Fundamenta Informaticae CCXXI (2011) 1–7
IOS Press

Dyck paths with parity restrictions for the final runs to the origin: a study of the height

Helmut Prodinger
Department of Mathematics
University of Stellenbosch
7602 Stellenbosch
South Africa hproding@sun.ac.za

Dedicated to Philippe Flajolet 1948–2011

Abstract. Stanley and Callan considered Dyck paths where the lengths of the run to the origin is always odd resp. the last one even, and the other ones odd. These subclasses are also enumerated by (shifted) Catalan numbers. We study the (average) height of these objects, assuming all such Dyck paths of length $2n$ to be equally likely, and find that it behaves like $\sim \sqrt{\pi n}$, as in the unrestricted case. This classic result for unrestricted Dyck paths is from de Bruijn, Knuth and Rice [2], and to this day, there are no simpler proofs for this, although more general results have been obtained by Flajolet and Odlyzko [4].

Keywords: Dyck paths, height, generating functions, asymptotic equivalent.

1. Introduction

It is a folklore result that the number of Dyck paths of length $2n$ (=semi-length n) is the Catalan number $\frac{1}{n+1}\binom{2n}{n}$. Two standard texts that describe this are [11, 7]. Callan [1] pointed out that there are interesting subclasses, which are enumerated by Catalan number $\frac{1}{n}\binom{2n-2}{n-1}$. The first one was already known to Stanley [11]: each (maximal) run back to the x-axis consists of an *odd* number of down steps. The second one is due to Callan himself: For the last sojourn, the (maximal) run back to the x-axis consists of an *even* number of down steps, for the other ones, these lengths are still odd. This works for a semi-length $n \geq 2$; the Stanley instance works for a semi-length $n \geq 1$, and we even allow the empty path.

In this paper, we confirm these results by generating functions, and then consider the *height* of them. For ordinary Dyck paths, it is a classical result [2] that the average height is asymptotic to $\sqrt{\pi n}$; in the cited paper, this was discussed in the equivalent notion of planar (=planted plane) trees.

Our generating functions are either w.r.t. length or semi-length, but that will always be clear from the context. To set them up, we start by marking a step by z, but when this is achieved, we switch back to the more convenient semi-length. Here are a few basic formulæ:

$$C(z) = \frac{1-\sqrt{1-4z}}{2z} = \sum_{n\ge0} \frac{1}{n+1}\binom{2n}{n} z^n$$

enumerates Dyck paths w.r.t. semi-length. The height of a Dyck path is the maximum y-coordinate of it. The generating function $C^{[h]}(z)$ where the coefficient of z^n is the number of Dyck paths of length $2n$ and height $\le h$ is given by

$$C^{[h]} = C^{[h]}(z) = \frac{\lambda_1^{h+1} - \lambda_2^{h+1}}{\lambda_1^{h+2} - \lambda_2^{h+2}}, \tag{1}$$

with

$$\lambda_{1,2} = \frac{1 \pm \sqrt{1-4z}}{2}.$$

Notice that in this formula, we can take the limit for $h \to \infty$, and get $C(z) = 1/\lambda_1(z)$. This counts Dyck paths without height restrictions, i. e., all Dyck paths. Here are a few references: [2, 9, 8, 10]

2. Dyck paths à la Stanley

A natural way to decompose Dyck paths is with respect to *sojourns*: A sojourn is a part of such a path between two consecutive visits to the x-axis. The natural decomposition of a sojourn (with the restrictions à la Stanley) is: an up step, a closed Dyck path, repeated i times (which leads us to level i), followed by an up step, and then the final $i+1$ down steps that bring us back to the x-axis. The level i has to be even, to satisfy the constraints.

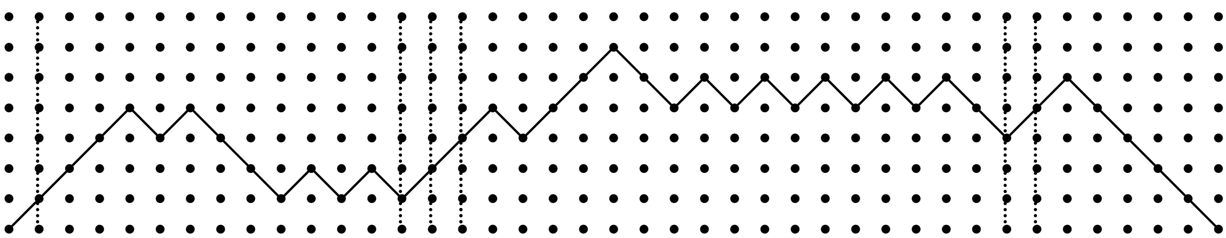

Figure 1. A lattice path of length 36 (semi-length 18) and height 6; the run back to the x-axis has length 5. The decomposition is indicated by vertical lines.

So the generating function (at the moment z is marking steps) of a sojourn is

$$D = \sum_{i\ge0} (zC)^{2i} \cdot z \cdot z^{2i+1} = \frac{z^2}{1 - z^4C^2}.$$

Eventually we need

$$\frac{1}{1-D} = 1 + z^2C = 1 + \sum_{n\ge1} \frac{1}{n}\binom{2n-2}{n-1} z^{2n},$$

since we can have an arbitrary number of sojourns. Notice that we also allow the empty path for convenience.

The decomposition also works if we look at paths whose height is restricted by h:

$$zC^{[h-1]}zC^{[h-2]}\dots zC^{[h-2i]}z^{2i+2}, \tag{2}$$

and $2i < h$. But $C^{[h]}$ is a known quantity (1), as mentioned in the introduction. So

$$zC^{[h-1]}zC^{[h-2]}\dots zC^{[h-2i]} = z^{2i}\frac{\lambda_1^{h-2i+1}-\lambda_2^{h-2i+1}}{\lambda_1^{h+1}-\lambda_2^{h+1}}.$$

We replace z^2 by z, which is not essential, but is slightly more pleasant. So we look at

$$z^{i}\frac{\lambda_1^{h-2i+1}-\lambda_2^{h-2i+1}}{\lambda_1^{h+1}-\lambda_2^{h+1}}z^{i+1} = \frac{u^{2i+1}}{(1+u)^{2i+2}}\frac{1-u^{h-2i+1}}{1-u^{h+1}},$$

with the principled substitution $z=\frac{u}{(1+u)^2}$.

And now we need the generating function $D^{[h]}(z)$ of a sojourn with height $\le h$; the enumeration is with respect to semi-length:

$$D^{[h]}(z) = \sum_{0\le i<h/2}\frac{u^{2i+1}}{(1+u)^{2i+2}}\frac{1-u^{h-2i+1}}{1-u^{h+1}}.$$

A routine computation leads to

$$D^{[h]} = \frac{u}{1-u^{h+1}}\left[\frac{1-\left(\frac{u^2}{(1+u)^2}\right)^{\lfloor\frac{h+1}{2}\rfloor}}{1+2u} - \frac{u^{h+1}-u^{[\![h \text{ is even}]\!]}\left(\frac{u^2}{(1+u)^2}\right)^{\lfloor\frac{h+1}{2}\rfloor}}{2u+u^2}\right]. \tag{3}$$

(We use Iverson's notation: $[\![h \text{ is even}]\!] = 1$ iff h is even, and 0 otherwise.)

Theorem 2.1. The generating function of Dyck paths of height $\le h$ with $h \ge 0$ where each run back to the x-axis has odd length, and z marks the semi-length, is given by

$$\frac{1}{1-D^{[h]}},$$

where $D^{[h]}$ is given in (3). The variables z and u are linked via

$$z=\frac{u}{(1+u)^2} \longleftrightarrow u = \frac{1-\sqrt{1-4z}}{2z}-1.$$

Now we discuss approximations, since it seems to be hopeless to derive anything explicit that is of reasonable complexity.

We are going to use singularity analysis of generating functions, as described in the book [7]. In the context of the analysis of the height, we refer to [8, 9, 6, 10, 4, 5].

The general method is that we must look at $u = 1$ ($\leftrightarrow z = \frac{1}{4}$, which is the dominating singularity) and approximate the relevant functions there. Once this is achieved, very general transfer theorems allow to translate this information to the asymptotics of the coefficients. —

We see that the ugly terms with exponent $\lfloor \frac{h+1}{2} \rfloor$ almost cancel out. Not only that, $u^2/(1+u)^2 \sim \frac{1}{4}$ when $u \sim 1$, so these terms don't contribute much, and we will leave them out from now on. (We use $\sim$ here for an expansion in a neighbourhood of $u = 1$, or, equivalently, $z = \frac{1}{4}$. Technicalities about such neighbourhoods can be found in [5].)

We will work with

$$\bar{D}^{[h]} = \frac{u}{1-u^{h+1}}\left[\frac{1}{1+2u} - \frac{u^h}{2+u}\right].$$

Note that the generating function of paths of height $> h$ is given by

$$\begin{aligned}\frac{1}{1-D} - \frac{1}{1-D^{[h]}} &= \frac{1+2u}{1+u} - \frac{1}{1-D^{[h]}} \sim \frac{1+2u}{1+u} - \frac{1}{1-\bar{D}^{[h]}}\\ &= \frac{(1-u)(1+2u)u^{h+1}}{(1+u)(2+u-(1+2u)u^{h+1})} \sim \frac{(1-u)u^{h+1}}{2(1-u^{h+1})}.\end{aligned}$$

The generating function of the *average height* is, apart from normalization, given by

$$E(z) = \sum_{h\ge 0}\left[\frac{1+2u}{1+u} - \frac{1}{1-D^{[h]}}\right] \sim \frac{1-u}{2}\sum_{h\ge 1}\frac{u^h}{1-u^h}.$$

It was computed before (in the Appendix, we sketch how this is done), that, with $u = e^{-t}$,

$$\sum_{h\ge 1}\frac{u^h}{1-u^h} \sim -\frac{\log t}{t} + \frac{\gamma}{t}.$$

Putting things together we get for $z \sim \frac{1}{4} \Leftrightarrow u \sim 1 \Leftrightarrow t \sim 0$

$$E(z) \sim -\frac{1}{2}\log t + K,$$

for an explicit constant K. Now we translate that into an expression in terms of z:

$$t \sim 2\sqrt{1-4z}.$$

Therefore

$$E(z) \sim -\frac{1}{4}\log(1-4z) + K''.$$

Singularity analysis now produces the asymptotic formula

$$[z^n]E(z) \sim -[z^n]\frac{1}{4}\log(1-4z) = \frac{4^{n-1}}{n}.$$

For the normalization, notice that

$$\frac{1}{n}\binom{2n-2}{n-1} \sim 4^{n-1}\sqrt{\pi}n^{-3/2}.$$

Dividing through, we find, that the average height is $\sim \sqrt{\pi n}$. This is the same behaviour as for the ordinary Dyck paths, which is perhaps not too surprising.

Theorem 2.2. The average height of Dyck paths where each run back to the x-axis has odd length, all such paths of length $2n$ being equally likely, is asymptotically given by

$$\sqrt{\pi n}.$$

3. Callan's version

Callan [1] considers Dyck paths were the last return to the x-axis has *even* length, but all the other ones (if any) have *odd* length. For semi-length $n \geq 2$, the enumeration leads to the shifted Catalan numbers as well. For that, we must consider the last sojourn

$$\sum_{i\geq 1}(zC)^{2i-1}\cdot z\cdot z^{2i} = \frac{z^4C}{1-z^4C^2}.$$

We replace, as before, z^2 by z, so we consider the semi-length instead of the number of steps. The overall generating function is then

$$(1+zC)\frac{z^2C}{1-z^2C^2} = zC - z = \sum_{n\geq 2}\frac{1}{n}\binom{2n-2}{n-1}z^n.$$

Now we want to consider again the average height of such paths. We must consider the last sojourn, as it is different from before.

The decomposition (2) also works if we look at path whose height is restricted by h:

$$zC^{[h-1]}zC^{[h-2]}\dots zC^{[h+1-2i]}z^{2i+1},$$

and $2i \leq h$. So

$$zC^{[h-1]}zC^{[h-2]}\dots zC^{[h+1-2i]} = z^{2i-1}\frac{\lambda_1^{h-2i+2}-\lambda_2^{h-2i+2}}{\lambda_1^{h+1}-\lambda_2^{h+1}}.$$

We replace again z^2 by z. So we look at

$$\frac{\lambda_1^{h-2i+2}-\lambda_2^{h-2i+2}}{\lambda_1^{h+1}-\lambda_2^{h+1}}z^{2i} = \frac{u^{2i}}{(1+u)^{2i+1}}\frac{1-u^{h-2i+2}}{1-u^{h+1}}.$$

And now we need

$$G^{[h]} = \sum_{1\leq i\leq h/2}\frac{u^{2i}}{(1+u)^{2i+1}}\frac{1-u^{h-2i+2}}{1-u^{h+1}}.$$

A routine computation leads to

$$G^{[h]} = \frac{1+u}{1-u^{h+1}}\left[\frac{1-\left(\frac{u^2}{(1+u)^2}\right)^{\lfloor\frac{h}{2}\rfloor}}{1+2u} - \frac{u^{h+2}-u^{2+[\![h \text{ is odd}]\!]}\left(\frac{u^2}{(1+u)^2}\right)^{\lfloor\frac{h}{2}\rfloor}}{2u+u^2}\right] - \frac{1}{1+u}\frac{1-u^{h+2}}{1-u^{h+1}}.$$

We approximate as before:

$$\bar{G}^{[h]} = \frac{1}{(1+u)(1-u^{h+1})}\left[\frac{u^2}{1+2u} - \frac{u^{h+1}}{2+u}\right].$$

Notice that without height restriction this leads to

$$\frac{\bar{G}^{[\infty]}}{1-\bar{D}^{[\infty]}} = \frac{1+2u}{1+u}\frac{u^2}{(1+u)(1+2u)} = \frac{u^2}{(1+u)^2}.$$

The generating function of paths à la Callan of height $> h$ is approximated by

$$\begin{aligned}\frac{\bar{G}^{[\infty]}}{1-\bar{D}^{[\infty]}} - \frac{\bar{G}^{[h]}}{1-\bar{D}^{[h]}} &= \frac{u^2}{(1+u)^2} - \frac{\bar{G}^{[h]}}{1-\bar{D}^{[h]}}\\ &= \frac{(1-u)(1+2u)u^{h+1}}{(1+u)(2+u-(1+2u)u^{h+1})} \sim \frac{(1-u)u^{h+1}}{2(1-u^{h+1})}.\end{aligned}$$

with

$$\bar{D}^{[h]} = \frac{u}{1-u^{h+1}}\left[\frac{1}{1+2u} - \frac{u^h}{2+u}\right].$$

Since we also obtained this approximation for the paths à la Stanley, the average height is here $\sim \sqrt{\pi n}$ as well.

Theorem 3.1. The average height of Dyck paths where each run back to the x-axis has odd length, except for the last one, whose run back to the x-axis has even length, all such paths of length $2n$ being equally likely, is asymptotically given by $\sqrt{\pi n}$.

Appendix

In order to use the powerful transfer machinery from [5], we need to establish a local expansion. This short description is borrowed from [3] and [8].

The behaviour of

$$f(t) := \sum_{h\ge 1} \frac{e^{-ht}}{1-e^{-ht}}$$

for $t \to 0$ is found using the *Mellin transform* [3]: One rewrites it as

$$f(t) = \sum_{h,k\ge 1} e^{-hkt}$$

notices that it is a *harmonic sum*, and transforms it:

$$f^*(s) = \sum_{h,k\ge 1} (hk)^{-s}\Gamma(s) = \zeta^2(s)\Gamma(s).$$

Now there is the *inversion formula*:

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

Shifting the line of integration to the left and collecting residues gives the asymptotic behaviour of $f(t)$ as $t \to 0$. For instance, the first encountered pole is at $s = 1$, and the residue evaluates to

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

where γ is Euler's constant.

References

[1] D. Callan. The 136th manifestation of C_n. *arXiv:math*, 0511010 (3 pages), 2005.

[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] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[4] P. Flajolet and A. Odlyzko. The average height of binary trees and other simple trees. *Journal of Computer and System Science*, 25:171–213, 1982.

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

[6] P. Flajolet and H. Prodinger. Register allocation for unary-binary trees. *SIAM J. Comput.*, 15:629–640, 1986.

[7] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2008.

[8] H. Prodinger. The height of planted plane trees revisited. *Ars Combinatoria*, 16 B:51–55, 1983.

[9] H. Prodinger. Some analytic techniques for the investigation of the asymptotic behaviour of tree parameters. *EATCS Bulletin*, 47:180–199, 1992.

[10] H. Prodinger. The location of the first maximum in the first sojourn of a Dyck path. *Discrete Mathematics and Theoretical Computer Science*, 10:125–134, 2008.

[11] R. Stanley. *Enumerative combinatorics. Vol. 2*. Cambridge University Press, Cambridge, 1999.