A NOTE ON THE DISTRIBUTION OF THE THREE TYPES OF NODES IN UNIFORM BINARY TREES

HELMUT PRODINGER

Abstract. We use Zeilberger's algorithm to compute some sums that came up in Mahmoud's analysis of the distribution of types of nodes in binary trees.

In [4], Mahmoud has considered uniform binary trees (counted by Catalan numbers $b_n = \frac{1}{n+1}\binom{2n}{n}$) and the three statistics $X_n^{(0)}$, $X_n^{(1)}$, $X_n^{(2)}$, counting the numbers of internal nodes having 0, 1, 2 internal nodes as successors, respectively. He obtained the following theorem (with one typo removed):

Theorem 1. [Mahmoud]

$$\Pr\left\{X_n^{(0)} = j\right\} = \frac{1}{b_n}\sum_{i=0}^{n-j}(-1)^{n-j-i} b_i \binom{i+1}{n-i}\binom{n-i}{j},$$

$$\Pr\left\{X_n^{(1)} = j\right\} = \frac{1}{b_n}\sum_{i=0}^{n-j}(-1)^{n-j-i} 2^{n-i} b_i \binom{n-1}{i-1}\binom{n-i}{j},$$

$$\Pr\left\{X_n^{(2)} = j\right\} = \frac{1}{b_n}\sum_{k=0}^{j}(-1)^{j-k} b_k \binom{n-k}{j-k}\sum_{i=0}^{n-k} 2^{n-k-i}\binom{n-1}{k+i-1}\binom{k+1}{i}.$$

Mahmoud used generating functions to get these results. For convenience, we sketch how to do it.

The generating function $B(z)$ of the binary trees is normally obtained via the equation $B = 1 + zB^2$. However, we can also get it via the equation $B = 1 + z + 2z(B-1) + z(B-1)^2$. Observe that $B(z) - 1$ is the generating function for the nonempty trees, and the recursion now distinguishes between zero, one, or two successors. Now, we can use additional variables u, v, w in order to count nodes with zero, one, or two successors. Then the equation is

$$B = 1 + zu + 2zv(B-1) + zw(B-1)^2,$$

with the solution

$$B(z; u, v, w) = \frac{1 - 2z(v-w) - \sqrt{1 - 4zv + 4z^2(v^2 - uw)}}{2zw}.$$

Date: November 1, 1996.

Now the probabilities from Theorem 1 are obtained by reading off appropriate coefficients:

$$\Pr\left\{X_n^{(0)} = j\right\} = \frac{1}{b_n}[z^n u^j]B(z;u,1,1)\,,$$
$$\Pr\left\{X_n^{(1)} = j\right\} = \frac{1}{b_n}[z^n v^j]B(z;1,v,1)\,,$$
$$\Pr\left\{X_n^{(2)} = j\right\} = \frac{1}{b_n}[z^n w^j]B(z;1,1,w)\,.$$

With Zeilberger's algorithm Ekhad (see [5]), I found the following explicit formulæ:

Theorem 2.

$$\Pr\left\{X_n^{(0)} = j\right\} = \frac{2^{n+1-2j}(n+1)!\,n!\,(n-1)!}{j!\,(j-1)!\,(n+1-2j)!\,(2n)!}\,,$$
$$\Pr\left\{X_{2n+1}^{(1)} = 2j\right\} = \frac{2^{2j-1}(2n+2)!\,(2n)!\,(2n)!}{(2j)!\,(n-j+1)!\,(n-j)!\,(4n+1)!}\,,$$
$$\Pr\left\{X_{2n}^{(1)} = 2j+1\right\} = \frac{2^{2j+1}(2n+1)!\,(2n)!\,(2n-1)!}{(2j+1)!\,(n-j)!\,(n-j-1)!\,(4n)!}\,,$$
$$\Pr\left\{X_n^{(1)} = j\right\} = 0 \quad \textit{for } n+j \textit{ even}\,,$$
$$\Pr\left\{X_n^{(2)} = j\right\} = \frac{2^{n-1-2j}(n+1)!\,n!\,(n-1)!}{(j+1)!\,j!\,(n-1-2j)!\,(2n)!}\,.$$

The quantity

$$\Pr\left\{X_n^{(0)} = j\right\} b_n = \frac{2^{n+1-2j}(n-1)!}{j!\,(j-1)!\,(n+1-2j)!}$$

is the number of binary trees of size n and j leaves. It is interesting to compare this quantity with the corresponding one for planted plane trees; the number of planted plane trees of size n with j leaves is the *Narayana* number

$$\frac{1}{n}\binom{n}{j}\binom{n-2}{j-1} = \frac{(n-1)!\,(n-2)!}{j!\,(j-1)!\,(n-1-j)!^2}\,,$$

see [3].

We provide a glimpse of what Zeilberger's algorithm does. Dealing with the first sum in Theorem 1, and setting

$$F(n,i) := \frac{1}{b_n}(-1)^{n-j-i}b_i\binom{i+1}{n-i}\binom{n-i}{j}\,,$$

where we treat j as a parameter, the algorithm computes

$$G(n,i) := \frac{(n+2)(n-2i)(n-2i-1)}{2(n+1-i-j)}\, F(n,i)\,,$$

such that

$$\begin{aligned}(n-2j+2)(2n+1)\,F(n+1,i) - (n+2)n\,F(n,i) &= \\ &= G(n,i+1) - G(n,i)\,.\end{aligned}$$

Denoting the sum on i by $F(n)$ and summing up, the right hand side telescopes, and we get

$$(n-2j+2)(2n+1)\,F(n+1) = (n+2)n\,F(n)\,.$$

Such a *first order recursion* can always be solved by *iteration*, leading to the announced formula.

If we do the same thing for the second sum, we find the recursion

$$\begin{aligned}(2n+3)(2n+1)(n-j+3)(n-j+1)\,F(n+2) &= \\ &= (n+3)(n+2)(n+1)n\,F(n)\end{aligned}$$

which is not of first order, but can, because of the lack of the term $F(n+1)$, still be solved by iteration. It turns out that four cases (n, j even or odd) have to be distinguished.

The third entry is a double sum and thus not so easily treatable. I recently showed it to F. Chyzak, and he is able to handle it with his package MGFUN (see [1]). However, because of the following argument, the double sum is reducible to the other instances. We will show that the three statistics are basically the same. Assume that there are i nodes with 0 successors, j nodes with 1 successor, and k nodes with 2 successors. Then we have by an elementary argument the equations

$$\begin{aligned} i+j+k &= n\,, \\ j+2k &= n+1\,, \end{aligned}$$

whence $k = i+1$ and furthermore $j = n-2i-1$. Thus, with

$$H(n,j) := \frac{2^{n+1-2j}(n+1)!\,n!\,(n-1)!}{j!\,(j-1)!\,(n+1-2j)!\,(2n)!}\,,$$

we get alternatively

$$\begin{aligned}
&\Pr\left\{X_n^{(0)}=j\right\}=H(n,j)\,,\\
&\Pr\left\{X_n^{(1)}=j\right\}=H(n,\tfrac{n-j+1}{2}), \qquad \text{for } n+j \text{ is odd}\,,\\
&\Pr\left\{X_n^{(1)}=j\right\}=0, \qquad \text{for } n+j \text{ even}\,,\\
&\Pr\left\{X_n^{(2)}=j\right\}=H(n,j+1)\,.
\end{aligned}$$

The fact, that the covariance matrix has rank 1 (see [4]) becomes perhaps a little bit more transparent by this combinatorial argument.

With Zeilberger's algorithm we can do even more: We can compute the s–th factorial moments explicitly. Let $n^{\underline{k}}=n(n-1)\dots(n-k+1)$, and

$$M_{n;s}^{(i)}:=\sum_j \Pr\left\{X_n^{(i)}=j\right\}j^{\underline{s}}$$

denote the s–th factorial moments for $i=0,1,2$, then:

Theorem 3.

$$\begin{aligned}
M_{n;s}^{(0)}&=\frac{(n+1)!\,n!\,(2n-2s)!}{(2n)!\,(n+1-2s)!\,(n-s)!}\,,\\
M_{n;s}^{(1)}&=\frac{2^s\,(n+1)!\,n!\,(n-1)!\,(2n-2s)!}{(2n)!\,(n-s+1)!\,(n-s)!\,(n-s-1)!}\,,\\
M_{n;s}^{(2)}&=\frac{n!\,(n-1)!\,(2n-2s)!}{(2n)!\,(n-1-2s)!\,(n-s)!}\,.
\end{aligned}$$

We end this little note by citing two other papers ([2], [6]) dealing with binary trees and statistics on the leaves.

References

[1] F. Chyzak. Holonomic systems and automatic proofs of identities. Research Report 2371, Institut National de Recherche en Informatique et en Automatique, October 1994.

[2] P. Kirschenhofer. On the height of leaves in binary trees. *Journal of Combinatorics, Information and System Sciences*, 8:44 –60, 1983.

[3] G. Kreweras. Sur les eventails de segments. *Cahiers du B.U.R.O*, 15:255–261, 1970.

[4] H. M. Mahmoud. The joint distribution of the three types of nodes in uniform binary trees. *Algorithmica*, 13:313–323, 1995.

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

[6] F. Ruskey. On the average shape of binary trees. *SIAM Journal on Algebraic and Discrete Methods*, 1:43 –50, 1980.

Institut für Algebra und Diskrete Mathematik, Technical University of Vienna, Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria.
E-mail address: Helmut.Prodinger@tuwien.ac.at
WWW-address: http://info.tuwien.ac.at/theoinf/proding.htm