elementary computation gives

$$\Delta^4\binom{2n-1}{n-k}=\frac{1}{2}\left(1+\frac{k}{n}\right)\Delta^4\binom{2n}{n-k}-\frac{2}{n}\Delta^3\binom{2n}{n-k-1}. \tag{19}$$

Now we use the following formula [5] (a fixed)

$$\begin{aligned}\frac{\binom{2n}{n-k-a}}{\binom{2n}{n}}=e^{-k^2/n}\Bigg[&1+\frac{a^2+a^4}{2n^2}-\frac{a^2}{n}+\left(\frac{-2a}{n}+\frac{2a^3+a}{n^2}\right)k\\ &+\left(\frac{4a^2+1}{2n^2}-\frac{12a^4+21a^2+1}{6n^3}\right)k^2-\frac{4a^3+5a}{3n^3}k^3\\ &+\left(\frac{16a^4+60a^2+9}{24n^4}-\frac{1}{6n^3}\right)k^4+\frac{a}{3n^4}k^5\\ &-\frac{20a^2+9}{60n^5}k^6+\frac{1}{72n^6}k^8+O(n^{-5/2+\varepsilon})\Bigg].\end{aligned} \tag{20}$$

This gives after laborious computations

$$\binom{2n}{n}^{-1}\Delta^4\binom{2n}{n-k}=\frac{1}{n^2}\,\mathrm{e}^{-k^2/n}\left[12-\frac{48k^2}{n}+\frac{16k^4}{n^2}+\mathrm{O}(n^{-1/2+\varepsilon})\right], \tag{21}$$

and

$$\binom{2n}{n}^{-1}\frac{k}{n}\Delta^4\binom{2n}{n-k} \quad \text{and} \quad \binom{2n}{n}^{-1}\frac{1}{n}\Delta^3\binom{2n}{n-k}$$

contribute only

$$\frac{1}{n^2}\,e^{-k^2n}\mathrm{O}(n^{-1/2+\varepsilon}).$$

An easy consequence of Lemma 3 is

LEMMA 4. *Let* $k=\mathrm{O}(n^{+1/2+\varepsilon})$.

$$\binom{2n}{n}^{-1}\sum_{s\geq 0}\Delta^4\binom{2n-1}{n-k-3s}=\frac{2}{n^2}\sum_{s\geq 0}\mathrm{e}^{-(k+3s)^2/n}\left[3-\frac{12(k+3s)^2}{n}+\frac{4(k+3s)^4}{n^2}\right]$$
$$\times(1+\mathrm{O}(n^{-1/2+\varepsilon})). \tag{22}$$

Let $\operatorname{erfc}(z)=2\pi^{-1/2}\int_z^\infty \mathrm{e}^{-t^2}\,\mathrm{d}t$ be the complement of the error function. Using the ideas of Feller [3] (compare [7]), one can prove

LEMMA 5. *Let* $k=\mathrm{O}(n^{+1/2+\varepsilon})$.

$$\sum_{s\geq 0}(k+3s)^b\,\mathrm{e}^{-(k+3s)^2/n}=\int_0^\infty (k+3s)^b\,\mathrm{e}^{-(k+3s)^2/n}\,\mathrm{d}s$$
$$\times(1+\mathrm{O}(n^{-1/2+\varepsilon})). \tag{23}$$

LEMMA 6.

$$\text{(a)}\quad \frac{2}{\pi^{1/2}}\int_z^\infty t^2\,\mathrm{e}^{-t^2}\,\mathrm{d}t=\frac{1}{2}\operatorname{erfc}(z)+\frac{z}{\pi^{1/2}}\,\mathrm{e}^{-z^2} \tag{24}$$

$$\text{(b)}\quad \frac{2}{\pi^{1/2}}\int_z^\infty t^4\,\mathrm{e}^{-t^2}\,\mathrm{d}t=\frac{3}{4}\operatorname{erfc}(z)+\frac{\frac{3}{2}z+z^3}{\pi^{1/2}}\,\mathrm{e}^{-z^2}. \tag{25}$$

PROOF. The functions

$$\operatorname{erfc}_n(z)=\frac{2}{\pi^{1/2}}\int_z^\infty\frac{(t-z)^n}{n!}\,\mathrm{e}^{-t^2}\,\mathrm{d}t \tag{26}$$

fulfill $\operatorname{erfc}_0(z)=\operatorname{erfc}(z)$, $\operatorname{erfc}_{-1}(z)=2\pi^{-1/2}\,\mathrm{e}^{-z^2}$ and the recursion

$$\operatorname{erfc}_n(z)=-\frac{z}{n}\operatorname{erfc}_{n-1}(z)+\frac{1}{2n}\operatorname{erfc}_{n-2}(z). \tag{27}$$

(Compare [1; p. 299].) From this the result follows after a long but elementary computation.

LEMMA 7.

$$\binom{2n}{n}^{-1} \sum_{s\geq 0} \Delta^4 \binom{2n-1}{n-k-3s} = \frac{2}{n^2} e^{-k^2/n} \left[-k + \frac{2}{3} \cdot \frac{k^3}{n} \right] (1 + O(n^{-1/2+\varepsilon})). \tag{28}$$

PROOF. Since for $k = O(n^{-1/2+\varepsilon})$

$$\sum_{s\geq 0} \frac{(k+3s)^{2b}}{n^b} e^{-(k+3s)^2/n} = \int_0^\infty \frac{(k+3s)^{2b}}{n^b} e^{-(k+3s)^2/n} \, ds (1 + O(n^{-1/2+\varepsilon}))$$

$$= \frac{n^{1/2}}{3} \int_{k/n^{1/2}}^\infty t^{2b} e^{-t^2/n} \, dt (1 + O(n^{-1/2+\varepsilon})) \tag{29}$$

the result follows by applying Lemma 6 to Lemma 4.

LEMMA 8.

$$\xi = \left(-\frac{2}{n^2} g_1(n) + \frac{4}{3n^3} g_3(n) \right) (1 + O(n^{-1/2+\varepsilon})) \tag{30}$$

where

$$g_b(n) = \sum_{k\geq 1} k^b \sigma(k) e^{-k^2/n}, \qquad b \geq 0 \quad \text{and} \quad \sigma(k) = \sum_{j|k} j. \tag{31}$$

PROOF. This is a simple rearrangement of (17) using the approximation (28).

LEMMA 9. *Let* $b \in \mathbb{N}_0$.

$$g_b(n) = \frac{\pi^2}{12} \Gamma\left(\frac{b+2}{2}\right) n^{(b+2)/2} - \frac{1}{4} \Gamma\left(\frac{b+1}{2}\right) n^{(b+1)/2} + O(1), \qquad n \to \infty. \tag{32}$$

PROOF. We use the Gamma-function method: Since

$$e^{-x} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(z) x^{-z} \, dz, \qquad x > 0, \quad c > 0 \tag{33}$$

and

$$\sum_{k\geq 1} \sigma(k) k^{-z} = \zeta(z) \zeta(z-1), \tag{34}$$

we have

$$g_b(n) = \sum_{k\geq 1} k^b \sigma(k) \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(z) \left(\frac{k^2}{n}\right)^{-z} dz$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(z) n^z \zeta(2z-b) \zeta(2z-b-1) \, dz. \tag{35}$$

By a well known method we can shift the line of integration to the left as far as we please if we only take the residues into account. From the simple pole of the ζ-function we have

$$\Gamma\left(\frac{b+2}{2}\right)n^{(b+2)/2}\zeta(2)\tfrac{1}{2} \tag{36}$$

and

$$\Gamma\left(\frac{b+1}{2}\right)n^{(b+1)/2}\tfrac{1}{2}\zeta(0). \tag{37}$$

Now $\zeta(2)=\pi^2/6$ and $\zeta(0)=-1/2$ [9], yielding the leading terms of the expansion. The simple poles of the Γ-function contribute $O(1)$.

Plugging this asymptotic expansion of $g_b(n)$ into equaton (30), we find

LEMMA 10.

$$\xi=\frac{1}{6n}(1+O(n^{-1/2+\varepsilon})), \qquad n\to\infty. \tag{38}$$

THEOREM 11.

$$\tau_n:=t^{-1}(n)\sum_{h\geqslant 0}[z^n]C_h(z)=\tfrac{2}{3}+O(n^{-1/2+\varepsilon}), \qquad n\to\infty. \tag{39}$$

PROOF. Since

$$t(n)=\frac{1}{n}\binom{2n-2}{n-1}=\frac{1}{4n}\binom{2n}{n}\left(1+O\left(\frac{1}{n}\right)\right),$$

we have to multiply ξ by $4n(1+O(1/n))$ to obtain the result.

Altogether we have proved:

THEOREM 12. *The average 2-height of a planted plane tree with n nodes is*

$$(\pi n)^{1/2}-\tfrac{7}{6}+O(n^{-1/2+\varepsilon}) \qquad \text{for } \varepsilon>0 \text{ and } n\to\infty. \tag{40}$$

Table 1 shows some values of τ_n of Theorem 11.

TABLE 1

n	τ_n	n	τ_n
1	1·0000	9	0·7664
2	2·0000	10	0·7532
3	1·5000	11	0·7430
4	1·2000	12	0·7348
5	1·0000	13	0·7282
6	0·8810	14	0·7228
7	0·8182	15	0·7184
8	0·7855	16	0·7146

References

1. M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*, Dover, New York, 1964.
2. N. G. de Bruijn, D. E. Knuth, and S. O. Rice. The average height of planted plane trees, in: *Graph Theory and Computing* (R. C. Read, Ed.), Academic Press, New York, 1972, pp. 15–22.
3. W. Feller, *An Introduction to Probability and its Applications*, Vol. 1, 2nd ed., John Wiley, New York, 1957.
4. Ph. Flajolet and A. Odlyzko. The average height of binary trees and other simple trees, *J. Comput. System Sci.* **25** (1982), 172–213.
5. R. Kemp. The average height of r-tuply rooted planted plane trees. *Computing* **25** (1980), 209–232.
6. D. E. Knuth. *The Art of Computer Programming*, Vol. 1, 2nd ed. Addison-Wesley, Reading, MA, 1973.
7. W. Panny and H. Prodinger. The expected height of paths for several notions of height, *Stud. Sci. Math. Hung.*, to appear.
8. A. Rényi and G. Szekeres. On the height of trees, *Austral. J. Math.* **7** (1967), 497–507.
9. E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*, 4th ed., Cambridge University Press, Cambridge, 1927.

Received 12 January 1981

H. Prodinger
Institut für Algebra und Diskrete Mathematik,
Technische Universität Wien,
A-1040 Wien, Gusshausstrasse 27–29, Austria