Words, Dyck paths, Trees, and Bijections

Helmut Prodinger
The John Knopfmacher Centre for Applicable Analysis and Number Theory
Department of Mathematics, University of the Witwatersrand
P. O. Wits, 2050 Johannesburg, South Africa
helmut@gauss.cam.wits.ac.za
http://www.wits.ac.za/helmut/index.htm

Dedicated to Gabriel Thierrin
on the occasion of his eightieth birthday

Abstract

In [1] the concept of nondecreasing Dyck paths was introduced. We continue this research by looking at it from the point of view of words, rational languages, planted plane trees, and continued fractions. We construct a bijection with planted plane trees of height ≤ 4 and compute various statistics on trees that are the equivalents of nondecreasing Dyck paths.

Personal reminiscences about Gabriel Thierrin

Gabriel Thierrin invited me to London, Ontario, for six weeks in February/March 1982. My memories about that trip are still very much alive since it was my first crossing of the atlantic ocean, and I was a very inexperienced traveller at that time, and everything was new to me. Snow storms, arctic temperatures, clear blue skies, frozen sidewalks! I had the opportunity to visit Waterloo, to give a talk, and to see the Niagara falls, and to learn that Canada has best proportion of great rock bands versus total population.

I am only guessing, but I think that Gabriel Thierrin was the referee of [4] and was attracted by the combination of formal languages and other mathematical concepts, perhaps not the standard ones in this context. I always liked the concepts of words, languages, grammars, and automata, but I also wanted to see them in a wider context, mostly in a combinatorial one. This is still true today, where I only occasionally bump into some (formal) languages.

Gabriel Thierrin invited me to his house several times, and he and his wife were extremely friendly and helpful. Once, he gave a party, and David Borwein also attended. Later in life I met his sons Jonathan and Peter.

We also wrote the paper [5] together.

I remember much more, more than about any other trip I guess, but perhaps I should rather stop here.

In the technical part of this paper, I want to demonstrate a charming interplay of Dyck paths (related to Dyck words, of course), certain rational languages and their associated generating functions (being best described as continued fractions), and some families of trees. The form of the generating functions cries out for bijections, and they are described in the sequel. Several characteristic parameters are also counted.

1 Introduction

In the paper [1], the Italian authors come up with the lovely new concept of *nondecreasing Dyck paths.* Dyck words are geometric renderings of Dyck paths where an open bracket is coded by an upward step, and a closing bracket by a downward step. The condition "nondecreasing" means roughly that the sequence of the altitudes of the *valleys* must be *nondecreasing.* We prefer to think about it in terms of *planted plane trees*; there is an obvious and well–known bijection, [2, 6].

In honour of one of the authors, we decide to call the corresponding trees *Elena trees*, or simply *Elenas*[1].

In [1] the generating function of nondecreasing Dyck paths of length $2n$ was already found to be $\frac{z(1-z)}{1-3z+z^2}$. We find it practical also to include

[1]In the literature, there are also *Patricia* trees (tries).

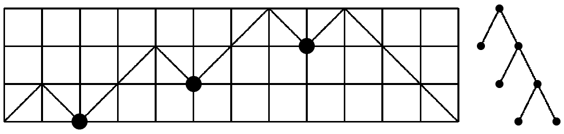


Figure 1: A nondecreasing Dyck path with valleys indicated and the corresponding planted plane tree

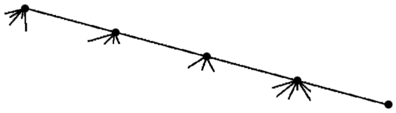


Figure 2: A typical Elena; the short lines indicate paths of various lengths

the *empty path*, which gives us

$$1+\frac{z(1-z)}{1-3z+z^2}=\frac{1-2z}{1-3z+z^2}\ .$$

Since the length $2n$ corresponds to an Elena of size (=number of nodes) $n+1$, we find the generating function of Elenas as

$$E(z)=\sum_{n\ge0}[\textsf{Number of Elenas of size } n]\,z^n=\frac{z(1-2z)}{1-3z+z^2}\ .$$

Now here is an easy argument to see that directly. We use the letter $\mathbf{p}$ to describe an arbitrary path of length ≥ 1 and the letter $\mathbf{a}$ which means '**a**dvance to next node on the rightmost branch.' Then the set of Elenas $\mathcal{E}$ is given by the symbolic equation (a rational language)

$$\mathcal{E}=\big(\mathbf{ap}^*\big)^*\mathbf{a}\ . \tag{1}$$

Now mapping $\mathbf{a}\mapsto z$ and $\mathbf{p}\mapsto\frac{z}{1-z}$ we find the generating function in the nice *continued fraction* form

$$E(z) = \cfrac{z}{1 - \cfrac{z}{1 - \cfrac{z}{1-z}}}.$$

The continued fraction form suggests a relation to planted plane trees of height ≤ 4; a bijection is constructed in the next section.

The following sections consider average values of several simple parameters of Elenas. For simplicity, we give only first order asymptotics, but explicit values (in terms of Fibonacci and Lucas numbers) and also variances should not be too hard to obtain.

Then we deal with the harder problem of the average height of (random) Elenas of size n.

We will use the number $\alpha = \frac{1+\sqrt{5}}{2}$ frequently in this paper, since it is prominent in the asymptotics of Fibonacci numbers (and thus also Elenas).

2 A bijection

The continued fraction representation for $E(z)$ is well known in tree enumeration; it enumerates the set of planted plane trees with height ≤ 4 (compare e.g. [2, 6]).

Now we will describe a bijection between Elenas and those trees. We start from the representation $(\mathbf{ap}^*)^*\mathbf{a}$ and give an alternative interpretation of the words in this set as height restricted trees.

First, a path with n nodes (coded by $\mathbf{p}_n$) will be interpreted as a root, followed by $n-1$ subtrees of size 1.

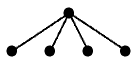


Figure 3: Interpretation of a path with 5 nodes

Then, a word $\mathbf{ap}\ldots\mathbf{p}$ will be interpreted as a root, followed by subtrees given by the $\mathbf{p}$'s.

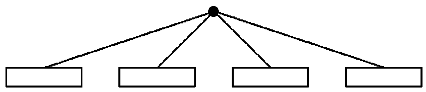


Figure 4: Interpretation of a **apppp**; the boxes are the interpretations of the respective paths

Finally, the last **a** will be the root, and the **ap** . . . **p**'s become subtrees of it. Figure 6 describes the process.

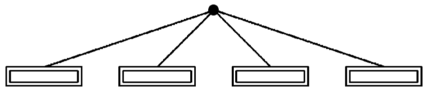


Figure 5: Interpretation of a $(\mathbf{ap}^*)(\mathbf{ap}^*)(\mathbf{ap}^*)(\mathbf{ap}^*)\mathbf{a}$; the boxes are the interpretations of the respective $(\mathbf{ap}^*)$'s; the last **a** serves as the root

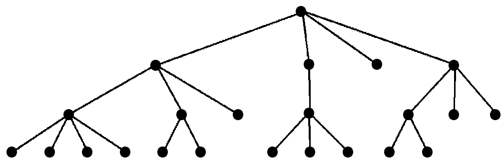


Figure 6: Interpretation of $(\mathbf{ap}_5\mathbf{p}_3\mathbf{p}_1)(\mathbf{ap}_4)(\mathbf{a})(\mathbf{ap}_3\mathbf{p}_1\mathbf{p}_1)\mathbf{a}$; $\mathbf{p}_i$ stands for a path with i nodes

Geometrically, one can imagine the process as follows. We consider the rightmost branch of an Elena, take its right node as a root, move the rest into horizontal position and rearrange the edges so that the nodes are successors of the root. The attached paths are then rearranged as described.

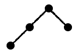
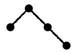
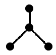
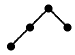
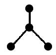
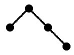
	aaaa	**ap$_2$a**	**ap$_1$aa**	**ap$_1$p$_1$a**	**aap$_1$a**
Elena					
Height restricted					

Figure 7: The bijection exemplified on trees with 4 nodes

3 Average degree of the root

We use a second variable u to label the degree of the root and obtain easily

$$T(z,u) = z + \frac{z}{1 - \dfrac{uz}{1-z}} \frac{uz(1-2z)}{1-3z+z^2} .$$

To compute the average value, we have to differentiate $T(z,u)$ with respect to u and then to set $u = 1$. This yields

$$\frac{\partial}{\partial u} T(z,u) \bigg|_{u=1} = \frac{z^2(1-z)^2}{(1-2z)(1-3z+z^2)} .$$

Around the (dominant) singularity $z = 1/\alpha^2$ we have

$$\frac{z^2(1-z)^2}{(1-2z)(1-3z+z^2)} \sim \frac{5-\sqrt{5}}{10} \frac{1}{1-z\alpha^2} ,$$

so that

$$[z^n] \frac{z^2(1-z)^2}{(1-2z)(1-3z+z^2)} \sim \frac{5-\sqrt{5}}{10} \alpha^{2n} .$$

Since

$$[z^n] \frac{z(1-2z)}{1-3z+z^2} \sim \left(1 - \frac{2}{\sqrt{5}}\right) \alpha^{2n}$$

the average degree of the root is asymptotic to

$$\frac{3+\sqrt{5}}{2} = 2.618033989 \; .$$

4 Average number of leaves

Replace $\mathbf{a} \mapsto z$ and $\mathbf{p} \mapsto \frac{zu}{1-z}$ in (1) and multiply the whole thing by u to get the bivariate generating function

$$\cfrac{zu}{1-\cfrac{z}{1-\cfrac{zu}{1-z}}} \; .$$

Differentiate w. r. t. u, then set $u = 1$ to get $\dfrac{z(1-5z+8z^2-3z^3)}{(1-3z+z^2)^2}$. Around the singularity $z = 1/\alpha^2$ we have

$$\frac{z(1-5z+8z^2-3z^3)}{(1-3z+z^2)^2} \sim \frac{-2+\sqrt{5}}{5}\frac{1}{(1-z\alpha^2)^2} \; ,$$

so that

$$[z^n]\frac{z(1-5z+8z^2-3z^3)}{(1-3z+z^2)^2} \sim \frac{-2+\sqrt{5}}{5} n\alpha^{2n} \; .$$

Since

$$[z^n]\frac{z(1-2z)}{1-3z+z^2} \sim \left(1-\frac{2}{\sqrt{5}}\right)\alpha^{2n}$$

the average number of leaves is asymptotic to

$$\frac{n}{\sqrt{5}} = 0.4472135956\, n \; .$$

5 Average number of paths

Replace $\mathbf{a} \mapsto z$ and $\mathbf{p} \mapsto \frac{zu}{1-z}$ to get the bivariate generating function

$$\cfrac{z}{1-\cfrac{z}{1-\cfrac{zu}{1-z}}}\,.$$

Differentiate w. r. t. u, then $u=1$ yields $\dfrac{z^3(1-z)}{(1-3z+z^2)^2}$. Hence $[z^n]\dfrac{z^3(1-z)}{(1-3z+z^2)^2} \sim \dfrac{-2+\sqrt{5}}{5}\,n\,\alpha^{2n}$. Thus the average number of paths is asymptotic to

$$\frac{n}{\sqrt{5}} = 0.4472135956\,n\ .$$

6 Average number of nodes 'a'

Replace $\mathbf{a} \mapsto zu$ and $\mathbf{p} \mapsto \frac{z}{1-z}$ to get the bivariate generating function

$$\cfrac{zu}{1-\cfrac{zu}{1-\cfrac{z}{1-z}}}\,.$$

Differentiate w. r. t. u, then $u=1$ yields $\dfrac{z(1-2z)^2}{(1-3z+z^2)^2}$. Hence

$$[z^n]\frac{z(1-2z)^2}{(1-3z+z^2)^2} \sim \frac{7-3\sqrt{5}}{10}\,n\,\alpha^{2n}\ .$$

Thus the average number of **a**'s is asymptotic to

$$\frac{5-\sqrt{5}}{10}n = 0.2763932022\,n\ .$$

As a corollary, we get that the average number of nodes lying in *paths* is asymptotic to

$$n - \frac{5-\sqrt{5}}{10}n = \frac{5+\sqrt{5}}{10}n = 0.7236067978\,n \ .$$

And furthermore the average number of nodes in *one* path is asymptotic to

$$\frac{5+\sqrt{5}}{10}n \Big/ \frac{n}{\sqrt{5}} = \frac{1+\sqrt{5}}{2} = 1.618033989 \ .$$

7 Number of descendants

The number of descendants of a node is the size of the subtree with this node as the root. The paper [3] deals e. g. extensively with this subject. We want to know the average number of descendants. This is an average over both, the Elenas, and the nodes in an Elena. Thus it is meaningful to define for an Elena t

$$\psi(t) := \sum_{v \text{ a node of } t} [\text{number of descendants of } v]$$

and

$$D(z,u) := \sum_{t \in \mathcal{E}} z^{|t|} u^{\psi(t)} \ ;$$

then we find the desired average as $\frac{1}{n}[z^n]\frac{\partial}{\partial u}D(z,u)\big|_{u=1}$. Now we want to derive a functional equation for this function $D(z,u)$. Of course we follow the general decomposition (1). The contribution of each path attached to the root is

$$Q(z,u) = \sum_{m \geq 1} z^m u^{\binom{m+1}{2}} \ .$$

The contribution of the root is zu^n, which is handled by first neglecting it and then substituting zu for z. Altogether we find

$$D(z,u) = zu + \frac{zu\,D(zu,u)}{1 - Q(zu,u)} \ .$$

Now let us differentiate this w. r. t. to u and plug in $u = 1$. We can also use the special values

$$D(z,1) = E(z) \quad \text{and} \quad \frac{\partial}{\partial z} D(z,1) = \frac{1-4z+5z^2}{(1-3z+z^2)^2}$$

as well as

$$Q(z,1) = \frac{z}{1-z} \text{ and } \frac{\partial}{\partial z} Q(z,1) = \frac{1}{(1-z)^2} \text{ and } \frac{\partial}{\partial u} Q(z,1) = \frac{z}{(1-z)^3} \; .$$

The resulting equation contains only one unknown function, $\frac{\partial}{\partial u} D(z,u)\big|_{u=1}$, and Maple solves it as

$$\frac{\partial}{\partial u} D(z,u)\bigg|_{u=1} = \frac{z(1-7z+20z^2-26z^3+11z^4)}{(1-z)(1-3z+z^2)^3} \sim \frac{7-3\sqrt{5}}{10} \frac{1}{(1-z\alpha^2)^3} \; .$$

Hence

$$\frac{1}{n}[z^n] \frac{\partial}{\partial u} D(z,u)\bigg|_{u=1} \sim \frac{7-3\sqrt{5}}{10} \frac{n}{2} \alpha^{2n} \; .$$

Dividing this quantity by the asymptotic equivalent for the total number, $\left(1 - \frac{2}{\sqrt{5}}\right)\alpha^{2n}$, we get the average number of descendants as

$$\frac{5-\sqrt{5}}{20} n = 0.1381966011\, n \; .$$

8 Number of ascendants

The number of ascendants of a node is defined to be the number of nodes on the path of the node to the root. It is also called the *depth*. And the sum over all depths (summed over all nodes in the Elena) is called the *path length*. It is very similar to the *area*, studied in the paper [1].

However, it is quite easy to see that the average number of ascendants equals the average number of descendants: Consider two nodes i and j

such that i lies on the path from the root to j. Then i appears in the count of j of the number of ascendants, and j appears in the count of i of the number of descendants. Since these quantities are summed over all nodes, we are done. (This argument is general and not restricted to Elenas.)

9 Average height of Elenas

The recursion $\mathcal{E} = a + (\mathbf{ap}^*)\mathcal{E}$ translates into $E = z + \frac{z(1-z)}{1-2z}E$ and also into the recursion for E_h, the generating functions of Elenas of height $\le h$,

$$E_h = z + \frac{z(1-z)}{1-2z+z^h}E_{h-1} \ .$$

Denoting the generating functions of Elenas of height $> h$ by U_h, we find by taking differences

$$(1-2z+z^h)U_h = \frac{(1-z)z^{h+2}}{1-3z+z^2} + z(1-z)U_{h-1} \ .$$

We find the average height as

$$\frac{[z^n]\sum_{h\ge 0} U_h(z)}{[z^n]E(z)} \ .$$

Now define $\mathcal{U}(z,w) := \sum_{h\ge 0} U_h(z)w^h$. Summing up we get

$$\begin{aligned} &(1-2z)\,\mathcal{U}(z,w) + \mathcal{U}(z,zw) \\ &= \frac{2z(1-z)(1-2z)}{1-3z+z^2} + \frac{z^3w(1-z)}{(1-3z+z^2)(1-wz)} + wz(1-z)\,\mathcal{U}(z,w) \ . \end{aligned}$$

The instance $w=1$ is of special interest;

$$(1-3z+z^2)\,\mathcal{U}(z,1) + \mathcal{U}(z,z) = \frac{z(2-6z+5z^2)}{1-3z+z^2} \ .$$

From this we see that $\mathcal{U}(z,1)$ has a double pole at the dominant singularity $z=\lambda:=1/\alpha^2$. Since for all $h\ge 0$

$$U_h(z)\sim\frac{\lambda(1-2\lambda)}{1-3z+z^2}\ ,$$

we infer that $\mathcal{U}(z,z)\sim\dfrac{\lambda(1-2\lambda)}{1-\lambda}\dfrac{1}{1-3z+z^2}$. Hence

$$\mathcal{U}(z,1)\sim\frac{47-21\sqrt{5}}{2}\frac{1}{(1-3z+z^2)^2}\sim\frac{7-3\sqrt{5}}{10}\frac{1}{(1-z\alpha)^2}$$

and

$$[z^n]\,\mathcal{U}(z,1)\sim\frac{7-3\sqrt{5}}{10}\,n\,\alpha^{2n}\ .$$

Dividing this by $\left(1-\frac{2}{\sqrt{5}}\right)\alpha^{2n}$, we find for the average height the asymptotic equivalent

$$\frac{5-\sqrt{5}}{10}n=0.2763932022\,n\ .$$

10 Conclusion

For the reader's convenience we collect our findings in a small table.

References

[1] E. Barcucci, A. Del Lungo, S. Fezzi, and R. Pinzani. Nondecreasing Dyck paths and q–Fibonacci numbers. *Discrete Mathematics*, 170:211–217, 1997.

[2] 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.

[3] C. Martínez, A. Panholzer, and H. Prodinger. Descendants and ascendants in random search trees. *Electronic Journal of Combinatorics*, 5 (R20), 1998.

Degree of root	$\frac{3+\sqrt{5}}{2}$
Number of leaves	$\frac{n}{\sqrt{5}}$
Number of paths	$\frac{n}{\sqrt{5}}$
Number of nodes on rightmost branch	$\frac{5-\sqrt{5}}{10}n$
Number of nodes in paths	$\frac{5+\sqrt{5}}{10}n$
Number of nodes in one path	$\frac{1+\sqrt{5}}{2}$
Number of ascendants	$\frac{5-\sqrt{5}}{20}n$
Number of descendants	$\frac{5-\sqrt{5}}{20}n$
Height	$\frac{5-\sqrt{5}}{10}n$

Table 1: Several averages on Elenas

[4] H. Prodinger. Congruences defined by languages and filters. *Information and Control*, 44:36–46, 1980.

[5] H. Prodinger and G. Thierrin. Towards a general concept of hypercodes. *Journal of Information and Optimization Sciences*, 4:255–268, 1983.

[6] R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms.* Addison-Wesley, 1996.