TOWARDS A MORE PRECISE ANALYSIS OF AN ALGORITHM TO GENERATE BINARY TREES : A TUTORIAL

ALOIS PANHOLZER AND HELMUT PRODINGER

ABSTRACT. For the analysis of an algorithm to generate binary trees [6], the behaviour of a certain sequence of numbers is essential. In the original paper, it was expressed by a recursion.

Here, we show how to solve this (and similar) recursions, both, explicitly and asymptotically. Some additional information about useful mathematical software is also provided.

1. INTRODUCTION

In the analysis of an algorithm to generate binary trees the authors of [6] came up with a double sequence $g_{n,k}$ and gave a recursion for it (Theorem 9, with notation changed). The sequence $g_{n,0}$ was of particular interest.

Since there was no analysis provided we think that it is appropriate to explain here how such recursions can be treated properly. Especially we want to discuss generating functions, explicit forms, asymptotic expansions, and convenient Computer software.

It is a little bit nicer to explain things for a slightly modified recursion that seems to be more natural. However, the original one can be treated in the same way, as we will also demonstrate.

Here is the recursion:

$$g_{n,k} = g_{n-1,k-1} + 2g_{n-1,k} + g_{n-1,k+1} + 1 \qquad \text{for } 0 \le k \le n \tag{1}$$

with the assumption that all values $g_{n,k}$ outside the range $0 \le k \le n$ are zero.

2. GENERATING FUNCTIONS

As in Knuth [3, page 537] we set up the generating function

$$G(x,z) = \sum_{0 \le k \le n} g_{n,k} x^k z^n.$$

Summing up the recursion (multiplied by $x^k z^n$) we find

$$G(x,z) = zxG(x,z) + 2zG(x,z) + \frac{z}{x}\big(G(x,z) - g(z)\big) + \sum_{0 \le k \le n} x^k z^n,$$

with $g(z) = G(0,z)$. Note that

$$\sum_{0 \le k \le n} x^k z^n = \sum_{k \ge 0} \sum_{j \ge 0} x^k z^{k+j} = \frac{1}{1-z} \sum_{k \ge 0} (zx)^k = \frac{1}{(1-z)(1-zx)}.$$

Hence

$$G(x,z) = \frac{zg(z) - \frac{x}{(1-z)(1-zx)}}{zx^2 + (2z-1)x + z}$$

This research was supported by the Austrian Research Society (FWF) under the project number P12599-MAT

Date: January 14, 1998.

Now note that

$$zx^2 + (2z-1)x + z = z\big(1 - xr_1(z)\big)\big(1 - xr_2(z)\big)$$

with

$$r_{1,2}(z) = -1 + \frac{1 \pm \sqrt{1-4z}}{2z}.$$

Now, again as Knuth, we argue that $G(x,z)$ has a power series expansion in z and x. Since the denominator vanishes for $x = r_2(z)$, the numerator should also vanish, which means

$$zg(z) = \frac{r_2(z)}{(1-z)(1-zr_2(z))},$$

which leads already to the solution

$$g(z) = \frac{r_2(z)}{z(1-z)(1-zr_2(z))}. \tag{2}$$

3. Explicit expressions

As it is perhaps not too surprising, there are some close connections with *Catalan numbers*, since

$$B(z) = \frac{1-\sqrt{1-4z}}{2z} = \sum_{n\ge 0} \frac{1}{n+1}\binom{2n}{n} z^n.$$

Note that

$$r_2(z) = -1 + B(z) = zB^2(z).$$

There is a convenient method to pull out coefficients, using (formal) residue calculus, compare e. g. [1]: On sets $z = u/(1+u)^2$, then

$$g(z) = \left[\frac{4}{3}\frac{1}{1+2u} - \frac{1}{3}\frac{(1+2u)(1-u)}{1-u^3}\right](1+u)^6.$$

Now

$$\begin{aligned}
[z^n]g(z) &= \frac{1}{2\pi i}\oint \frac{dz}{z^{n+1}} g(z)\\
&= \frac{1}{2\pi i}\oint \frac{du(1-u)(1+u)^{2n+2}}{(1+u)^3 u^{n+1}} g\big(z(u)\big)\\
&= [u^n](1-u)(1+u)^{2n-1} g\big(z(u)\big)\\
&= [u^n](1-u)(1+u)^{2n-1}\left[\frac{4}{3}\frac{1}{1+2u} - \frac{1}{3}\frac{(1+2u)(1-u)}{1-u^3}\right](1+u)^6\\
&= [u^n](1+u)^{2n+5}\left[\frac{4}{3}\frac{1-u}{1+2u} - \frac{1}{3}\frac{(1+2u)(1-u)^2}{1-u^3}\right].
\end{aligned}$$

Note that

$$[u^{n-k}](1-u)(1+u)^{2n+5} = 2\frac{k+3}{n+k+6}\binom{2n+5}{n-k}$$

and $(1+2u)(1-u)^2 = 1 - 3u^2 + 2u^3$ and further

$$\begin{aligned}
&[u^{n-3k}](1-3u^2+2u^3)(1+u)^{2n+5}\\
&\qquad = \frac{2(3n^2 - 54nk^2 - 153nk - 87n + 54k^3 + 54k^2 - 177k - 168)}{(n+8+3k)\,(n+7+3k)\,(n+6+3k)}\binom{2n+5}{n-3k}.
\end{aligned}$$

Hence

$$[z^n]g(z) = \frac{8}{3}\sum_{0\le k\le n}(-2)^k\frac{k+3}{n+k+6}\binom{2n+5}{n-k}$$
$$+\frac{2}{3}\sum_{0\le 3k\le n}\frac{3n^2-54nk^2-153nk-87n+54k^3+54k^2-177k-168}{(n+8+3k)\,(n+7+3k)\,(n+6+3k)}\binom{2n+5}{n-3k}. \tag{3}$$

This is admittedly not too nice, but it has only one (level of) summation, and we cannot hope to do better than that.

Another approach would be based on the formula

$$g(z) = \frac{1}{z^2(1-z)}\sum_{k\ge 1}\big(zB(z)\big)^{2k},$$

but that would lead to a double summation of the coefficients.

4. Asymptotics

The explicit form (2) suggests to expand $g(z)$ around its (dominant) singularity $z=\frac{1}{4}$, viz.

$$g(z) \sim \frac{64}{9} - \frac{512}{27}\sqrt{1-4z},$$

with more terms available.

Now singularity analysis of generating functions [2] tells us that

$$g_{n,0} = [z^n]g(z) \sim \frac{256}{27\sqrt{\pi}}\,4^n\,n^{-3/2}.$$

5. A recursion

With Zeilberger's algorithm [4] we can first *guess* a linear recursion with polynomial coefficients for the values $g_n := g_{n,0}$, which can be *proved* afterwards, just by looking at the corresponding linear differential equation with polynomial coefficients for the generating function $g(z)$.

In this manner we get for g_n the recursion

$$(n+6)g_{n+3} + (\tfrac{9}{2}n-21)g_{n+2} + (\tfrac{3}{2}n+6)g_{n+1} + (2n+9)g_n = 0$$

with initial values $g_0 = 1$, $g_1 = 3$ and $g_2 = 9$.

The corresponding differential equation for $g(z)$ can be written as

$$\frac{1}{2}z(z-1)(z+2)(4z-1)g'(z) + \frac{1}{2}(18z^3+9z^2-24z+6)g(z) - 3 = 0$$

with initial condition $g(0) = 1$. Now this differential equation can be solved (or, we might just check that its solution is (2)) and therefore the validity of the above recursion is easily verified.

A very useful tool in this context is the program Gfun [5]. It would also guess the above recursion, based on the first few values. Or, it would transform the explicit form of $g(z)$ into an algebraic equation for it, and this equation into an equivalent differential equation, from which the recursion with polynomial coefficients would be available.

6. The original recursion

The original recursion from [6] that we want to treat is

$$\widetilde{g}_{n,k} = \widetilde{g}_{n-1,k-1} + 2\widetilde{g}_{n-1,k} + \widetilde{g}_{n-1,k+1} + 1 \quad \text{for } 0 \le k < n \text{ and } n \ge 2 \tag{4}$$

with the assumption that all other values $\widetilde{g}_{n,k}$ are zero.

When we introduce the bivariate generating function

$$\widetilde{G}(x,z) = \sum_{0\le k\le n} \widetilde{g}_{n,k} x^k z^n \,,$$

this recursion leads by multiplication with $z^n x^k$ and summing up to the equation

$$G(x,z) = zxG(x,z) + 2zG(x,z) + \frac{z}{x}\big(G(x,z) - g(z)\big) + \sum_{\substack{0\le k<n,\\ n\ge 2}} x^k z^n,$$

with $\widetilde{g}(z) = \widetilde{G}(0,z)$. Of course we have

$$\sum_{\substack{0\le k<n,\\ n\ge 2}} x^k z^n = \sum_{0\le k<n} x^k z^n - z = \frac{1}{(1-z)(1-xz)} - \frac{1}{1-xz} - z\,.$$

This leads to the relation

$$\widetilde{G}(z,u) = \frac{z\widetilde{g}(z) - \frac{x}{(1-z)(1-xz)} + \frac{x}{1-xz} + xz}{zx^2 + (2z-1)x + z}\,.$$

Analogously to the previous derivation we get the solution for $\widetilde{g}(z)$:

$$\widetilde{g}(z) = \frac{r_2(z)}{(1-z)(1-zr_2(z))} - \frac{r_2(z)}{z(1-zr_2(z)} - r_2(z)\,.$$

To get an explicit expression for the coefficients of $\widetilde{g}(z)$ we also need

$$[z^n]r_2(z) = \frac{1}{n+1}\binom{2n}{n} \quad \text{for } n \ge 1$$

and

$$[z^n]\frac{r_2(z)}{z(1-zr_2(z))} = [u^n](1-u)(1+u)^{2n-1}\frac{(1+u)^4}{2u-1}\,.$$

This yields

$$[z^n]\frac{r_2(z)}{z(1-zr_2(z))} = [u^n]\frac{(1-u)(1+u)^{2n+3}}{2u+1} = \sum_{0\le k\le n}(-2)^k\frac{2k+4}{n+k+4}\binom{2n+3}{n-k}\,.$$

Combining the above formulæ we thus get an explicit expression for the coefficients of $\widetilde{g}(z)$ and therefore for the required parameter. For $n \ge 1$ we have

$$\begin{aligned}[z^n]\widetilde{g}(z) = {}& \frac{8}{3}\sum_{0\le k\le n}(-2)^k\frac{k+3}{n+k+6}\binom{2n+5}{n-k} - \sum_{0\le k\le n}(-2)^k\frac{2k+4}{n+k+4}\binom{2n+3}{n-k} - \frac{1}{n+1}\binom{2n}{n}\\ &+ \frac{2}{3}\sum_{0\le 3k\le n}\frac{-87n-177k+3n^2+54k^3+54k^2-54nk^2-168-153nk}{(n+8+3k)\,(n+7+3k)\,(n+6+3k)}\binom{2n+5}{n-3k}\,.\end{aligned} \tag{5}$$

To evaluate this expression asymptotically, we expand $\widetilde{g}(z)$ around its algebraic singularity $z = \frac{1}{4}$, and we obtain

$$\widetilde{g}(z) \sim \frac{7}{9} - \frac{74}{27}\sqrt{1-4z}\,,$$

which leads to

$$[z^n]\widetilde{g}(z) \sim \frac{37}{27\sqrt{\pi}} 4^n n^{-3/2}\,.$$

It is also of interest to present a linear recursion with polynomial coefficients for the numbers $\widetilde{g}_n := \widetilde{g}_{n,0}$. We get here

$$\begin{aligned} 0 &= \frac{1}{17}\,(n+6)\,(17n+25)\,\widetilde{g}_{n+4} - \frac{1}{238}\left(1357n^2+7775n+7890\right)\widetilde{g}_{n+3} \\ &\quad + \frac{3}{119}\left(274n^2+941n+605\right)\widetilde{g}_{n+2} + \frac{1}{238}\left(47n^2+1897n+2790\right)\widetilde{g}_{n+1} \\ &\quad - \frac{1}{119}\,(143n+315)\,(2n+1)\,\widetilde{g}_n\,. \end{aligned}$$

This assertion can be proved by translating the recursion for $\widetilde{g}_n$ to the corresponding differential equation for the generating function $\widetilde{g}(z)$. The validity of the resulting equation

$$\begin{aligned} 0 &= z^2(143z-119)(z-1)(4z-1)(z+2)\widetilde{g}''(z) \\ &\quad + 2z(1059z^4-925z^3-357z^2+495z-56)\widetilde{g}'(z) \\ &\quad + 2(315z^4-470z^3+543z^2-1611z+602)\widetilde{g}(z) - 6z^2(105z^2-128z+84) \end{aligned}$$

is then easy to check.

7. Some numerical data

The following Table 1 contains the analysed parameters g_n and $\widetilde{g}_n$ for a few values of n. We compare the exact values with the asymptotic equivalents $a_n := \frac{256}{27\sqrt{\pi}} 4^n n^{-3/2}$ resp. $\widetilde{a}_n := \frac{37}{27\sqrt{\pi}} 4^n n^{-3/2}$, and give also the ratios $\frac{g_n}{a_n}$ and $\frac{\widetilde{g}_n}{\widetilde{a}_n}$.

References

[1] N. G. De Bruijn, D. E. Knuth, and S. O. Rice. The average height of planted plane trees. In R. C. Read, editor, *Graph Theory and Computing*, pages 15–22. Academic Press, 1972.

[2] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[3] D. E. Knuth. *The Art of Computer Programming*, volume 1: Fundamental Algorithms. Addison-Wesley, 1968. Third edition, 1997.

[4] M. Petkovšek, H. Wilf, and D. Zeilberger. $A = B$. A.K. Peters, Ltd., 1996.

[5] B. Salvy and P. Zimmermann. GFUN: a Maple package for the manipulation of generating and holonomic functions in one variable. *ACM Transactions on Mathematical Software*, 20(2):163–167, 1994.

[6] L. Xiang, C. Tang, and K. Ushijima. Grammar–oriented enumeration of binary trees. *The Computer Journal*, 40:278–291, 1997.

	g_n	$\widetilde{g}_n$	$\frac{g_n}{a_n}$	$\frac{\widetilde{g}_n}{\widetilde{a}_n}$
0	1	0		
1	3	0	0.140	0
2	9	1	0.298	0.228
3	27	4	0.407	0.419
4	84	13	0.489	0.524
5	270	42	0.550	0.594
6	892	138	0.599	0.642
7	3012	463	0.638	0.675
8	10350	1582	0.666	0.706
9	36074	5488	0.696	0.727
10	127218	19278	0.716	0.749
20	$0.551 \cdot 10^{11}$	$0.818 \cdot 10^{10}$	0.838	0.862
50	$0.178 \cdot 10^{29}$	$0.261 \cdot 10^{28}$	0.926	0.938
100	$0.828 \cdot 10^{58}$	$0.120 \cdot 10^{58}$	0.961	0.968
200	$0.479 \cdot 10^{118}$	$0.695 \cdot 10^{117}$	0.980	0.983
500	$0.509 \cdot 10^{298}$	$0.736 \cdot 10^{297}$	0.990	0.994

TABLE 1.