GENERATING FUNCTIONS AND RANDOM WALKS

SIMON RUBINSTEIN-SALZEDO

1. An illustrative example

Before we start studying generating functions properly, let us look an example of their use. Consider the numbers a_n, defined by the recurrence relation

$$a_{n+1} = 2a_n + 1, \qquad a_0 = 0.$$

We would like a general formula for a_n. We might try working out a few terms to see if we can detect a pattern:

$$a_0 = 0, \quad a_1 = 1, \quad a_2 = 3, \quad a_3 = 7, \quad a_4 = 15, \quad a_5 = 31.$$

If we're lucky, we notice that it looks like $a_n = 2^n - 1$. Then, we can try to prove this formula with induction.

But that only worked because we got lucky and happened to notice and guess the pattern. How could we figure out a general formula without any guesswork?

Let's solve the above example, without any guesswork. What we will do is to set up an object that carries all the information about the recurrence, in an easy-to-use format. To do this, let us define an object

$$A(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots.$$

We want to use the recurrence relation in order to learn more about $A(x)$. So, let us take the relation

$$a_{n+1} = 2a_n + 1,$$

multiply by x^n to get

$$a_{n+1} x^n = 2a_n x^n + x^n,$$

and then sum from $n = 0$ to ∞:

$$\sum_{n=0}^{\infty} a_{n+1} x^n = 2\sum_{n=0}^{\infty} a_n x^n + \sum_{n=0}^{\infty} x^n.$$

Some of the terms in this sum are familiar objects: the first term on the right side is $2A(x)$, and the second term is a geometric series, which sums to $\frac{1}{1-x}$.[1] But what about the left side?

Date: 20110.11.09.

[1] In case you're worried about convergence, don't be. When we use generating functions, we like to trouble ourselves as little as possible with issues of convergence. This will only be an issue if we ever try to plug something in for x. Sometimes we will do that, and then we will need to be careful, but for now it will not be necessary.

We can express that in terms of $A(x)$ as well, as $\frac{A(x)-a_0}{x}$. Since $a_0 = 0$, this means that the left side is $\frac{A(x)}{x}$. Thus, we have

$$\frac{A(x)}{x} = 2A(x) + \frac{1}{1-x}.$$

We can now solve for $A(x)$ to get

$$A(x) = \frac{x}{(1-x)(1-2x)}.$$

Now, we've figured out what $A(x)$ is. How do we recover the a_n's in a useful form? The usual thing to do at this point is to use partial fractions, i.e., to try to write

$$A(x) = \frac{c}{1-x} + \frac{d}{1-2x},$$

for some constants c and d. There are many ways to find c and d, but one of the easiest is to take the above equation and clear denominators, to get

$$x = c(1-2x) + d(1-x) \tag{1}$$

and plug in nice values for x like 1 and $\frac{1}{2}$, so that we get $c = -1$ and $d = 1$, so that

$$A(x) = \frac{1}{1-2x} - \frac{1}{1-x}.$$ [2]

Finally, we have expressed $A(x)$ as a sum of two terms, each of which is the sum of a geometric series. So, we unsum them to get

$$A(x) = \sum_{n=0}^{\infty} 2^n x^n - \sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} (2^n - 1)x^n = \sum_{n=0}^{\infty} a_n x^n.$$

Equating coefficients of the x^n terms,[3] we find that $a_n = 2^n - 1$.

2. Generating functions

The power series $A(x) = \sum_{n=0}^{\infty} a_n x^n$ that we used above is called the **generating function** for the sequence (a_n). As we saw above, they are really useful for solving problems involving recurrences. For example, by slightly modifying the above approach, we can prove **Binet's formula** for Fibonacci numbers:

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right). \tag{2}$$

(It's a little bit annoying, since we have to do partial fractions with terms that have $\sqrt{5}$'s in them, but a slightly uglier computation is the main extra difficulty we encounter. I wouldn't want to do it on the board with people watching me.)

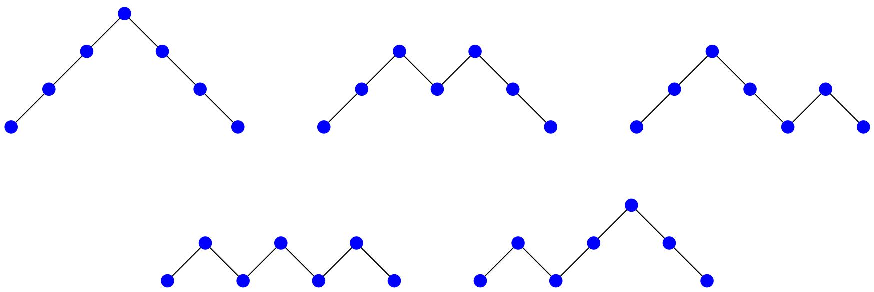
---

[2] I hear you complaining that I promised not to plug things in for x since we cavalierly disregarded convergence issues earlier, and then I did so anyway. If this bothers you, feel free to solve for c and d by equating constant and linear terms in equation (1).

[3] We can do this, since two such power series are equal if and only if all their coefficients are equal.

However, some recurrences have a more complicated form than the linear recurrences defining a_n above or the Fibonacci numbers. Many of these come from sequences of combinatorial interest.

Example. The n^{th} **Catalan number** C_n is the number of paths from $(0,0)$ to $(2n,0)$ consisting of n steps of the form $(1,1)$ and n steps of the form $(1,-1)$ that never pass below the x-axis.[4] Such paths are called Dyck paths. Here are all the Dyck paths with $n = 3$:



The first few Catalan numbers are

$$C_0 = 1, \quad C_1 = 1, \quad C_2 = 2, \quad C_3 = 5, \quad C_4 = 14, \quad C_5 = 42.$$

The Catalan numbers have a nice recurrence relation:

Theorem 1. *The Catalan numbers satisfy the recurrence relation*

$$C_{n+1} = C_0C_n + C_1C_{n-1} + \cdots + C_{n-1}C_1 + C_nC_0 = \sum_{i=0}^{n} C_iC_{n-i}.$$

Based on this recurrence, we can compute the generating function $C(x) = \sum_{n=0}^{\infty} C_n x^n$ for the Catalan numbers. But before we do that, we should investigate what happens when we add and multiply generating functions:

Theorem 2. *Let $A(x) = \sum_{n=0}^{\infty} a_n x^n$ and $B(x) = \sum_{n=0}^{\infty} b_n x^n$ be two generating functions. Then*

$$A(x) + B(x) = \sum_{n=0}^{\infty} (a_n + b_n)x^n,$$

$$A(x)B(x) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right) x^n.$$

Let us now work out the generating function for the Catalan numbers, based on the recurrence relation: doing the usual trick of multiplying by x^n and summing over n gives

$$\sum_{n=0}^{\infty} C_{n+1}x^n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} C_kC_{n-k} \right) x^n.$$

[4]This is very far from the only way of defining the Catalan numbers. Richard Stanley, in his book *Enumerative Combinatorics, Volume 2*, gives 66 interpretations of the Catalan numbers in his (in)famous Exercise 6.19. He now lists several hundreds more on his website, at `http://www-math.mit.edu/~rstan/ec/catadd.pdf`, and also in his new book, simply titled *Catalan Numbers*.

The left side is $\frac{C(x)-1}{x}$, and the right side is $C(x)^2$. Hence we have

$$\frac{C(x)-1}{x} = C(x)^2.$$

Solving for $C(x)$ using the quadratic formula gives

$$C(x) = \frac{1 \pm \sqrt{1-4x}}{2x}.$$

Ah, but we're not quite done, because we don't know whether it's the positive or negative root. But we can figure this out, in one of two ways:

(1) The honest way: We can do this by computing some terms and seeing which one is right. Using the positive root and applying the binomial theorem, we would get $C(x) = \frac{1}{x} - 1 - x - 2x^2 + \cdots$, so we would have $C_{-1} = 1$, $C_0 = -1$, and so forth. But this is wrong. If we instead try the negative root, we would get $C(x) = 1 + x + 2x^2 + 5x^3 + \cdots$, which is right. Hence we have

$$C(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

(2) Cheating: If we plug $x = 0$ into $C(x)$, we should get

$$C(0) = \sum_{n=0}^{\infty} C_n 0^n = C_0 = 1.$$

So, let's try it. If we try it with the positive root, we get $C(0) = \frac{2}{0}$, which is disastrous. On the other hand, if we try it with the negative root, we get $C(0) = \frac{0}{0}$. Much better!

Generating functions are very powerful when tackling combinatorial problems. Indeed, there is an enormous amount to be said about generating functions and combinatorics, but I won't say it here, because I'd like to move on to random walks and probability. (See the list of references at the end.)

3. Random walks

George Pólya was the first person to study random walks. He became interested in them when he repeatedly ran into the same couple while taking his afternoon walks, despite choosing different routes. Embarrassed that this coincidence might be construed as snooping, he wondered whether this would happen by chance, and the subject of random walks was born.

Let us set up the problem. A random walker starts at the origin in $\mathbb{R}^d$. At every second, the walker moves one unit in a direction parallel to one of the coordinate axes. (So, the walker takes a step of the form $(0, 0, \ldots, 0, \pm 1, 0, \ldots, 0)$, each with equal probability.) Since there are $2d$ possible steps, each one occurs with probability $\frac{1}{2d}$.

The problem is to determine the probability that the walker eventually returns to the origin.

It will not be easy to treat all dimensions at once, so we shall begin with a 1-dimensional random walk. Here, the walker moves one step to the right with probability $1/2$, and one step to the left with probability $1/2$. Let us define p_n as the probability that the walker is at the origin after time n, and let f_n be the probability that the walker has returned to the origin for the first time at time n. Note that, if n is odd, then $p_n = f_n = 0$.

There are two ways of proceeding.

First method: Let's work out what f_{2n} is. In order for the walker to have returned to the origin for the first time at time $2n$, the walker must have stayed entirely positive or entirely negative before that. Let us assume that the walker stayed entirely in the positives. In order for this to happen, the walker must have made one positive step at the beginning, then done a Dyck path of length $2n-2$, then taken one negative step at the end. The number of ways for this to happen is C_{n-1}, and there are 2^{2n} possible paths in total. Hence, the probability that this happens is $\frac{C_{n-1}}{2^{2n}}$. But since the walker could have stayed entirely negative as well, we have to multiply this number by 2. So, the probability that the walker returns for the first time at time $2n$ is $f_{2n} = \frac{C_{n-1}}{2^{2n-1}}$.

What is the probability that the walker ever returns to the origin? Well, in any walk that eventually returns to the origin, there must be some first return, so the probability of a return is simply the sum of the f_n's: $\sum_{n=1}^{\infty} f_{2n}$. In terms of the Catalan numbers, this is

$$\sum_{n=1}^{\infty} \frac{C_{n-1}}{2^{2n-1}} = \sum_{n=0}^{\infty} \frac{C_n}{2^{2n+1}}.$$

From the generating function for the Catalan numbers, we can work this out as a number, since it's almost what we get when we plug in $x = \frac{1}{4}$ in $C(x)$:

$$C(1/4) = \sum_{n=0}^{\infty} \frac{C_n}{4^n}.$$

Hence the desired probability is $\frac{1}{2}C(1/4) = 1$. Thus we have proved:

Theorem 3. *In a one-dimensional random walk, the walker returns to the origin with probability 1.*

Second method: This method involves meshing the f's and the p's better. To do this, we find a recurrence relation involving the f's and p's: if we're at the origin after $2n$ steps, then we must have had a first return to the origin at $2k$ steps, for some $k \leq n$, and then made a path that goes back to the origin in $2n-2k$ steps. Hence, we have

$$p_{2n} = f_0 p_{2n} + f_2 p_{2n-2} + \cdots + f_{2n-2} p_2 + f_{2n} p_0,$$

when $n \geq 1$. (When $n = 0$, this relation doesn't work, because there is no first return.) Let us set $F(x) = \sum_{n=0}^{\infty} f_{2n} x^n$ and $P(x) = \sum_{n=0}^{\infty} p_{2n} x^n$. We can find a relation between F and P by using the recurrence: multiply by x^n and sum over $n \geq 1$ to get

$$\sum_{n=1}^{\infty} p_{2n} x^n = \sum_{n=1}^{\infty} \left(\sum_{k=0}^{n} f_{2k} p_{2n-2k} \right) x^n.$$

The left side is $P(x) - p_0 = P(x) - 1$, and the right side is $F(x)P(x) - f_0 p_0 = F(x)P(x)$ (since $f_0 = 0$). Thus we have

$$P(x) - 1 = F(x)P(x),$$

or

$$F(x) = 1 - \frac{1}{P(x)}.$$

Now, we compute $P(x)$. To compute p_{2n}, we note that, in order to be back at the origin after $2n$ steps, we must have taken n positive steps and n negative steps, and these can be arranged in any order. Each possible sequence of n positive and n negative steps occurs with probability $1/4^n$, so

$$p_{2n} = \frac{1}{4^n}\binom{2n}{n}.$$

Hence

$$P(x) = \sum_{n=0}^{\infty} \binom{2n}{n} \left(\frac{x}{4}\right)^n.$$

By the binomial theorem and some manipulations, we find that this is equal to $\frac{1}{\sqrt{1-x}}$. Hence

$$F(x) = 1 - \sqrt{1-x},$$

so $F(1) = 1$, once again.

So, we have now shown that with probability 1, the walker returns home. But how long does it take, on average?

Let us determine how to work out the average value of a random variable. It is easiest to understand what this means by first doing an example: the average value of a die roll. The average value of a die roll is $\frac{7}{2} = \frac{1}{6}(1+2+3+4+5+6)$. There is a probability of 1/6 of each result, so we multiply the result by the probability and sum over all the possibilities. Hence, in this case, the average time it takes to get back to the origin for the first time is

$$\sum_{n=1}^{\infty} 2n f_{2n}.$$

There is a standard trick to evaluate such sums: take a derivative! If $A(x) = \sum_{n=0}^{\infty} a_n x^n$, then $A'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1}$. In particular, if we have a random variable that takes on only nonnegative integer values, with a_n being the probability of getting n (so that in particular $\sum_{n=0}^{\infty} a_n = 1$), then the expected or mean value of the random variable is $A'(1)$.

In our case,

$$F'(x) = \sum_{n=0}^{\infty} n f_{2n} x^{n-1}.$$

Hence, the average time to the first return home is $2F'(1)$, with the 2 coming because we multiplied f_{2n} by x^n rather than x^{2n}. We can compute this directly, since $F(x) = 1-\sqrt{1-x}$; we get $F'(x) = \frac{1}{2\sqrt{1-x}}$, so it looks like $2F'(1)$ is infinite. This means that, even though the walker returns in finite time with probability 1, the *average* amount of time it takes is infinite!

4. Higher-dimensional random walks

Now that we've taken care of the one-dimensional random walk, let's move on to higher dimensions. There are surprises waiting for us here!

Let's do the 2-dimensional case. Here, there are four possible steps: $(1, 0)$, $(-1, 0)$, $(0, 1)$, and $(0, -1)$. We define $f_{2n}^{(2)}$ and $p_{2n}^{(2)}$ to be the probability of a first return or return (respectively) to the origin at time $2n$.

We will compute $f_{2n}^{(2)}$ and $p_{2n}^{(2)}$ by following Method 2 in the one-dimensional case. We start by computing $p_{2n}^{(2)}$. In order to return to the origin after $2n$ steps, we must take k steps right, k steps left, $n-k$ steps up, and $n-k$ steps down, for some k with $0 \le k \le n$. The probability of taking k steps right, k steps left, $n-k$ steps up, and $n-k$ steps down is

$$\frac{1}{4^{2n}} \frac{(2n)!}{k!^2(n-k)!^2} = \frac{1}{4^{2n}} \frac{(2n)!n!^2}{n!^2k!^2(n-k)!^2} = \frac{1}{4^{2n}} \binom{2n}{n} \binom{n}{k}^2.$$

Thus

$$p_{2n}^{(2)} = \sum_{k=0}^{n} \frac{1}{4^{2n}} \binom{2n}{n} \binom{n}{k}^2 = \frac{1}{4^{2n}} \binom{2n}{n}^2,$$

where the last equality follows from problem 9.

Now, the recurrence relating the f's and the p's still holds, so if we let $F^{(2)}$ be the generating function for the f's, and $P^{(2)}$ the generating function for the p's, we have once again

$$F^{(2)}(x) = 1 - \frac{1}{P^{(2)}(x)}.$$

This time it won't be so easy to evaluate $P^{(2)}(x)$ exactly, so we'll have to resort to an approximation. But what we care about is whether we will necessarily return. That is, is $F^{(2)}(1) = 1$? This will happen if and only if $P^{(2)}(1)$ is infinite. So, let us try to work that out. We do that by applying Stirling's formula (problem 11), which says that

$$n! \approx \sqrt{2\pi n}\, n^n e^{-n}.$$

We can use this to approximate $\binom{2n}{n} = \frac{(2n)!}{n!^2}$ as well: we get

$$\binom{2n}{n} \approx \frac{2^{2n}}{\sqrt{\pi n}}.$$

Hence,

$$p_{2n}^{(2)} \approx \frac{1}{4^{2n}} \frac{(2^{2n})^2}{\pi n} = \frac{1}{\pi n}.$$

Now,

$$P_{2n}^{(2)}(1) = \sum_{n=0}^{\infty} p_{2n}^{(2)} \approx \sum_{n=1}^{\infty} \frac{1}{\pi n},$$

which diverges. Hence $P_{2n}^{(2)}(1)$ also diverges, so $F_{(2n)}^{(2)}(1) = 1$, so again, the walker always returns to home.

We might now be ready to conjecture that the walker always returns home. But there is an important rule in conjecture-making: whenever you're about to make a conjecture, try one more case.[5] So, let's do the 3-dimensional case.

The setup in three dimensions is very similar. We now compute $p_{2n}^{(3)}$, and we find it to be

$$p_{2n}^{(3)} = \frac{1}{6^{2n}} \sum_{j,k} \frac{(2n)!}{j!^2 k!^2 (n-j-k)!^2} = \frac{1}{2^{2n}} \binom{2n}{n} \sum_{j,k} \left(\frac{1}{3^n} \frac{n!}{j!k!(n-j-k)!} \right)^2.$$

This is hard to evaluate exactly, but we can bound it: the term in parentheses is maximized when $j, k, n-j-k$ are all around $n/3$, so we can replace all the terms with what we get when we plug in $n/3$ for j and k. (And it's no particular harm to assume that n is divisible by 3, but we won't be that careful, since we're just going to use Stirling's formula anyway.) When we do this, we find that the largest value M of $\frac{n!}{j!k!(n-j-k)!}$ is roughly $\frac{c}{n}$, for some constant c.

We now bound $p_{2n}^{(3)}$:

$$\begin{aligned} p_{2n}^{(3)} &\le \frac{1}{2^{2n}} \binom{2n}{n} \sum_{j,k} \left(\frac{M}{3^n} \frac{n!}{j!k!(n-j-k)!} \right) \\ &= \frac{1}{2^{2n}} \binom{2n}{n} \frac{M}{3^n} 3^n \\ &= \frac{1}{2^{2n}} \binom{2n}{n} M \\ &\approx \frac{1}{\sqrt{\pi n}} M \\ &\approx \frac{c}{\sqrt{\pi} n^{3/2}}. \end{aligned}$$

But the sum $\sum_{n=1}^\infty \frac{c}{\sqrt{\pi} n^{3/2}}$ *converges*, and hence so does the smaller sum $\sum_{n=0}^\infty p_{2n}^{(3)}$. Thus $\sum_{n=0}^\infty f_{2n}^{(3)}$ is strictly less than 1, i.e. the walker does not always return home!

So, surprisingly, we see different behavior in three dimensions from one or two dimensions. In fact, the two-dimensional case is critical: in two dimensions or fewer, the walker returns home, whereas in more than two dimensions, the walker does not necessarily return. The reason for putting the cutoff at 2, rather than at 3, is that if we were to come up with a sensible definition for a random walk in d dimensions, where d is not necessarily an integer, the walker would always return if $d \le 2$, but not if $d > 2$. (Indeed, one can come up with sensible definitions of this.) The reason for this is that the sum in question that we have to evaluate is quite close to

$$\sum_{n=1}^\infty \frac{1}{n^{d/2}}.$$

[5] Okay, so I just made that up. But I still think it's a good rule!

5. References

The best place to get started with generating functions is undoubtedly Herbert Wilf's *generatingfunctionology*, which can be downloaded for free from his website.[6] Looking back at the book after many years (I read it when I was in high school), I can see that, without even trying, I came up with several of the same examples as he did, presumably because that book so strongly shaped how I think about generating functions.

A good place to learn about the analytic theory of generating functions (plugging stuff into them), especially its applications to number theory, is the first chapter of Donald Newman's *Analytic Number Theory*.

The material on random walks can be found in many places, including Charles Grinstead and Laurie Snell's *Introduction to Probability*.

6. Problems

(1) Prove Binet's formula (equation (2)) for the Fibonacci numbers using generating functions.
(2) Find a closed formula for the Catalan numbers.
(3) Prove Theorem 1.
(4) Use generating functions to solve the recurrence $a_{n+1} = 2a_n + n$.
(5) Let $p(n)$ denote the number of *partitions* of n, i.e., the number of ways of writing n as the sum of positive integers, discounting the order of the terms. (For example, $p(5) = 7$: the 7 ways are 5, 4+1, 3+2, 3+1+1, 2+2+1, 2+1+1+1, 1+1+1+1+1.)
   (a) Show that
$$\sum_{n=0}^{\infty} p_n x^n = \prod_{n=1}^{\infty} \frac{1}{1-x^n}.$$
   (b) Show that the number of partitions of n into all odd parts is the same as the number of partitions of n into distinct parts, using generating functions. (For example, when $n = 5$, there are 3 of each. The partitions into odd parts are 5, 3+1+1, 1+1+1+1+1; the partitions into distinct parts are 5, 4+1, 3+2.)
   (c) Do part (b) by giving a bijection. Observe that this is substantially harder!
(6) Is it possible to relabel the sides of two dice with positive integer labels so that the probability of getting any sum from 2 to 12 is the same as that of ordinary dice?
(7) Is it possible to put weights on the sides of two ordinary dice so that the probability of getting each sum from 2 to 12 is $\frac{1}{11}$?
(8) (Putnam 2003) For a set S of nonnegative integers, let $r_S(n)$ denote the number of ordered pairs (s_1, s_2) such that $s_1, s_2 \in S$, $s_1 \neq s_2$, and $s_1 + s_2 = n$. Is it possible to partition the nonnegative integers into two sets A and B (i.e., so that $A \cup B = \mathbb{N}$, and $A \cap B = \varnothing$) in such a way that $r_A(n) = r_B(n)$ for all n?
(9) Show that
$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

[6] `http://www.math.upenn.edu/~wilf/DownldGF.html`

(10) Evaluate the sum
$$\sum_{n=0}^{\infty} \frac{n^2}{2^n}.$$
(11) (a) Show that $n^n e^{-n} < n! < n^{n+1} e^{-n}$.
(b) **Harder:** Show that $n! \approx C\sqrt{n}\, n^n e^{-n}$, for some constant C.
(c) **Even harder:** Show that $n! \approx \sqrt{2\pi n}\, n^n e^{-n}$. This is **Stirling's approximation**.
(12) In a one-dimensional random walk, what is the expected number of visits to the origin by time $2n$?
(13) Show that, in a one- or two-dimensional random walk, we eventually land on each lattice point with probability 1.
(14) Let d be a positive integer. Suppose we have two walkers on the integer lattice $\mathbb{Z}^d$, both starting from the origin, both doing independent random walks.
(a) For which values of d is it the case that two walkers will land on the same vertex at the same time (after the initial meeting) with probability 1?
(b) For which values of d is it the case that the two walkers will both land on the same point other than the origin, possibly at different times, with probability 1?

Euler Circle, Palo Alto, CA 94306
E-mail address: simon@eulercircle.com