Using the KISS method to Count Restricted Dyck Paths

Shalosh B. EKHAD and Doron ZEILBERGER

In memory of our hero Richard Guy (1916-2020)

Abstract. In his classic essay "The Strong Law of Small Numbers", Richard Guy gave numerous *cautionary tales* where one can't 'jump to conclusions' from the first few terms of a sequence. But if you are cautious enough you can find many families of enumeration problems where it is very safe to deduce the general pattern from the first few cases, obviating the need for either the human or the computer to think too hard, and by using the **'Keep It Simple Stupid'** principle (**KISS** for short), one can easily derive many deep enumeration theorems by doing exactly what Richard Guy told us **not** to do, computing the first few terms of the sequence and deducing the formula for the general term. We admit that often 'few' should be replaced by 'quite a few', but it is still much less painful than trying to figure out the intricate combinatorial structure by 'conceptual' means.

The Maple package

This article is accompanied by the Maple package

`Dyck.txt` available, along with ample input and output files from the front of thos article

`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/kiss.html` .

Enumerating Dyck Paths

Recall that a **Dyck path** of **semi-length** n is a walk in the 2-dimensional plane, from the origin $(0,0)$ to $(2n,0)$ with **atomic steps** $U := (1,1)$ and $D := (1,-1)$ that **never** goes below the $x-$axis, i.e. that always stays in $y \leq 0$.

For example, the five Dyck paths of semi-length 3 are

$$UUUDDD \quad , \quad UUDUDD \quad , \quad UUDDUD \quad , \quad UDUUDD \quad , \quad UDUDUD \quad .$$

The number of Dyck paths of semi-length n is famously the **Catalan number** $\frac{(2n)!}{n!(n+1)!}$ the most popular, and important sequence in enumerative combinators (with no offense to Fibonacci), the subject of a whole book by Guru Richard Stanely [St].

When we searched (on May 14, 2020) our favorite website, the OEIS [Sl] for the phrase "Dyck paths" we got back 1095 hits. Of course we did not have patience to read all of them but a random browsing revealed that they enumerate Dyck paths with various **restrictions**. We will soon see how to quickly enumerate 'infinitely' many such classes, but first let's recall one of the many proofs of the fact that the number of Dyck paths is indeed the venerable OEIS sequence A108, $\frac{(2n)!}{n!(n+1)!}$.

Let $a(n)$ be the desired number, i.e. the number of Dyck paths of semi-length n, and cosider the **ordinary generating function**

$$f(x) \,=\, \sum_{n=0}^{\infty} a(n)\, x^n \quad ,$$

which is the **weight-enumerator** of the set of all Dyck path, with $weight(P) := x^{SemiLength(P)}$.

It is readily seen that any Dyck path P is either empty or can be written **uniquely** (i.e. **unambiguously**) as $P = U\,P_1\,D\,P_2$, where P_1 and P_2 are shorter Dyck paths, and vica versa, for any Dyck paths P_1, P_2, $U\,P_1\,D\,P_2$ is a Dyck path on its own right. Let $\mathcal{P}$ be the **totality** of all Dyck paths, then we have the **grammar**

$$\mathcal{P} \,=\, EmptyPath \cup U\,\mathcal{P}\,D\,P \quad .$$

Applying the **weight** functional we get

$$f(x) = 1 + xf(x)^2 \quad .$$

To deduce that $a(n) = (2n)!/(n!(n+1)!)$, you can, *inter alia*

• (i) Solve the quadratic and use Newoton's binomial theorem .

• (ii) Differentiate both sides getting a differential equation for $f(x)$ that translates to a first-order recurrence for $a(n)$.

• (iii) Use Largrange Inversion (see [Z1] for a brief and lucid exposition).

How it All Started: Vladimir Retakh's Question

Volodia Retakh asked whether there is a proof of the fact that the number of Dyck paths of semi-length n such that the height all of all peaks is either 1 or even is given by the also famous Motzkin numbers (OEIS sequence A1006), whose generating function satisfies

$$f(x) = 1 \,+\, xf(x) \,+\, x^2 f(x)^2 \quad .$$

We first tried to find a 'conceptual' proof generalizing the above proof, and indeed we found one, by adapting the above classical proof enumerating all Dyck paths.

Let $\mathcal{P}_1$ be the set of Dyck paths whose peak-heights are never in $\{3, 5, 7, \ldots\}$, and let $f_1 = f_1(x)$ be its weight enumerator.

Let P_1 be any member of $\mathcal{P}_1$ then, it is either empty, or we can write

$$P_1 \,=\, U\,P_2\,D\,P_1' \quad ,$$

where $P_1' \in \mathcal{P}_1$ but P_2 has the property that none of its peak-heights is in $\{2, 4, 6, \ldots\}$. Let $\mathcal{P}_2$ be the set of such Dyck paths.

we can write the 'grammar'

$$\mathcal{P}_1 \ = \ EmptyPath \cup U\,\mathcal{P}_2\,D\,P_1 \quad .$$

Let $f_2 = f_2(x)$ be the weight-enumerator of $\mathcal{P}_2$.

Taking weights above, we have the equation

$$f_1 = 1 + x f_2 f_1 \quad .$$

Alas, now we have to put-up with $\mathcal{P}_2$ and f_2. Let P_2 be any member of $\mathcal{P}_2$. Then either it is empty, or it can be written as

$$U\,P_3\,D\,P_2' \quad ,$$

where $P_2' \in \mathcal{P}_2$ but P_3 is a Dyck path whose peak-heights are never in $\{1,3,5,7,\ldots\}$.

Let $\mathcal{P}_3$ be the set of such Dyck paths. We have the grammar

$$\mathcal{P}_2 \ = \ EmptyPath \cup U\,\mathcal{P}_3\,D\,P_1' \quad .$$

Let $f_3 = f_3(x)$ be the weight-enumerator of $\mathcal{P}_3$.

Taking weights we have another equation

$$f_2 = 1 + x f_3 f_2 \quad .$$

It looks like we are doomed to have **infinite regress**, but let's try one more time.

Let P_3 be any member of $\mathcal{P}_3$. It is either empty, or We can write

$$P_3 \ = \ U\,P_4\,DP_3'$$

where $P_3' \in \mathcal{P}_3$ and P_4 is a **non-empty** path avoiding peak-heights in $\{2,4,6,8,\ldots$. But this looks familiar, so the set of P_4 is really $\mathcal{P}_2 \backslash \{EmptyPath\}$, and we have the grammar

$$\mathcal{P}_3 \ = \ EmptyPath \cup U\,(\mathcal{P}_2 \backslash \{EmptyPath\})\,D\,\mathcal{P}_3 \quad .$$

Taking weight, we get

$$f_3 = 1 + x(f_2 - 1) f_3 \quad .$$

We have a system of three algebrai equations

$$\{ f_1 = 1 + x f_2 f_1 \quad , \quad f_2 = 1 + x f_3 f_2 \quad , \quad f_3 = 1 + x(f_2 - 1) f_3 \quad \} \quad ,$$

in the unknowns

$$\{f_1, f_2, f_3\} \quad .$$

Eliminiating f_2, f_3 yields the following algebraic equation for our object of desire f_1.

$$f_1(x) = 1 + x f_1(x) + x^2 f_1(x)^2 \quad ,$$

proving Volodia's Retakh's claim.

The KISS way

Now that we know that such an argument **exists**, and that the desired generating function $f(x)$, satisfies an algebraic equation of the form $P(x, f(x)) = 0$ for some bivariate polynomial $P(x, y)$, why not **keep it simple**, and rather than wrecking our brains (either human or machines) we can collect sufficiently many terms of the desired sequence, and then use Maple's command `gfun[listtoalgeq]` (or our own home-made version) to **guess** that polynomaial $P(x, f(x))$.

Numerical Dynamical Programming to the rescue

Suppose that we don't know anything, and want to compute the number of Dyck paths of semi-length n, i.e. the number of walks using the fundamental steps $U = (1, 1)$ and $D = (1, -1)$. A natural approach is to consider the more general quantity $d(m, k)$, the number of walks from $(0, 0)$ to (m, k) staying weakly above the x-axis and **ending at a down step**. If the length of that downward run is r, then the pervious peak was at $(m - r, k + r)$, and we need to introduce the auxiliary function $u(m, k)$ the number of such paths thatend at (m, k) and end at an up step.

We have

$$d(m, k) = \sum_{r=1}^{m} u(m - r, k + r)$$

Analously

$$u(m, k) = \sum_{r=1}^{m} d(m - r, k - r)$$

Of course we have the obvious initial conditions $d(0, 0) = 1$, conditions $d(m, k) = 0$ and $u(m, k)$ if $k > m$.

Here is the short Maple code that does it

```
u:=proc(m,k) local r:  option remember:  if k>m then RETURN(0):  fi:  if m=0 then 0:
else add(d(m-r,k-r),r=1..k):  fi:  end:

d:=proc(m,k) local r:  option remember:  if k>m then RETURN(0):  fi:

if m=0 then if k=0 then RETURN(1):  else RETURN(0):  fi:  fi:  add(u(m-r,k+r),r=1..m):
end:
```

To the the desried sequence enumerating all Dyck paths of semi-length n for n from 1 to N, in other words $\{d(2n,0)\}_{n=1}^{N}$ for any desired N we type

```
seq(d(2*n,0),n=1..N);:
```

Now, recall that we had to work much harder to, *logically* and *conceptually* to find the algebraic equation for the Dyck paths considered by Volodia Retakh. To get the the analogous sequence we only need to change the program by **one line**. Let's call the analogous quantities $u_1(m,k)$ and $d_1(m,k)$.

```
u1:=proc(m,k) local r:  option remember:  if ( k>m or k>1 and k mod 2=1) then RETURN(0):
fi:  if m=0 then 0:  else add(d1(m-r,k-r),r=1..k):  fi:  end:
```

```
d1:=proc(m,k) local r:  option remember:  if k>m then RETURN(0):  fi:  if m=0 then
if k=0 then RETURN(1):  else RETURN(0):  fi:  fi:  add(u1(m-r,k+r),r=1..m):  end:
```

In other words, just declaring that $u_1(m,k)=0$ if the elevation k is an odd integer larger than 1.

Typing

```
seq(d1(2*n,0),n=1..N);
```

will let us get, very fast the first N terms, that would enable us to guess the algebaric equation satisfied by the the generating function, that we can justify, *a posteriori* since we know that it **exists**, saving us the mental agony of doing it logically, by figuring out the intricate 'grammmer'.

The general case

Since it is so easy to tweak the numerical dynamical programming procedure, why not be as general as can be? Let A, B, C, D be *arbitrary* sets of positive integers, either finite sets, or infinite sets (like in Retakh's case) that are arithmetical progressions (or unions thereof). We are interested in counting Dyck paths that obey the following restrictions

- No peak can be of a height that belongs to A
- No valley can be of a height that belongs to B
- No upward run can be of a height that belongs to C
- No downward run can be of a height that belongs to D

Then we declare that $u(m,k)=0$ if $k \in A$ and $d(m,k)=0$ if $k \in B$ and otherwise

$$d(m,k) \;=\; \sum_{\substack{1 \leq r \leq m \\ r \notin C}} u(m-r,k+r)$$

Analously

$$u(m,k) \; = \sum_{\substack{1 \le r \le m \\ r \notin D}} d(m-r, k-r) \quad .$$

Then we get, very fast, sufficiently many terms to guess an algebraic equation, by finding $\{d(2n,0)\}_{n=1}^{N}$.

Guessing linear recurrences

It is well-known (see [KP], Theorem 6.1) that if $f(x)$ is an algebaric formal power series (like in our case), then it satisfies a linear differential equation with polynomial coefficients (i.e. it is D-finite, and hence its sequence of coefficients, $a(n)$ satisfies a linear recurrence equation with polynomial coefficients, i.e. is P-recursive. While there are easy algorithms for finding these, they do not always give the minimal recurrence, and once again, let's keep it simple! Just guess such a recurrence using *undetermined coefficients*, and we are guaranteed by the background 'general nonsense' that everything is rigorously proved, and we don't have to worry about Richard Guy's Strong Law of Small Numbers.

The Maple package `Dyck.txt`

Everything here, and more is implemented in the Maple package available from the front of this article

`https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/kiss.html` .

There you would also find long web-books with many deep enumeration theorems. Let us illustrate just one **random** example.

Typing `Theorem({1},{},{2},{1},60,P,x,n,a,20,1000);` gives

Sample Theorem: Let $a(n)$ be the number of Dyck paths of semi-length n obeying the following restrictions The height of a peak can not belong to $\{1\}$, no upward-run can belong to $\{2\}$, and no downward-run can belong to $\{1\}$, then the generating function

$$f(x) \; := \sum_{n=1}^{\infty} a(n) x^n \quad ,$$

satisfies the algebraic equation

$$1 + \left(x^4 + x^3 + x^2 + x\right)\left(P\left(x\right)\right)^2 + \left(-x^2 - x - 1\right) P\left(x\right) = 0 \quad ,$$

and the sequence $a(n)$ satisfies the following linear recurrence

$$a\left(n\right) = \frac{\left(n-2\right) a\left(n-1\right)}{n+1} + 2\,\frac{\left(n-2\right) a\left(n-2\right)}{n+1} + \frac{\left(4\,n-11\right) a\left(n-3\right)}{n+1}$$

$$+\frac{\left(8\,n-25\right) a\left(n-4\right)}{n+1} + 6\,\frac{\left(n-4\right) a\left(n-5\right)}{n+1} + \frac{\left(5\,n-22\right) a\left(n-6\right)}{n+1} + 3\,\frac{\left(n-5\right) a\left(n-7\right)}{n+1} \quad ,$$

subject to the initial conditions

$$a(1) = 0, a(2) = 0, a(3) = 1, a(4) = 2, a(5) = 3, a(6) = 7, a(7) = 17 \quad .$$

Just for fun

$$a(1000) =$$

$$50324963656379550676833478709504097109157017645222822767746751572436038025828662982112312106530063940708331389763348225007597598917857819768291846376430383787883303304173561096755566724223651076112624984594489328879621306432162789206819579174880699300406347337450543154482201080856180030329544504922488122762391456888251217173587699286352230512754854626916997290850041113153875267233420342398420706028366390785696361604283302005101154378 \quad .$$

Conclusion

While Richard Guy's cautionary tales [G1][G2] should be taken seriously, it is often safe to ignore them. This is the case here, where *keeping it simple* is so much more efficient, and painless, than doing it the 'thinking way'.

References

[G1] Richard K. Guy, *The strong law of small numbers*, Amer. Math. Monthly, **95** (1988), 697–712.

[G2] Richard K. Guy, *The second strong law of small numbers*, Math. Mag. **63** (1990), 3–20.

[KP] Manuel Kauers and Peter Paule, *"The Concrete Tetrahedron"*, Springer, 2011.

[PW] P. Peart and W.-J. Woan, *Dyck Paths With No Peaks at Height k*, J. Integer Sequences, **4** (2001), #01.1.3.

[Sl] Neil Sloane, *The On-Line Encyclopedia of Integer Sequences*, `https://oeis.org` .

[St] Richard Stanely, *" Catalan Numbers"*, Cambridge University Press, 2015.

[Z1] Doron Zeilberger, *Lagrange Inversion Without Tears (Analysis) (based on Henrici)*, The Personal Journal of Shalosh B. Ekhad and Doron Zeilberger,
`https://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/lag.html` .

Shalosh B. Ekhad, c/o D. Zeilberger, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA.
Email: `ShaloshBEkhad at gmail dot com` .

Doron Zeilberger, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA.
Email: `DoronZeil at gmail dot com` .

Exclusively published in the Personal Journal of Shalosh B. Ekhad and Doron Zeilberger and arxiv.org .

Written: May 20, 2020.