THE HEIGHT OF PLANTED PLANE TREES REVISITED

HELMUT PRODINGER

Technical University Vienna, Institute for Algebra and Discrete Mathematics, A-1040 Vienna, Gußhausstraße 27-29, Austria

The family of planted plane trees (trees for short) $\mathcal{B}$ is defined by the formal equation.

$$\mathcal{B} = \cdot + \underset{\mathcal{B}}{\overset{\cdot}{|}} + \underset{\mathcal{B}\;\;\mathcal{B}}{\overset{\cdot}{/\backslash}} + \ldots$$

from which it follows that the generating function B(z) of trees satisfies $B=z/(1-B)$ and is given by

$$B(z) = \frac{1-\sqrt{1-4z}}{2} \quad ; \quad [z^n]\, B = -\frac{1}{2}\,(-4)^n \binom{1/2}{n} = \frac{1}{n}\binom{2n-2}{n-1} .$$

$[z^n]f$ denotes, as usual, the coefficient of z^n in the series f.

The average height $\overline{h_n}$ (i.e. maximal number of nodes in a chain connecting the root and a leaf) of trees of size n, where all trees with n nodes are assumed to be equally likely, satisfies

$$\overline{h_n} = \sqrt{\pi n} - \frac{1}{2} + O(n^{-1/2}) \quad .$$

This has been derived in the pioneering paper [1].

We give here an alternative derivation which has the advantage that the coefficients in the asymptotic series for $\overline{h_n}$ can be computed more easily. To illustrate this, we compute two further terms (one appears already in [5],[6]). The method is based on a complex variable approach which is explained in more detail in a forthcoming paper of Flajolet and the author about register allocation problems [3]. Here, we just give the computational steps; a rigorous derivation can be made "à la Odlyzko" ([2],[7]) or by a Darboux-type argument where a generating function f(z) has a singularity

ρ on its circle of convergence and behaves like

$$f(z) = f_1(z).\log(1-\frac{z}{\rho}) + f_2(z).(1-\frac{z}{\rho})^{1/2} + f_3(z)$$

with f_1, f_2, f_3 analytic in a larger area [4].

Such a local expansion can then be translated into an expansion of the Taylor coefficients of f(z).

$\overline{h_n}$ is given by

$$\overline{h_n} = \frac{[z^n]\ E(z)}{[z^n]\ B(z)} ,$$

where E(z) is the generating function of the sum of heights of all trees with n nodes. Using the substitution [1]

$$z = \frac{u}{(1+u)^2} \quad \longleftrightarrow \quad u = \frac{1-r}{1+r} \text{ with } r = \sqrt{1-4z},$$

the singularity of z=1/4 turns into u=1. So we want to know about a local expansion of E(z) about u=1. It is convenient to set $u=e^{-t}$, where t tends to 0. It is known [1] that

$$E(z) = \frac{1-u}{1+u} \cdot \sum_{k\geq 1} d(k)\ u^k ,$$

where d(k) is the number of divisors of k. Now

$$\frac{1-u}{1+u} = \frac{1-e^{-t}}{1+e^{-t}} = \frac{t - t^2/2 + t^3/6 + \dots}{2 - t + t^2/2 + \dots}$$

$$= \frac{t}{2} - \frac{t^3}{24} + \dots ,$$

so that we can turn to the other factor of E(z), which we call

$$V(t) = \sum_{k\geq 1} d(k)\ e^{-tk} .$$

The Mellin transform (see [4]) of V(t) is readily derived to be

$$V^*(s) = \zeta^2(s)\ \Gamma(s) ;$$

the Mellin inversion formula tells us

$$V(t) = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \zeta^2(s)\,\Gamma(s)\,t^{-s}\,ds\,.$$

By shifting the line of integration to the left as far as we please and taking the residues into account, we obtain an asymptotic series for $V(t)$:

There is a double pole at $s=1$; the residue is

$$\frac{\gamma}{t} - \frac{\log t}{t}\,.$$

There are simple poles at $s=-k$, $k \in \mathbb{N}_0$; the residues are

$$\frac{B_{k+1}^2}{(k+1)^2}\cdot\frac{(-1)^k}{k!}\,t^k\,,$$

B_i denoting Bernoulli numbers. Thus

$$V(t) \sim \frac{\gamma}{t} - \frac{\log t}{t} + \frac{1}{4} - \sum_{j\geq 1}\frac{B_{2j}^2}{(2j)^2}\cdot\frac{t^{2j-1}}{(2j-1)!}\,.$$

Hence

$$E(z) \sim \left[\frac{t}{2} - \frac{t^3}{24} + \ldots\right]\left[\frac{\gamma}{t} - \frac{\log t}{t} + \frac{1}{4} - \frac{t}{144} + \ldots\right]$$

$$= \frac{\gamma}{2} - \frac{\log t}{2} + \frac{t}{8} - \frac{t^2}{288} - \frac{t^2\gamma}{24} + \frac{t^2\log t}{24} - \frac{t^3}{96} + \ldots$$

Now

$$t = -\log\frac{1-r}{1+r} = 2r + 2\frac{r^3}{3} + \ldots\,,$$

yielding

$$E(z) \sim \frac{\gamma}{2} - \frac{1}{2}\log 2r - \frac{1}{2}\frac{r^2}{3} + \frac{r}{4} + \frac{r^3}{12} - \frac{r^2}{72} - \frac{r^2}{24}$$

$$+ \frac{r^2\log 2r}{6} - \frac{r^3}{12} + \ldots$$

$$= K_1 - \frac{1}{4}\log(1-4z) + \frac{1}{4}\sqrt{1-4z} + K_2.(1-4z)$$

$$+ \frac{1}{12}(1-4z)\log(1-4z) + 0.(1-4z)^{3/2} + \ldots$$

To get $[z^n]E(z)$ we have to find $[z^n]\log(1-4z)$, $[z^n]\sqrt{1-4z}$ and so on:

$$[z^n]E(z) \sim \frac{1}{4}\cdot\frac{4^n}{n} + \frac{1}{4}\cdot(-4)^n\binom{1/2}{n} + \frac{1}{12}\cdot\frac{4^n}{n(n-1)} + 0 + \ldots$$

Now

$$\binom{1/2}{n} = \frac{-1}{2n-1}\,\frac{(-1)^n}{4^n}\binom{2n}{n}$$

and

$$\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}\cdot\left(1 - \frac{1}{8n} + \ldots\right) ,$$

thus

$$\binom{1/2}{n} \sim \frac{(-1)^{n+1}}{2n}\left(1 + \frac{1}{2n} + \ldots\right)\frac{1}{\sqrt{\pi n}}\left(1 - \frac{1}{8n} + \ldots\right)$$

$$= \frac{(-1)^{n+1}}{2\sqrt{\pi}n^{3/2}}\left(1 + \frac{3}{8n} + \ldots\right) .$$

This gives

$$\overline{h_n} = \frac{[z^n]E(z)}{[z^n]B(z)} = \frac{[z^n]E(z)}{-\frac{1}{2}(-4)^n\binom{1/2}{n}}$$

$$= \frac{1}{4}\,\frac{4^n}{n}\,\frac{2}{4^n}\,2\sqrt{\pi}n^{3/2}\left(1-\frac{3}{8n}+\ldots\right) - \frac{1}{2}$$

$$+ \frac{1}{12}\,\frac{4^n}{n^2}(1+\ldots)\frac{2}{4^n}2\sqrt{\pi}n^{3/2}(1+\ldots) + \frac{0}{n} + \ldots$$

$$= \sqrt{\pi n} - \frac{3}{8}\sqrt{\frac{\pi}{n}} - \frac{1}{2} + \frac{1}{3}\sqrt{\frac{\pi}{n}} + \frac{0}{n} + \ldots$$

$$= \sqrt{\pi n} - \frac{1}{2} - \frac{1}{24}\sqrt{\frac{\pi}{n}} + \frac{0}{n} + \ldots$$

Remark that the nature of the expansion of $E(z)$ tells us, that the asymptotic series for $\overline{h_n}$ is in powers of $1/\sqrt{n}$.

REFERENCES

[1] N.G.DE BRUIJN, D.E.KNUTH, S.O. RICE : The average height of planted plane trees, in: Graph Theory and Computing (R.C.Read, Ed.), 15-22, New York-London: Academic Press, 1972.

[2] P.FLAJOLET, A.ODLYZKO : The average height of binary trees and other simple trees, J. of Comput. Syst. Sci. 25(1982),171-213.

[3] P.FLAJOLET, H.PRODINGER : The average number of registers to evaluate unary-binary trees, in preparation, 1983.

[4] P.HENRICI : Applied and Computational Complex Analysis, Vol. 2, John Wiley, New York, 1972.

[5] R.KEMP : On the average stack size of regularly distributed binary trees, in Proc. of the 6th ICALP (Graz, Austria, 1979), Lecture Notes in Comp. Sci., Springer 1979.

[6] D.E.KNUTH : The Art of Computer Programming, Vol.1, 2nd ed., Reading, Mass., Addison-Wesley, 1973.

[7] A.ODLYZKO : Periodic oscillations of coefficients of power series that satisfy functional equations, Adv. in Math. 44(1982),180-205.

$$\sim \frac{1}{3}t^{-2} - \frac{\zeta(-1)}{\log 2}\log t \qquad (a=2)$$

$$\sim \frac{1}{3}t^{-2} - \frac{1}{2}t^{-1} + \frac{\zeta(-1)}{2\log 2}\log t \qquad (a=3).$$

Hence

$$N(z) \sim \tfrac{3}{4}\zeta(-1)\log_2 t \sim -\tfrac{1}{16}\log_2 \varepsilon.$$

Theorem 5. *The average number of nodes "above" the node which first equals the register value is asymptotic to*

$$\frac{\sqrt{\pi}}{32\log 2}\sqrt{n}, \qquad n\to\infty. \quad \approx 0.079909699\sqrt{n}$$

7. Epilogue

From the explicit formulae for the generating functions mentioned in the introduction we find for instance

$$S_{p+1}(z) = (B(z)-1)^{2^p} R_p(z).$$

We will sketch a simple combinatorial argument for that: Take a tree with register function p and consider its unique largest subtree in the forest of critical nodes. This tree is a complete binary tree with 2^p leaves. If we replace each leaf by an arbitrary nonempty tree (counted by $B(z)-1$), we thereby obtain a tree with strictly larger register function. This mapping is injective; it is surjective as well. So we have a bijection and therefore the announced formula. (The inverse mapping could be described in a clumsy way by cutting down a tree with register function $>p$ in a certain sense of maximality, yielding the 2^p nonempty trees and a tree with register function p.)

We will now prove the explicit formulae starting from our just established equality. This is therefore a second easy derivation of the explicit formulae. (The first one is due to P. Kirschenhofer and H. Prodinger, see [12]). We have

$$S_{p+1} = u^{2^p} R_p \quad \text{and} \quad S_p = u^{2^{p-1}} R_{p-1};$$

taking differences we see:

$$R_p = u^{2^{p-1}} R_{p-1} - u^{2^p} R_p$$

$$= R_{p-1} \frac{u^{2^p - 1}}{1+u^{2^p}}$$

$$= \frac{u^{1+2+4+\ldots+2^p-1}}{(1+u^2)(1+u^4)\ldots(1+u^{2^p})}$$

$$= \frac{u^{2^p-1}}{(1-u^{2^{p+1}})/(1-u^2)} = \frac{1-u^2}{u} \frac{u^{2^p}}{1-u^{2^{p+1}}}.$$

References

[1] P. Flajolet, Analyse d'algorithmes de manipulation d'arbres et de fichiers, Cahiers de BURO, 34-35 (1981), 1–209.

[2] P. Flajolet, J. -C. Raoult, J. Vuillemin, On the average number of registers required for evaluating arithmetic expressions, FOCS-Conference 1977, 196–205.

[3] P. Flajolet, J. -C. Raoult, J. Vuillemin, The number of registers required for evaluating arithmetic expressions, Theoretical Computer Science 9 (1979), 99–125.

[4] P. Flajolet, A. Odlyzko, The average height of binary trees and other simple trees, J. Comput. Syst. Sci. 25 (1982), 142–158.

[5] P. Flajolet, H. Prodinger, Register allocation for unary-binary trees, SIAM J. on Computing, 15 (1986), 629–640.

[6] P. Flajolet, M. Régnier, R. Sedgewick, Mellin transform techniques for the analysis of algorithms, in preparation.

[7] J. Francon, Sur le nombre de registres nécessaires à l'evaluation d'une expression arithmétique, RAIRO Inf. Theor. 18 (1984), 355–364.

[8] R. Kemp, The average number of registers to evaluate a binary tree optimally, Acta Informatica 11 (1979), 363–372.

[9] A. Meir, J. W. Moon, J. R. Pounder, On the order of random channel networks, SIAM J. Alg. Discr. Meth. 1 (1980), 25–33.

[10] J. W. Moon, On Horton's law for random channel networks, Annals of Discrete Mathematics 8 (1980), 117–121.

[11] H. Prodinger, *Abzählprobleme bei Bäumen*, Publ. de l'IRMA (Straßburg, G. Baron, P. Kirschenhofer, Ed.), (1983), 167–173.

[12] H. Prodinger, Die Bestimmung gewisser Parameter bei binären Bäumen mit Hilfe analytischer Methoden, Lecture Notes in Mathematics 1114 (1985), 118–133.

[13] H. Prodinger, *Quelques techniques analytiques pour l'etude asymptotique des parametres dans les arbres*, Proceedings Ecole de Printemps Ile de Ré, P. Flajolet, Ed.

[14] E. T. Whittaker, G. N. Watson, *A course of modern analysis*, Cambridge University Press, 1927.