Topics in generating functions

Qiaochu Yuan
Massachusetts Institute of Technology
Department of Mathematics

Written for the Worldwide Online Olympiad Training program
http://www.artofproblemsolving.com

April 7th, 2009

1 Introduction

Suppose we want to study a sequence $a_0, a_1, a_2, ...$ Such a sequence might be defined by a recurrence relation we're given, or it might count some family of sets. There are many specific classes of sequences with very different properties, so what general methods exist to study sequences? The general technique we'll discuss here is that of studying a sequence by studying its *Z-transform*, or *generating function*

$$A(x) = a_0 + a_1 x + a_2 x^2 + ... = \sum_{n \geq 0} a_n x^n. \tag{1}$$

An obvious advantage of this representation is that it's a very compact way of describing many simple sequences: for example, geometric series can be written as

$$a + arx + ar^2 x^2 + ... = \frac{a}{1 - rx}$$

which is just the statement of the geometric-series-summation formula. You can think of the introduction of the parameter x as useful because it allows you to restrict to x small enough so that we have convergence regardless of the value of r, but you don't really need calculus to understand generating functions: for most purposes it's more convenient to regard generating functions as *formal* and ignore questions of convergence. The ring of formal power series, denoted $\mathbb{C}[[x]]$, consists of infinite sums of precisely the above form, with addition defined by

$$(a_0 + a_1 x + ...) + (b_0 + b_1 x + ...) = (a_0 + b_0) + (a_1 + b_1)x + ... \tag{2}$$

and multiplication defined by

$$(a_0 + a_1 x + ...)(b_0 + b_1 x + ...) = c_0 + c_1 x + ..., c_n = \sum_{k=0}^{n} a_k b_{n-k}. \tag{3}$$

(c_n) is called the *convolution* of the sequences (a_n) and (b_n). One big advantage of the generating function approach is that convolution is a natural operation on many sequences of combinatorial interest and that talking about multiplying functions is easier than writing down convolutions. But as we'll see, the value of the generating functions approach is much deeper than this.

Example Denote the probability of rolling a sum of n with d six-sided dice by $p_{n,d}$. Clearly one rolls a sum of n by rolling a sum of $n - k$ with the first $d - 1$ dice and rolling a sum of k on the last die; in other words, we have the recurrence

$$p_{n,d} = \frac{p_{n-1,d-1} + p_{n-2,d-1} + ... + p_{n-6,d-1}}{6}.$$

This recurrence is difficult to work with until we realize that it is a convolution and equivalent to the following identity:

$$\begin{aligned}\sum_{n\geq 0} p_{n,d}x^n &= \left(\frac{x+x^2+...+x^6}{6}\right)\left(\sum_{n\geq 0} p_{n,d-1}x^n\right)\\ &= \left(\frac{x+x^2+...+x^6}{6}\right)^d.\end{aligned}$$

Note that, as expected, this identity tells us that the probability of rolling a sum less than d is zero and that the probability of rolling a sum of either d or $6d$ is $\frac{1}{6^d}$. The generating function $\frac{x+x^2+...+x^6}{6}$ is just $p_{1,1}x + p_{2,1}x^2 + ...$, the function that describes the probability of rolling each face on a six-sided die.

Now here's a computation you really don't want to do without generating functions: the factorization

$$x+x^2+...+x^6 = x\left(\frac{(x^3-1)(x^3+1)}{x-1}\right) = x(x+1)(x^2+x+1)(x^2-x+1),$$

which implies the following factorization:

$$\begin{aligned}(x+x^2+...+x^6)^2 &= x(x^2+1)(x^2+x+1)\cdot x(x^2+1)(x^2+x+1)(x^2-x+1)^2\\ &= (x+2x^2+2x^3+x^4)(x+x^3+x^4+x^5+x^6+x^8).\end{aligned}$$

What this means, interpreted combinatorially, is as follows: the probability distribution of rolling two normal six-sided dice is the same as the probability distribution of rolling a die with sides $1, 2, 2, 3, 3, 4$ and sides $1, 3, 4, 5, 6, 8$, and in fact substituting $x = 1$ and playing around with the factors above should convince you that this is the *only other pair of six-sided dice* for which this is true. A fact like this, which might seem to require a lot of tedious casework to verify, follows directly from the factorization of polynomials of the form $x^n - 1$ into their irreducible factors, known as *cyclotomic polynomials*.

In harder combinatorial problems, the sequence of interest won't have a nice formula even though it has a nice generating function. For example, the number of partitions $p(n)$ of a positive integer n into a sum of other positive integers (ignoring order) has the beautiful generating function

$$\sum_{n\geq 0} p(n)x^n = \frac{1}{(1-x)(1-x^2)(1-x^3)...}.$$

While sequences like $p(n)$ don't have "nice" closed forms, we can learn two very interesting things from this generating function: multiplying out the denominator (which is harder than it sounds), Euler obtained his *pentagonal number theorem*, which implies the beautiful recursion

$$p(k) = p(k-1) + p(k-2) - p(k-5) - p(k-7) + p(k-12) + p(k-15) - - + +....$$

While it is possible to give a combinatorial proof of this result, it's hard to imagine that it could've been discovered without generating functions. And the power of the generating function doesn't step there: analytic arguments allow us to deduce the asymptotic

$$p(n) \sim \frac{1}{4\sqrt{3}n} e^{\pi\sqrt{\frac{2n}{3}}}.$$

This beautiful result, due to Hardy and Ramanujan, is uniquely analytic. While we won't discuss the methods by which such estimates are obtained, it's good to keep in mind that generating functions can give us genuinely new information; see [6].

For combinatorialists, generating functions make the proof of certain combinatorial identities so easy in some cases that there are various combinatorial identities whose *only* proofs are via generating functions and for which a combinatorial proof isn't known. Hence generating functions also provide us with a rich source of difficult and interesting identities to explain.

1.1 What a "problem-solver" needs to know

There's a good reason to learn how to use generating functions properly, even if they don't show up too often on Olympiads: if you figure out what generating function to use on a problem, it tends to become very easy (at least, compared to other Olympiad problems). So it's a good idea to learn how to solve them in case they do show up!

This is not to say that it is always possible to apply these techniques automatically. Often figuring out the generating function becomes the main challenge of the problem. Fortunately, there are fairly systematic ways to do this, which we will attempt to cover. Keep the following general strategies and principles in mind as you read the specific strategies in the rest of this paper.

1. Be familiar with the simplest generating functions so you can recognize their coefficients when they appear in problems. (You'll know what these are once you're done reading.)

2. See if the problem statement implies a recursion. Recursions imply generating function identities.

3. Many natural operations on generating functions (multiplying by x, differentiating) are linear operators. If a computation seems difficult to do for a specific function F but it is easy to do for a class of functions $\mathcal{F}$ and it is linear, see if you can write F as a sum of functions in $\mathcal{F}$. (This might be called the Fourier-analytic point of view.)

4. If a summation seems symmetric, it might be the result of a product of generating functions. The simplest example is the sum $\sum a_k b_{n-k}$, which can be written more symmetrically as $\sum_{i+j=n} a_i b_j$. The sum $\sum_{i+j+k=n} a_i b_j c_k$, for example, is a product of three generating functions.

5. If you know the closed form of a single generating function F, you know the closed form of any generating function you can get by manipulating F and you can compute any sum you can get by substituting specific values into any of those generating functions.

6. Special cases are harder than general cases because structure gets hidden. If you introduce extra variables, you might figure out what the general case is and you can solve that instead. On the other hand, there is more than one way to introduce a variable into a problem.

7. Exchanging orders of summation can turn difficult computations into simple ones. [6] has some very good examples.

8. Sequences don't have to be numbers.

2 Basic results

Given a sequence (a_n), we call the associated function $A(x) = \sum_{n\geq 0} a_n x^n$ its *ordinary generating function*, or ogf for short. There are many other kinds of generating function, but we'll explore this case first. Given a function $A(x)$, the notation $[x^n]A(x)$ denotes the coefficient a_n of x^n. Adding generating functions is easy enough, but multiplication is worth discussing.

Definition Given two generating functions $A(x) = \sum_{n\geq 0} a_n x^n, B(x) = \sum_{n\geq 0} b_n x^n$, their product AB is the generating function $C(x) = \sum_{n\geq 0} c_n x^n$ with coefficients

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}. \tag{4}$$

(c_n) is called the *Cauchy product* or convolution of (a_k) and (b_k).

The following suggests a combinatorial motivation for this definition.

Proposition 2.1. *If A is a family of sets and a_k is the number of sets of "weight" k in A and B is a family of sets such that b_k is the number of sets of "weight" k in B, then c_n is the number of pairs of a set from A and a set from B of total "weight" n. We can therefore write*

$$A(x) = \sum_{a\in A} x^{|a|}, B(x) = \sum_{b\in B} x^{|b|}, C(x) = \sum_{c=(a,b)\in A\times B} x^{|c|} \tag{5}$$

where $|a|$ is the weight of a and $|c| = |a| + |b|$ (by definition). We call $A(x)$ the weight enumerator of A.

Note that the definition of weight is arbitrary: that's what makes this idea so powerful. Algebraically, this is the definition we would expect if we simply required that formal series multiply like polynomials, but as the combinatorial interpretation of the definition shows, it is both very general and usually the "correct" multiplication to think about combinatorially. If you think of the sets A, B as being "combinatorial sets," then C can be thought of as their "combinatorial Cartesian product"; this is the point of view taken up by *species theory*, which we will not discuss further, but the interested reader can consult [1].

Example Let A be the family of subsets of an n-element set X with weight the number of elements in a subset, and let B be the family of subsets of an m-element set Y, likewise. An ordered pair of a subset of X and a subset of Y determines a subset of the *disjoint union* $X \sqcup Y$, which has $m + n$ elements. Thus

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}.$$

In generating function terms, $(1 + x)^{m+n} = (1 + x)^m(1 + x)^n$. This is known as *Vandermonde's identity*, and we will return to it later.

Note that convolution is very different from the pointwise or *Hadamard* product $\sum_{n\geq 0} a_n b_n x^n$, which is in general very difficult to compute given only A, B.

While multiplication is natural to look at combinatorially, there are more natural operations on a single sequence.

Proposition 2.2. *Let $A(x) = \sum_{n\geq 0} a_n x^n$ be the generating function of (a_n) and define $s_n = a_0 + ... + a_n$. Then*

$$S(x) = \sum_{n\geq 0} s_n x^n = \frac{A(x)}{1-x} = A(x)\left(1 + x + ...\right). \tag{6}$$

Corollary 2.3. *Define $d_0 = a_0, d_n = a_n - a_{n-1}$. Then*

$$D(x) = \sum_{n\geq 0} d_n x^n = (1-x)A(x). \tag{7}$$

We call the transformation $(a_n) \mapsto (d_n)$ the *(backward) finite difference operator* and will write $d_n = \nabla a_n$. Later we will discuss both of these transformations in more depth, but first an application.

Example The harmonic numbers H_n are defined by $H_n = 1 + \frac{1}{2} + ... + \frac{1}{n}$. Compute their generating function.

Proof. We simply need to take the partial sums of the generating function $\sum_{n\geq 1} \frac{x^n}{n}$. With some familiarity with calculus, you might recognize this as the generating function of $\ln \frac{1}{1-x}$; it then follows that

$$\sum_{n\geq 1} H_n x^n = \frac{1}{1-x} \ln \frac{1}{1-x}.$$

□

Let $P(x) = \sum_{n\geq 0} p_n x^n$ describe the probabilities of rolling various faces of a die (which can be finite or infinite; that is, P can be either a polynomial or a formal series); p_n is the probability of rolling face n, and we require that $p_n \geq 0 \forall n$ and $P(1) = 1$. We can therefore deduce the following from definition, which is very convenient.

Proposition 2.4. *The expected value of a dice roll with generating function P is $P'(1)$.*

Generally, given a generating function for a sequence (a_n) we can easily compute the generating function for the sequence (na_n).

Proposition 2.5. *Given $A(x) = \sum_{n\geq 0} a_n x^n$, the generating function for na_n is $x\frac{d}{dx}A(x)$.*

Thus to multiply a sequence by a polynomial (equivalently, to take a Hadamard product of a sequence with a polynomial) all we have to do is repeatedly differentiate and multiply by x. We often abbreviate the derivative as D, and so the above can be understood as applying

an operator xD repeatedly. To multiply by higher-degree polynomials, apply xD repeatedly (that is, take it to various powers) and add the results.

Applying xD repeatedly to the generating function $\frac{1}{1-x}$ tells us how to compute the generating function of any polynomial.

Corollary 2.6. *The generating function of a polynomial $p(x)$ is $p(xD)\left(\frac{1}{1-x}\right)$.*

If $p(x) = \sum_{i=0}^{n} p_i x^i$, then $p(xD)$ is shorthand for the operator $\sum_{i=0}^{n} p_i (xD)^i$, which means "for each i, apply the differential operator xD i different times, multiply by p_i, and add the results obtained." This is an annoying computation to perform for polynomials of even moderate degree. In the next section we present an alternate method for computing the generating function of polynomials.

Example You have a coin that flips heads with probability p and tails with probability $1-p$. Flip the coin until you flip tails. What is the expected value of the square of the number of flips necessary for this to occur?

Proof. The probability of flipping heads n times and flipping tails is $p^n(1-p)$, which gives

$$P(x) = \sum_{k \geq 0} p^k (1-p) x^k = \frac{1-p}{1-px}.$$

Note that $P(1) = 1$ as is necessary. (This is a probabilistic proof of the geometric series summation formula.) We now have to compute $\sum k^2 p^k (1-p)$. We readily compute that $xD\left(\frac{1}{1-px}\right) = \frac{px}{(1-px)^2}$ and

$$xD\left(\frac{px}{(1-px)^2}\right) = \frac{px}{(1-px)^2} + \frac{2p^2x^2}{(1-px)^3}$$

hence we find that

$$\sum_{k \geq 0} k^2 p^k (1-p) x^k = \frac{(1-p)(px + p^2x^2)}{(1-px)^3}.$$

Substituting $x = 1$ gives $\frac{p(1+p)}{(1-p)^2}$. □

A remark is in order. The derivative on formal power series is purely a formal operation. It does not require any notion of a limit; it is defined by what it does to each term, and as such it is valid for coefficients in any ring. This observation will be of use later. In any case, if the derivative can be defined formally, then so can the integral, which has the following consequence.

Proposition 2.7. *Given $A(x) = \sum_{n \geq 0} a_n x^n$, the generating function for $\frac{a_n}{n+1}$ is $\int_0^x A(t)\,dt$.*

This observation is the basis for concocting rather difficult-looking identities.

Example Compute $\sum_{k=0}^{n} \frac{1}{k+1}\binom{n}{k}$. (This is one of the less difficult-looking ones.)

Proof. We merely have to integrate the generating function for the binomial coefficients, which gives

$$\int_0^1 (1+t)^n \, dt = \frac{(1+t)^{n+1} - 1}{n+1},$$

and substituting $t = 1$ we obtain the answer $\frac{2^{n+1}-1}{n+1}$. □

2.1 Exercises

Generally the exercises will vary wildly in difficulty and not really be arranged in order of increasing difficulty.

1. What is the expected size of a random subset of $\{1, 2, ...n\}$?

2. USAMO 1989 #1: For each positive integer n, let

$$S_n = 1 + \frac{1}{2} + ... + \frac{1}{n}$$

$$T_n = S_1 + S_2 + ... + S_n$$

$$U_n = \frac{T_1}{2} + \frac{T_2}{3} + ... + \frac{T_n}{n+1}.$$

   Find, with proof, integers $0 < a, b, c, d < 10000$ such that $T_{1988} = aS_{1989} - b$ and $U_{1988} = cS_{1989} - d$.

3. For a, b positive integers, compute

$$\sum_{i=0}^{b} (-1)^{b-i} \frac{1}{a+b-i} \binom{b}{i}.$$

4. (a) (Almost) USAMO 1996 #6: Determine (with proof) whether there is a subset X of the non-negative integers with the following property: for any non-negative integer n there is exactly one solution of $a + 2b = n$ with $a, b \in X$.

   (b) Putnam 2003 A6: For a set S of non-negative integers let $r_S(n)$ denote the number of ordered pairs of distinct elements $s_1, s_2 \in S$ such that $s_1 + s_2 = n$. Is it possible to partition the non-negative integers into disjoint sets A and B such that $r_A(n) = r_B(n)$ for all n?

5. USAMO 1991 #2: For any nonempty set S of numbers, let $\sigma(S)$ and $\pi(S)$ denote the sum and product, respectively, of the elements of S. Prove that

$$\sum \frac{\sigma(S)}{\pi(S)} = (n^2 + 2n) - \left(1 + \frac{1}{2} + ... + \frac{1}{n}\right)(n+1)$$

   where $\sum$ denotes a sum involving all nonempty subsets of $\{1, 2, ...n\}$.

3 Polynomials, finite differences, and summations

If functions of the form $\frac{1}{(1-x)^n}$ have coefficients which are polynomials, perhaps it would be worthwhile to figure out 1) exactly what polynomials these are, and 2) how to write an arbitrary polynomial as a linear combination of such polynomials. This would, in principle, be easier than repeated differentiation. If we could do so, it would be very easy to figure out the answer to evaluate sums such as

$$\sum_{k=0}^{n} k^3$$

since all we would have to do is compute $\sum_{k\geq 0} k^3 x^k$ and multiply it by $\frac{1}{1-x}$ as we have seen. A generalization of Proposition 2.3 turns out to be exactly what we need. Recall that $\nabla a_n = a_n - a_{n-1}$.

Proposition 3.1. *Let* $\nabla^0 a_n = a_n, \nabla^{k+1} a_n = \nabla(\nabla^k a_n)$. *If* $A(x) = \sum_{n\geq 0} a_n x^n$, *then*

$$\sum_{n\geq 0} \nabla^k a_n x^n = (1-x)^k A(x). \tag{8}$$

Corollary 3.2. *Let* $\mathbf{S}a_n = a_0 + a_1 + ... + a_n$, *and* $\mathbf{S}^0 a_n = a_n, \mathbf{S}^{k+1} a_n = \mathbf{S}(\mathbf{S}^k a_n)$. *Then*

$$\sum_{n\geq 0} \mathbf{S}^k a_n x^n = \frac{A(x)}{(1-x)^k} \tag{9}$$

The proposition explains why taking repeated finite differences of a generic sequence gets you binomial coefficients, if you have ever observed this pattern without understanding it. But if we understand the coefficients of $(1-x)^n$, what are the coefficients of $\frac{1}{(1-x)^n}$? When $n = 0$, this is the generating function of the sequence $1, 0, 0, 0, ...$ Repeatedly taking partial sums gives us the following family of sequences:

$$\begin{array}{cccccc} 1 & 1 & 1 & 1 & 1 & \dots \\ 1 & 2 & 3 & 4 & 5 & \dots \\ 1 & 3 & 6 & 10 & 15 & \dots \\ 1 & 4 & 10 & 20 & 35 & \dots \\ 1 & 5 & 15 & 35 & 70 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

The obvious conjecture immediately presents itself: this is Pascal's triangle turned sideways! This matrix is referred to as the (infinite) *symmetric Pascal matrix*. The construction we presented above is equivalent to the usual construction of Pascal's triangle, and it implies the following result.

Proposition 3.3. *The coefficient of* x^k *in* $\frac{1}{(1-x)^n}$ *is* $\binom{n+k-1}{n-1}$, *a polynomial of degree* $n-1$.

Corollary 3.4.

$$\sum_{n\geq 0}\binom{n}{k}x^n = \frac{x^k}{(1-x)^{k+1}}.$$

Pascal's triangle, however, is not the end of the story. The product of $(1+x+x^2+...)$ with itself n times counts the number of solutions, in non-negative integers, to $a_1+a_2+...+a_n = k$. This is because we can think of evaluating this product as choosing x^{a_1} from the first factor, x^{a_2} from the second factor, and so forth. We can then understand the above result as follows: there are n distinguishable urns into which we want to place k indistinguishable balls; the numerical value of a_i is the number of balls in urn i. To solve the problem combinatorially, consider a string of symbols, one for each of the k balls and $n-1$ other divider symbols. There are clearly $\binom{n+k-1}{n-1}$ ways of arranging these symbols. On the other hand, given any arrangement of balls and dividers we can take the set of balls before the first divider to belong in the first urn, the balls between the first and second divider to belong in the second urn, and so forth.

$\binom{n+k-1}{k}$ is also the number of *multisets* of k elements among n elements of a set. A multiset is a set into which an element may be placed more than once; the number of multisets of k elements from n elements is denoted $\left(\binom{n}{k}\right)$. Now, an application.

Example Compute the probability of rolling a sum of 18 on 4 six-sided dice.

Proof. This is the coefficient of x^{18} in $\left(\frac{x+x^2+...+x^6}{6}\right)^4$, but how do we actually compute it without going through a lot of tedious expansion? First, we can recognize that by the symmetry of the coefficients the coefficient of x^{18} is equal to the coefficient of x^{10}: in balls-and-urns terms, instead of solving the equation $a+b+c+d = 18$ where $1 \leq a,b,c,d \leq 6$ we can equivalently solve $(7-a)+(7-b)+(7-c)+(7-d) = 10$. Now for the important step: factor as

$$\frac{x^4}{6^4}\left(\frac{1-x^6}{1-x}\right)^4.$$

Now we only have to compute the coefficient of x^6 in $(1-x^6)^4 \cdot \frac{1}{(1-x)^4}$, and we know the coefficients in both of those now! The x^6 term of the right factor is $\binom{6+4-1}{3} = \binom{9}{3}$ and there is an additional term $-4x^6$ from the left factor, so our final answer is

$$\frac{\binom{9}{3}-4}{6^4}.$$

The combinatorial interpretation of the above argument is essentially inclusion-exclusion: first we count all solutions to $a+b+c+d = 16$ (or 12) by balls-and-urns, then we subtract the solutions where one of a,b,c,d is greater than 6, and so forth. Of course, this generalizes. □

Thinking of $\frac{1}{(1-x)^n}$ as $(1-x)^{-n}$ suggests the following generalization of the binomial theorem, which was proven by Newton.

Theorem 3.5. *For $\alpha \in \mathbb{C}$ and a non-negative integer k, define*

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)...(\alpha-(k-1))}{k!}. \tag{10}$$

Then

$$(1+x)^\alpha = \sum_{n\geq 0} \binom{\alpha}{n} x^n. \tag{11}$$

The polynomial $\binom{\alpha}{k}$ is sometimes called a Newton polynomial of degree k. It specifies to the usual binomial coefficient when α is a non-negative integer.

Proof. The k^{th} derivative of $(1+x)^\alpha$ is $k!\binom{\alpha}{k}(1+x)^{\alpha-k}$ by induction, so this follows by the familiar Taylor series formula. Ideally, we'd like to prove a version of this theorem that holds for actual values of x, but for now we will only require this identity to hold formally, which means it is really a definition of the function $(1+x)^\alpha$ which we will assume (but can prove) has all of the properties we expect it to have. □

Now, Corollary 5.6 tells us that the generating function of the Newton polynomial $\binom{n}{k}$ (with k fixed!) is $\frac{x^k}{(1-x)^{k+1}}$. To develop the theory of Newton polynomials it will be conceptually nicer to replace the backward finite difference operator with the *forward* finite difference operator $\Delta a_n = a_{n+1} - a_n$. Note that unlike the backward difference operator, the forward difference operator is not invertible, and for that reason it more closely resembles the derivative.

Proposition 3.6. *Given $A(x) = \sum_{n\geq 0} a_n x^n$,*

$$\sum_{n\geq 0} \Delta a_n x^n = \frac{A(x) - A(0)}{x} - A(x) = \frac{1-x}{x} A(x) - \frac{A(0)}{x}. \tag{12}$$

But this now implies that $\Delta\binom{n}{k} = \binom{n}{k-1}$, which is just Pascal's identity again. If we write $k!\binom{n}{k} = (n)_k = n(n-1)...(n-(k-1))$, the *falling factorial*, this now implies that

$$\Delta(n)_k = k(n)_{k-1}$$

which bears a striking resemblance to the similar rule for the polynomials n^k under differentiation. The similarity is not coincidental; identities of this kind are the subject of *umbral calculus*, which we will not discuss further, but see Rota's book [2] for an illuminating account. (Rota has some beautiful ideas about generating functions that are far beyond the scope of this article.) The point here is that we can prove an analogue of Taylor expansion.

Theorem 3.7. *Let $P(n)$ be a polynomial of degree d, and define $p_k = (\Delta^k P(n))_{n=0}$. Then*

$$P(n) = \sum_{k=0}^{d} p_k \binom{n}{k} = \sum_{k=0}^{d} p_k \frac{(n)_k}{k!}. \tag{13}$$

Corollary 3.8.

$$\sum_{n\geq 0} P(n)x^n = \sum_{k=0}^{d} p_k \frac{x^k}{(1-x)^{k+1}} = \frac{1}{1-x} F\left(\frac{1}{1-x}\right) \tag{14}$$

where $F(x) = \sum_{k\geq 0} p_k x^k$. (In fact, this holds for arbitrary F!)

Corollary 3.9. *Let $Q(n)$ be the polynomial of degree $d+1$ such that $Q(0) = 0$ and $Q(n) = P(0) + ... + P(n)$. Then*

$$Q(n) = \sum_{k=0}^{d} p_k \binom{n+1}{k+1}.$$

Corollary 3.10. *A polynomial $P(n)$ takes on integer values for integer values of n if and only if it is an integer-linear combination of the polynomials $\binom{n}{k}$.*

Proof. Let $R(n) = \sum_{k=0}^{d} p_k \binom{n}{k}$ and observe that $R(0) = p_0$ and that

$$\Delta^i R(n) = \sum_{k=i}^{d} p_k \binom{n}{k-i}$$

and therefore that $(\Delta^i R(n))_{n=0} = p_i$. This is exactly analogous to the proof of the Taylor formula; now $P(n) - R(n)$ has all of its first finite differences up to the d^{th} difference zero. But recall that the d^{th} difference of a polynomial of degree d is constant (which we can also prove by generating function methods using Corollary 2.6), hence all of its finite differences are zero and $P(n) - R(n) = 0$ as desired. □

Observe that the fact that $\binom{n}{k} = 0$ if $n < k$ implies the following:

$$P(0) = p_0$$

$$P(1) = p_0 + p_1$$

$$P(2) = p_0 + 2p_1 + p_2$$

and so forth. Generally, the following matrix identity holds:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 2 & 1 & 0 & \dots & 0 \\ 1 & 3 & 3 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{d}{0} & \binom{d}{1} & \binom{d}{2} & \binom{d}{3} & \dots & \binom{d}{d} \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_d \end{bmatrix} = \begin{bmatrix} P(0) \\ P(1) \\ P(2) \\ P(3) \\ \vdots \\ P(d) \end{bmatrix}.$$

The matrix in question here is a *(finite) lower-triangular Pascal matrix*, and lower-triangular matrices are clearly invertible, so we have a second affirmation of the notion that P is determined by $p_0, ...p_d$. Now, at last, it is time for some applications.

Example Compute $\sum_{k=0}^{n} k^3$.

Proof. $P(k) = k^3$ has first four terms $0, 1, 8, 27$, which gives the finite differences $1, 7, 19$, followed by $6, 12$, followed by 6. It follows that

$$k^3 = 6\binom{k}{3} + 6\binom{k}{2} + \binom{k}{1} \Leftrightarrow$$

$$\sum_{k=0}^{n} k^3 = 6\binom{n+1}{4} + 6\binom{n+1}{3} + \binom{n+1}{2}$$

which simplifies to the usual answer. □

Example A polynomial P of degree d satisfies $P(k) = 2^k, k = 0, 1, 2, ...d$. Compute $P(d+1)$.

Proof. Every finite difference is of the form $1, 2, 4, ...$, which gives the quite elegant

$$P(n) = \binom{n}{0} + \binom{n}{1} + ... + \binom{n}{d}.$$

Setting $n = d + 1$ we compute that $P(d + 1) = 2^n - 1$. □

A generalization is considered in the exercises.

3.1 Exercises

1. Compute
$$\sum_{k=0}^{n} F_k \binom{n}{k}$$
where F_k is the Fibonacci sequence (without using Binet's formula).

2. P is a polynomial of degree d satisfying $P(k) = q^k, k = 0, 1, ...d$, where $q \in \mathbb{C}$. Compute $P(d+1)$. Using this result, solve the following two problems.

3. Putnam 2008 A5: Let $n \geq 3$ be an integer. Let $f(x)$ and $g(x)$ be polynomials with real coefficients such that the points $(f(1), g(1)), ...(f(n), g(n))$ in $\mathbb{R}^2$ are the vertices of a regular n-gon in counterclockwise order. Prove that at least one of f, g has degree greater than or equal to $n-1$.

4. USAMO 1984 #5: $P(x)$ is a polynomial of degree $3n$ such that
$$P(0) = P(3) = ... = P(3n) = 2$$
$$P(1) = P(4) = ... = P(3n-2) = 1$$
$$P(2) = P(5) = ... = P(3n-1) = 0$$
$$P(3n+1) = 730.$$
Find n.

4 Binomial coefficients and lattice paths

The generalized binomial theorem alone, combined with the rule for products, is already powerful enough to prove several interesting identities. First, the observation that $(1+x)^\alpha(1+x)^\beta = (1+x)^{\alpha+\beta}$ gives us the following.

Proposition 4.1. *For any $\alpha, \beta \in \mathbb{C}$,*

$$\binom{\alpha+\beta}{k} = \sum_{i=0}^{k} \binom{\alpha}{i}\binom{\beta}{k-i}. \tag{15}$$

This is a generalization of the Vandermonde identity to arbitrary α, β. The tricky thing to do is apply this identity in cases where α, β aren't non-negative integers.

Corollary 4.2. *For any non-negative integers n, m, and in multichoose notation,*

$$\left(\binom{n+m}{k}\right) = \sum_{i=0}^{k} \left(\binom{n}{i}\right)\left(\binom{m}{k-i}\right). \tag{16}$$

Proof. Recall that $\binom{-n}{k} = (-1)^n\left(\binom{n}{k}\right)$; then this follows from Vandermonde's identity upon setting $\alpha = -n, \beta = -m$. We can also interpret the statement combinatorially in the obvious way: a multiset from an n-element set and a multiset from an m-element set determines a multiset of their disjoint union. Or stated in terms of balls-and-urns, putting k balls into $m+n$ urns is the same thing as putting i balls into the first n urns and putting $k-i$ balls into the last m urns for some i. □

Besides negative integers, there's one more particularly nice choice of α that occurs surprisingly often in combinatorial problems.

Proposition 4.3. *The generating function for the central binomial coefficients $\binom{2n}{n}$ is*

$$\sum_{n\geq 0} \binom{2n}{n} x^n = \frac{1}{\sqrt{1-4x}}. \tag{17}$$

Proof. The coefficient of x^n on the RHS is, by the general binomial theorem,

$$(-4)^n\binom{-\frac{1}{2}}{n} = (-4)^n\frac{(-1)(-3)(-5)...(1-2n)}{2^n n!} = \frac{(2^n n!)1\cdot 3\cdot 5\cdot ...\cdot(2n-1)}{n!n!} = \frac{(2n)!}{n!n!}$$

as desired. □

Integrating the above generating function, one obtains (nearly) the generating function for the *Catalan numbers* $C_n = \frac{1}{n+1}\binom{2n}{n}$. Because this sequence is so ubiquitous in combinatorics, we treat it separately instead. Catalan numbers count the following classes of objects:

1. Rooted ordered binary trees with $n+1$ leaves.

2. Rooted unordered binary trees with n internal vertices such that every vertex has either 0 or 2 children.

3. *Ballot sequences* of length $2n$: a sequence of n 1s and n -1s such that the partial sums of the sequence are always non-negative. Equivalently, Dyck words (where 1 is A and -1 is B) or Dyck paths: paths from $(0,0)$ to (n,n) going right or up that never cross the diagonal.

4. Number of triangulations of an $n+2$-gon with diagonals.

Stanley (see [4]) has compiled a list of combinatorial interpretations of the Catalan numbers that currently stands at at least 168; 66 are included in his book. For now, we will content ourselves with presenting the classic derivation of their generating function. To that end, let $C(x) = \sum_{n\geq 0} C_n x^n$. Given a rooted ordered binary tree with $n+2$ leaves, delete the root. What remains is a tree with $k+1$ leaves on the left and $n-k+1$ leaves on the right for some k, unless we started with the empty tree (which consists of a root alone). It follows that

$$C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k} \Leftrightarrow$$

$$C(x) = 1 + xC(x)^2.$$

This is because the coefficients of $C(x)^2$ count the number of *ordered pairs* of binary trees with $n+2$ leaves. Now, this is a quadratic equation, which we can solve. One of the roots doesn't have a power series expansion about 0 and the other one does, so we take the only possible choice, which is

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

Multiplying by x and taking the derivative gives

$$(xC(x))' = \frac{1}{\sqrt{1-4x}}$$

so $C_n = \frac{1}{n+1}\binom{2n}{n}$ as desired.

Alright, so how do we actually prove binomial coefficient identities? The first technique we'll investigate is based on the following results.

Proposition 4.4. $\binom{m+n}{m}$ *is the number of* lattice paths *on* $\mathbb{Z}^2$ *from* $(0,0)$ *to* (m,n) *that only go to the right or up, i.e. in steps of the form* $(0,1)$ *or* $(1,0)$.

Corollary 4.5. $\binom{2n}{n}$ *is the number of lattice paths from* $(0,0)$ *to* $(2n,2n)$ *that only go right or up. Equivalently, it is the number of lattice paths from* $(0,0)$ *to* $(2n,0)$ *in steps of the form* $(1,1)$ *or* $(1,-1)$.

Corollary 4.6. *$C_n = \frac{1}{n+1}\binom{2n}{n}$ is the number of lattice paths from $(0,0)$ to (n,n) that only go right or up and that don't cross the diagonal $y = x$. Equivalently, it is the number of lattice paths from $(0,0)$ to $(n,0)$ in steps of the form $(1,1)$ or $(1,-1)$ that don't cross the x-axis. (This is equivalent to the ballot-sequence definition.)*

Proof. Such a path must consist of m steps to the right and n steps up arranged in some order. But of course the number of ways to arrange $m+n$ symbols, m of which are of one type and n of which are of the other, is just $\binom{m+n}{m}$ as desired. This result has a pretty geometric interpretation: suppose we write on each lattice point in the first quadrant the number of ways to get there from $(0,0)$. Then we have written down an infinite Pascal matrix! The entries on the diagonal $y = x$ are the central binomial coefficients. It is also pretty easy to prove things like Vandermonde's identity this way: a path to $(k, m+n-k)$ has to pass through a point of the form $(i, m-i)$ for some i, i.e. the i^{th} row of Pascal's triangle.

The generating functions interpretation is quite deep. To study lattice paths we'll introduce a *two-variable generating function*

$$F(x,y) = \sum_{m,n \geq 0} l(m,n) x^m y^n \tag{18}$$

where $l(m,n)$ is the number of paths from $(0,0)$ to (m,n). Clearly you can only get to (m,n) from either $(m-1,n)$ or $(m,n-1)$, hence we have the recurrence

$$l(m,n) = l(m-1,n) + l(m,n-1)$$

where $l(m,n) = 0$ if (m,n) is not in the first quadrant. Along with the "boundary conditions" $l(m,0) = l(0,n) = 1$, this recurrence is equivalent to

$$F(x,y) = xF(x,y) + yF(x,y) + 1 \Leftrightarrow$$

$$F(x,y) = \frac{1}{1-x-y} = \sum_{k \geq 0} (x+y)^k$$

and the binomial theorem gives us the desired result. Alternately, you should be able to convince yourself directly that $(x+y)^k$ is the generating function for the number of paths of length k: a choice of x or y from each factor is precisely a choice to move to the left or right, respectively. □

The general method, then, is to interpret a binomial coefficient identity as counting some family of paths and, once you've figured out what the constraints on the steps are, turn it into a generating function. Here we will give an example of doing things the other way around, to remove the mystery.

Example Let's count the number of lattice paths starting at $(0,0)$ in steps of the form $(1,1)$ or $(2,0)$ ending on the line $x = n$. On the one hand, we don't even need all of the machinery

we just built up: at any point during the walk, the x-coordinate of the current location of the walk incremented by either 1 or 2, so if S_n is the number of such paths then

$$S_n = S_{n-1} + S_{n-2}$$

and we're just dealing with the Fibonacci recurrence! In particular, $S_1 = 1, S_2 = 2$ gives $S_n = F_{n+1}$.

On the other hand, all the machinery we just built up gives us stronger results. For a given point $(n, n-2j)$ on the line (note that $x-y$ is even at every step of the walk) we need to take $n-2j$ steps of the form $(1,1)$ and hence j steps of the form $(2,0)$; the number of ways we can do this is $\binom{n-j}{j}$, hence we have the beautiful identity

$$F_{n+1} = \sum_{j \geq 0} \binom{n-j}{j}$$

with a combinatorial explanation built in.

On the *third hand*, the generating function for taking k steps of the form $(1,1)$ or $(2,0)$ is $(xy + x^2)^k$, and summing over all k we obtain

$$\frac{1}{1 - xy - x^2} = \sum_{n \geq 0} \sum_{j \geq 0} \left(\binom{n-j}{j} y^{n-2j} \right) x^n.$$

The polynomial $F_{n+1}(y) = \sum_{j \geq 0} \binom{n-j}{j} y^{n-2j}$ is called a *Fibonacci polynomial*, and it generalizes the Fibonacci numbers. The Fibonacci polynomials satisfy $F_1(y) = 1, F_2(y) = y, F_{n+2}(y) = yF_{n+1}(y) + F_n(y)$, and specialize to the usual Fibonacci numbers when $y = 1$, which gives in particular

$$\sum_{n \geq 0} F_{n+1} x^n = \frac{1}{1 - x - x^2}.$$

Beautiful! We'll return to this generating function later.

The second general method of proving binomial coefficient identities we will discuss is what Wilf in [6] calls the *snake-oil* or "external" method. He contrasts this with the "internal" method, which is about manipulating the individual terms of a summation via various identities until the desired identity is proven. Wilf's approach is much slicker in some situations (no pun intended!) and can be carried out with surprisingly little knowledge of the combinatorial motivation for an identity. It is also very easy to describe: given a binomial identity which depends on a parameter n, multiply both sides by x^n and sum over all n, and then *exchange the order of summation.*

Example Show that

$$\sum_{k=0}^{n} (-1)^k \binom{n+k}{2k} C_k = 0$$

for all positive integers n.

Proof. This identity isn't as straightforward as a simple product of two generating functions; we'll just have to multiply by x^n and hope everything goes for the best. Now,

$$\sum_{n\geq 0}\sum_{k=0}^{n}(-1)^k\binom{n+k}{2k}C_k x^n = \sum_{k\geq 0}(-1)^k C_k x^k \sum_{n\geq 0}\binom{n+k}{n-k}x^{n-k}$$

which looks promising; the inner sum looks like a familiar generating function, except that we need to set $r = n - k$. Then the sum is

$$\sum_{k\geq 0}(-1)^k C_k x^k \sum_{n\geq 0}\binom{r+2k}{2k}x^r = \sum_{k\geq 0}(-1)^k C_k \frac{x^k}{(1-x)^{2k+1}}$$

which looks very promising: if we set $y = \frac{-x}{(1-x)^2}$ and factor out the $\frac{1}{1-x}$, then this is just

$$\frac{1}{1-x}\sum_{k\geq 0}C_k y^k = \frac{1}{1-x}\frac{1-\sqrt{1-4y}}{2y}.$$

Then some beautiful simplification occurs: $\sqrt{1-4y} = \frac{1+x}{1-x}$, and the above simplifies to

$$\frac{1-\frac{1+x}{1-x}}{\frac{-2x}{1-x}} = \frac{-2x}{-2x} = 1$$

so indeed the coefficient of x^n is 0 for all positive integers n as desired.

Can we translate this into a combinatorial proof? This is harder - we haven't discussed what it means combinatorially to substitute one generating function into another, so we'll prove as follows instead: following our Fibonacci discussion, $\binom{n+k}{2k}$ counts the number of lattice paths with $n-k$ steps of the form $(2,0)$ and $2k$ steps of the form $(1,1)$. Then C_k is the number of ways to *change the* $(1,1)$ *steps into a Dyck path* - in other words, $\binom{n+k}{2k}C_k$ counts the number of lattice paths with $n-k$ steps of the form $(2,0)$ and $2k$ steps of the form either $(1,1)$ or $(1,-1)$ ending at $(2n,0)$ and not crossing the x-axis. What we want to show is that there exists a bijection between the number of such paths with k even and the number of such paths with k odd, and that will prove the identity.

The bijection is as follows: find the first spot at which there is either a $(1,1)$ step followed by a $(1,-1)$ step or a $(2,0)$ step, and switch them: if it's the former, change it to the latter, and vice versa. This spot is uniquely determined by the path, so this bijection is its own inverse, and every path has a spot like this: since it has to start and end at the x-axis, it has to go up then down, so if it doesn't have a $(2,0)$ step at any point then at some point it attains a "local maximum." And we can readily see that this bijection changes the parity of k, which either increases or decreases by one. □

4.1 Exercises

1. In the style of the Vandermonde identity $(1+x)^\alpha(1+x)^\beta = (1+x)^{\alpha+\beta}$, give one-line proofs of the following identities.

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k} \text{ (Pascal's identity)}$$

$$\sum_{i=0}^{n} \binom{i+k-1}{k-1} = \binom{n+k}{k} \text{ (the Hockey-stick identity)}$$

$$\sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k}^2 = \begin{cases} 0 \text{ if n is odd} \\ (-1)^{n/2}\binom{n}{n/2} \text{ if n is even} \end{cases}$$

$$\sum_{i=0}^{k} (-1)^i \binom{n}{i}\binom{n+k-i-1}{k-i} = \begin{cases} 1 \text{ if } k=0 \\ 0 \text{ otherwise} \end{cases}$$

$$\sum_{k=0}^{n} \binom{2k}{k}\binom{2(n-k)}{n-k} = 4^n.$$

$$[x^k](1+x+x^2+x^3)^n = \sum_{j=0}^{n} \binom{n}{j}\binom{n}{k-2j} \text{ (Putnam 1992 B2)}$$

   Try to also give combinatorial proofs wherever possible.

2. How many lattice paths are there from $(0,0)$ to (m,n) of length l if we allow up, down, left, and right steps?

3. Compute

$$\sum_{k\geq 0} \binom{n}{3k}.$$

   (Hint: think of this as $\sum_{k\geq 0}\binom{n}{k}e_k$ where $e_k = 1$ if $k \equiv 0 \bmod 3$ and 0 otherwise.)

4. (a) Putnam 2005 B4: For positive integers m, n, let $f(m,n)$ denote the number of n-tuples $(x_1, ...x_n)$ of integers such that $|x_1| + ... + |x_n| \leq m$. Show that $f(m,n) = f(n,m)$.

   (b) Show that

$$\sum_{n\geq 0} f(n,n)x^n = \frac{1}{\sqrt{1-6x+x^2}}.$$

5 Linear recurrences, matrices, and walks on graphs

An important and well-understood class of sequences are those defined by a particularly simple kind of recurrence.

Definition A sequence (s_n) satisfies a *linear homogeneous recurrence* of degree d if there exist coefficients $a_{d-1}, ...a_0$ such that

$$s_{n+d} = a_{d-1}s_{n+d-1} + a_{d-2}s_{n+d-2} + ... + a_0 s_n. \tag{19}$$

The polynomial $P(x) = x^d - a_{d-1}x^{d-1} - ... - a_0$ is called the *characteristic polynomial* of (s_n).

Example Every geometric series $a_n = r^n$ satisfies $a_{n+1} = ra_n$. They have characteristic polynomial $x - r$.

Example Let $t_1 < t_2 < ... < t_k$ be a sequence of distinct integers and let T_n be the number of lattice paths from $(0, 0)$ to $(n, 0)$ in steps of the form $(t_i, 0)$ for some i. Then

$$T_n = T_{n-t_1} + T_{n-t_2} + ... + T_{n-t_k}.$$

When $t_1 = 1, t_2 = 2$ we recover the Fibonacci recursion. T_n can also be thought of as the number of tilings of a $1 \times n$ board with tiles of size $1 \times t_i$, or as the number of words from the alphabet $\{t_1, ...t_k\}$ with sum of "digits" equal to n. T_n has characteristic polynomial $x^{t_k} - x^{t_k - t_1} - ... - 1$.

Example Let p_n denote the probability that, if you roll a six-sided die continually and add up the partial sums, one of the partial sums will be equal to n. Then

$$p_n = \frac{p_{n-1} + p_{n-2} + ... + p_{n-6}}{6}.$$

Note the similarity to the previous example. p_n has characteristic polynomial $x^6 - \frac{x^5 + ... + 1}{6}$.

Example A polynomial sequence q_n of degree d has constant $(d+1)^{th}$ difference, hence satisfies

$$q_{n+d+1} - \binom{d+1}{1} q_{n+d} + \binom{d+1}{2} q_{n+d-1} \mp ... + (-1)^{d+1} q_n = 0.$$

q_n has characteristic polynomial $(x-1)^{d+1}$.

So we see that 1) sequences defined by linear homogeneous recurrences subsume both polynomials and geometric series, two of the simplest types of sequences, and 2) the behavior of such a sequence seems intimately related to the behavior of its characteristic polynomial. The basic facts making this precise can be proven by elementary means, but what we're looking for is generating functions!

Proposition 5.1. *Suppose s_n satisfies a recurrence with characteristic polynomial $P(x)$ of degree d and let $Q(x) = x^d P\left(\frac{1}{x}\right)$ be the polynomial whose coefficients are the coefficients of P in reverse order. Then there is a polynomial $R(x)$ of degree less than d such that*

$$\sum_{n \geq 0} s_n x^n = \frac{R(x)}{Q(x)}.$$

Proof. If you think about this hard enough, multiplying by $Q(x)$ tells you everything you need to know, but a good way to understand this result is via the following lemma.

Lemma 5.2. *Given $S(x) = \sum_{n \geq 0} s_n x^n$, the generating function for s_{n+1} is*

$$\frac{S(x) - S(0)}{x}.$$

The operation $s_n \mapsto s_{n+1}$ is called the left shift operator, which we will denote by $\mathbf{L}a$. What we need is the following idea: linear recurrences can be thought of as "differential equations" where the derivative is replaced with $\mathbf{L}$! To be more precise, the recurrence

$$s_{n+d} = a_{d-1} s_{n+d-1} + ... + a_0 s_n$$

is equivalent to the statement

$$\mathbf{L}^d s = a_{d-1} \mathbf{L}^{d-1} s + ... + a_0 s \Leftrightarrow$$

$$P(\mathbf{L}) s = 0.$$

where P is the characteristic polynomial of (s_n). Note that the 0 on the RHS means the zero *sequence*. Here we are thinking about $\mathbf{L}s$ as a sequence in the abstract - $\mathbf{L}$ is a linear operator on the space of sequences just as differentiation is a linear operator on the space of (nice) functions. The analogy with differential equations can be made precise if (s_n) is interpreted as the Taylor series of some function, but we will not pursue this interpretation. The point is that our lemma lets us translate this statement into a statement about generating functions: since $\mathbf{L}$ is basically division by x, up to the inclusion of an extra term, what we have is, after multiplying out by x^d,

$$x^d P\left(\frac{1}{x}\right) S(x) = R(x)$$

for some polynomial $R(x)$ determined by the $-\frac{S(0)}{x}$ terms accumulated from repeated application of $\mathbf{L}$ that we don't need to write down. □

We've already seen a special case of this result at work: in the one-dimensional lattice walk / tiling example above, the number of ways to put together m tiles of size $t_1, t_2, ...t_k$ is just

$$(x^{t_1} + x^{t_2} + ... + x^{t_k})^m$$

so summing over all m we obtain

$$\sum_{n\geq 0} T_n x^n = \frac{1}{1 - x^{t_1} - x^{t_2} - ... - x^{t_k}}$$

exactly as the more general result suggests, and in exact agreement with our specific discussion of the Fibonacci numbers. In fact, we can do better than this: suppose each tile, in addition to coming in different sizes, also comes in c_i different *colors*. Then it's not hard to see, either by looking at the recursion our summing over all m as above, that

$$\sum_{n\geq 0} T_n x^n = \frac{1}{1 - c_1 x^{t_1} - ... - c_k x^{t_k}}.$$

We can go even further: think of the c_is as *probabilities* that at any given step a particular tiling will be used, i.e. we are rolling a die with sides $t_1, t_2, ...t_k$ and probabilities $c_1, c_2, ...c_k$ of rolling each side, and we obtain a generalization of the dice example.

Okay, so what now? Well, if you're familiar with the theory of ODEs like the one we just wrote down, what we do is write the answer as a sum of exponentials. If why this works has never been explained to you, don't worry - we'll get it as a corollary of what we're about to do. The important viewpoint here is to think of exponentials as *eigenvectors* of the derivative operator, since of course

$$\frac{d}{dx} e^{rx} = re^{rx}$$

so we can think of e^{rx} as an eigenvector with eigenvalue r, and every eigenvector takes this form. In terms of their Taylor series, this is very easy to see: the Taylor series of e^{rx} just has coefficients $1, r, r^2, ...$, which gets sent to $r, r^2, r^3, ...$ What's the analogous function for the left shift? That's also straightforward:

$$\frac{A(x) - A(0)}{x} = rA(x) \Leftrightarrow$$

$$A(x) = \frac{A(0)}{1 - rx},$$

precisely the functions which are generating functions of $1, r, r^2, ...$! (The connection here is provided by the *Laplace transform*, which is not important for our purposes.) In other words, the corresponding thing to do for the left-shift operator is *partial fraction decomposition*. The standard example here is the Fibonacci numbers: as we saw earlier,

$$\sum_{n\geq 0} F_n x^n = \frac{x}{1 - x - x^2}.$$

This is just x times the generating function we deduced using lattice paths, and you can also deduce it using left-shifts. The partial fraction decomposition of the RHS is

$$\frac{1}{\phi - \varphi}\left(\frac{1}{1 - \phi x} - \frac{1}{1 - \varphi x}\right)$$

where ϕ, φ are the positive and negative roots of the characteristic polynomial $x^2 = x+1$, and by expanding $\frac{1}{1-rx}$ we recover Binet's formula. Note that although the denominator of the generating function has roots which are the reciprocal of the roots of the characteristic polynomial, partial fraction decomposition tells us the answer in terms of the reciprocal of those, i.e. the original roots of the characteristic polynomial as we suspected all along.

Now, just as in the ODE theory, we run into a slight problem when the characteristic polynomial has multiple roots. In the ODE theory this is resolved by multiplying the exponentials by a polynomial factor (in x), which is the same thing as multiplying the Taylor series by a polynomial factor (in n) - and in terms of $\mathbf{L}$, this just means that now we're looking at the functions $\frac{1}{(1-rx)^k}$, which in the language of linear algebra are the *generalized eigenvectors*. And these, if you're familiar with partial fraction decomposition, are all we need. Similarly, just as in the ODE theory, we will sometimes want to solve recurrences of the form

$$P(\mathbf{L})s = t$$

where t_n is some other sequence, such as 2^n or n^2. The secret here is that t_n is almost always, at least in problems of this type that I have seen, a sequence that satisfies some other recursion $Q(\mathbf{L})t = 0$, so it's straightforward to see that

$$Q(\mathbf{L})P(\mathbf{L})s = 0$$

and we proceed as before.

So how do we actually compute partial fraction decompositions? If you've been doing this all your life by solving a system of linear equations, the following might come as a pleasant surprise.

Proposition 5.3. *Let $\frac{A(x)}{B(x)}$ be a rational function such that r is a factor of $B(x)$ with multiplicity 1 and suppose that c_r is the coefficient of $\frac{1}{x-r}$ in the partial fraction decomposition of $\frac{A}{B}$. Then $c_r = \frac{A(r)}{B'(r)}$.*

Corollary 5.4. *Let $p(x) = \sum_{i \geq 0} p_i x^i$ be the probability distribution of some event where event i occurs with probability p_i and let*

$$\frac{1}{1-p(x)} = \sum_{n \geq 0} q_n x^n$$

be the generating function for the probability q_n that at some point, the sum of multiple trials of p will add up to exactly n. Suppose, in addition, that $\lim_{n\to\infty} q_n$ exists. Then

$$\lim_{n\to\infty} q_n = \frac{1}{p'(1)}$$

where $p'(1)$ is the expected value of p. (This is just the coefficient of $\frac{1}{1-x}$ in the partial fraction decomposition.)

Corollary 5.5. *Let $B(x) = \prod_{i=1}^{d}(x - x_i)$ where the x_i are distinct and let*

$$\frac{A(x)}{B(x)} = \sum_{i=1}^{d} \frac{y_i}{B'(x_i)(x - x_i)}$$

Then A is the unique polynomial of degree less than d such that $A(x_i) = y_i$. This is known as Lagrange interpolation.

Corollary 5.6. *Let $P(x) = \prod_{i=1}^{d}(x - r_i)$. Then*

$$\frac{P'(x)}{P(x)} = \sum_{i=1}^{d} \frac{1}{x - r_i}.$$

This last corollary, appropriately manipulated, is equivalent to Newton's sums. It's also a handy reminder of the value of logarithmic diferentiation as a "shortcut" to the product rule and can be surprisingly useful.

Proof. A slick way to see that this is true is by l'Hopital's rule. Simply compute

$$\lim_{x \to r} \frac{A(x)(x - r)}{B(x)} = \lim_{x \to r} \frac{(x - r)A'(x) + A(x)}{B'(x)} = \frac{A(r)}{B'(r)}.$$

In the partial fraction decomposition on the RHS, multiplying by $x - r$ isolates c_r and everything else is sent to zero (since r occurs with multiplicity 1). A similar statement is true for higher multiplicities, except that one must take more derivatives of B. Equivalently, write $B(x) = \prod_{i=1}^{d}(x - r_i)$. Then

$$\lim_{x \to r} \frac{A(x)(x - r)}{B(x)} = \lim_{x \to r} \frac{A(x)}{\prod_{i=1, r_i \neq r}^{d}(x - r_i)} = \frac{A(r)}{\prod_{i=1, r_i \neq r}^{d}(r - r_i)}$$

which, by the product rule, agrees with the above. □

It's a little messier to figure out the corresponding result for repeated roots. If r has multiplicity m, then the coefficient of $\frac{1}{(x-r)^m}$ is just $\frac{A(r)}{B^{(m)}(r)}$, but the rest of the coefficients are at best coefficients in the Taylor expansion of $\frac{A(x)(x-r)^m}{B(x)}$ about $x = r$, and these are a little more tedious to figure out.

One useful thing about these coefficients that we've figured out is that they can be used to extract some (admittedly coarse) asymptotics.

Example For large n, approximately how many ways are there to make change for n cents using pennies, nickels, dimes, and quarters?

Proof. The relevant generating function is

$$\sum_{n \geq 0} c_n x^n = \frac{1}{(1-x)(1-x^5)(1-x^{10})(1-x^{25}).}$$

We don't really want to compute the entire partial fraction decomposition (although we can make things a little easier for ourselves if we wanted to by setting $y = x^5$ and multiplying by $1 - x$, then putting it back in again), but the important thing here is that the roots of the denominator are all roots of unity, so they contribute periodic terms with constant or polynomial coefficients. (Such a function is called a *quasi-polynomial*, and as this example shows they are quite common in combinatorics.) The fifth roots of unity have multiplicity 3, so they contribute a periodic quadratic term; the other roots of unity have multiplicity 1; but $x = 1$ has multiplicity 4 and contributes a non-periodic cubic term, so this cubic term dominates in the limit. To compute it, we simply compute

$$\lim_{x \to 1} \frac{1}{(1 + ... + x^4)(1 + ... + x^9)(1 + ... + x^{24})} = \frac{1}{5 \cdot 10 \cdot 25}$$

which is the coefficient of $\frac{1}{(1-x)^4} = \sum_{n \geq 0} \binom{n+3}{3} x^n$. It follows that

$$c_n \sim \frac{n^3}{5 \cdot 10 \cdot 25 \cdot 3!}.$$

A simple geometric argument actually shows that this is obvious: c_n is the number of solutions to $x_1 + 5x_2 + 10x_3 + 25x_4 = n$ in non-negative integers, which is also the number of solutions to $5x_2 + 10x_3 + 25x_4 \leq n$. The set of all points (x_2, x_3, x_4) in the positive octant such that this is true form a tetrahedron with side lengths $\frac{n}{5}, \frac{n}{10}, \frac{n}{25}$ - which has area precisely the leading term we computed! The periodic terms, then, are a consequence of a higher-dimensional generalization of Pick's theorem called an *Ehrhardt quasi-polynomial* that we won't discuss further. □

5.1 Matrices and walks on graphs

Although many, many types of counting problems are described by linear recurrences, it's not always obvious how to write them down. Sometimes it'll be more natural to define a system of recurrences. The goal of this section is to describe a systematic way of doing so, and from there a systematic way to write down a single recurrence that describes the original problem.

Example (AIME 1990 #9) A fair coin is to be tossed 10 times. Let $\frac{i}{j}$ be the probability, in lowest terms, that heads never occurs on consecutive tosses. Find $i + j$.

Proof. Call a string of n tosses *good* if heads never occurs on consecutive tosses. Let H_n be the number of strings of n good tosses ending in heads and let T_n be the number of strings

of n good tosses ending in tails. The flip before the last heads in a good toss must be a tails, whereas the flip before the last tails in a good toss is arbitrary, so

$$H_n = T_{n-1}$$

$$T_n = H_{n-1} + T_{n-1}.$$

Combining this information, we find that $T_n = T_{n-1} + T_{n-2}$ and similarly $H_n = H_{n-1} + H_{n-2}$, so the total number of good tosses $G_n = T_n + H_n$ satisfies the same recurrence. But this is just the Fibonacci recurrence! In fact, since $H_1 = 2, H_2 = 3$ we find precisely that $H_n = F_{n+2}$. In particular, $H_{10} = 144$, so

$$\frac{144}{2^{10}} = \frac{9}{64}$$

and the answer is 73. □

Although this particular system of recurrences was easy to translate into a single recurrence, we want a general method to do this. One way to think aboutt his problem is to introduce two generating functions

$$H(x) = \sum_{n \geq 0} H_n x^n$$

$$T(x) = \sum_{n \geq 0} T_n x^n$$

and translate the system into a system of linear equations for H, T. This method is very general, since it can handle a system of recurrences each of which has arbitrary order, but in combinatorial questions of the above sort we often find that each of our recurrences has degree 1 since we write them down by considering what would happen at any given step. Perhaps there's a more specialized theory. And indeed, observe that the above has the matrix form

$$\begin{bmatrix} H_n \\ T_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} H_{n-1} \\ T_{n-1} \end{bmatrix}$$

which suggests that what we want are the tools of linear algebra. If we like, we can even define a matrix generating function. Let $\mathbf{T}$ be the above matrix; then

$$(\mathbf{I} - \mathbf{T}x)^{-1} = \sum_{n \geq 0} \mathbf{T}^n x^n.$$

Since $\mathbf{I} - \mathbf{T}x$ is invertible unless x is the inverse of an eigenvalue and we can take x to be very small, this operation makes sense analytically, but again we'll think of it as formal in x for our purposes. Now, we can think of the above as formal in both x and $\mathbf{T}$, but there's no need.

Theorem 5.7. *For every $n \times n$ matrix $\mathbf{M}$ there exists a polynomial P of degree at most n, its characteristic polynomial, such that $P(\mathbf{M}) = 0$. This polynomial is precisely*

$$P(t) = \det(t\mathbf{I}_n - \mathbf{M}). \tag{20}$$

The characteristic polynomial of $\mathbf{T}$ is $\mathbf{T}^2 = \mathbf{T} + \mathbf{I}$, precisely the characteristic polynomial of the Fibonacci sequence (why?). This suggests that we think of $\mathbf{T}$ as an "algebraic number" (more formally, an element of $F[t]/(t^2 - t - 1)$ where $F = \mathbb{C}(x)$) and try to divide directly. To "rationalize the denominator" here, we need to figure out the "conjugate" of $\mathbf{T}$, which is just $\mathbf{I} - \mathbf{T}$, hence

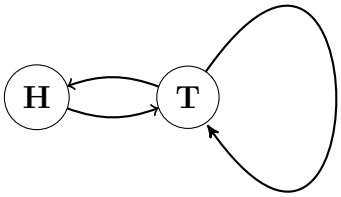
$$\frac{\mathbf{I} - (\mathbf{I} - \mathbf{T})x}{(\mathbf{I} - \mathbf{T}x)(\mathbf{I} - (\mathbf{I} - \mathbf{T})x)} = \frac{\mathbf{I}(1-x) + \mathbf{T}x}{\mathbf{I}(1 - x - x^2)} = \begin{bmatrix} \frac{1-x}{1-x-x^2} & \frac{x}{1-x-x^2} \\ \frac{x}{1-x-x^2} & \frac{1}{1-x-x^2} \end{bmatrix}.$$

Hence $\mathbf{T}^n = \begin{bmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{bmatrix}$. It's not hard to see that the denominator we'll end up getting here, in terms of x, is the same as the denominator we'd get if we solved for either H or T individually and took generating functions. But rather than discuss how to do this computation in general (it gets a little messy for higher than quadratic characteristic polynomials), we'll focus on the fact that computing the characteristic polynomial of a matrix that describes the sequence we want is the same thing as figuring out a recurrence for it. So how do we write these matrices down in the first place?

To, we can think about what we did in the coin problem algorithmically: we build a good toss by deciding, at each step, what letter to add to the end of a string (H or T). At every point in this algorithm, we keep track of the last letter of the string because it determines the next letter we can add. In other words, we're behaving like a finite state machine: the two states are "current string ends in H" and "current string ends in T," and the allowable transitions are from the first state to the second, from the second state to the first, and from the second state to itself. But rather than using the language of automata, it'll be more natural to use the language of graph theory.

Definition A finite directed graph with multiple edges $G = (V, E)$ is a set V of vertices $v_1, v_2, ...v_n$ together with a multiset E of ordered pairs (v_i, v_j) of directed edges, vertices, or arcs. (A graph without multiple edges has the requirement that E is a set.) A walk of length l is a sequence of vertices $w_0, ...w_l$ such that $(w_i, w_{i+1}) \in E$ for every i. (Note that a walk of length l is a word of length $l + 1$.)

Example The graph that describes the previous problem is as follows:



A walk on this graph is precisely a sequence of Hs and Ts such that H never appears twice in a row. I like to call this graph the Fibonacci graph.

The *adjacency matrix* $\mathbf{A}(G)$ of a graph G is the matrix with entries a_{ij} equal to the number of edges from v_i to v_j. For example, $\mathbf{T}$ is the adjacency matrix of the Fibonacci graph. A graph is *undirected* if its adjacency matrix is symmetric; that is, (v_i, v_j) is an arc if and only if (v_j, v_i) is an arc. An undirected graph is *simple* if $a_{ij} = 0$ or 1 and in addition $a_{ii} = 0$. Not only does the adjacency matrix of a graph completely describe it, but it turns out to be the natural way to talk about walks.

Proposition 5.8. *The number of walks of length l from v_i to v_j is $(\mathbf{A}(G))^l_{ij}$.*

The proof is by induction and the definition of matrix multiplication. In fact, one can regard this as the definition of matrix multiplication if an arbitrary matrix is regarded as a *weighted* graph where the number of edges between two vertices is replaced by a "flow." Such graphs occur in the study of electrical circuits and other networks, but we will not be concerned with them here. What we are concerned with is that, once constructed, the adjacency matrix automatically encodes a system of linear recurrences.

Proposition 5.9. *Let G be a graph with vertices $v_1, v_2, ...v_n$ and for each i let $s_{i,l}$ denote the number of walks of length l ending at v_i. Then*

$$\mathbf{s}_l = \begin{bmatrix} s_{1,l} \\ s_{2,l} \\ \vdots \\ s_{n,l} \end{bmatrix} = \mathbf{A}(G)^l \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{21}$$

It follows that $\mathbf{s}_{l+1} = \mathbf{A}\mathbf{s}_l$.

Corollary 5.10. *For any i, j, the sequence $a_l = (\mathbf{M})^l_{ij}$ satisfies a linear recurrence with characteristic polynomial the characteristic polynomial of $\mathbf{M}$.*

So in fact when we found the roots of the characteristic polynomial of a recurrence, we were really studying the eigenvalues of the adjacency matrix of a corresponding graph! (There is a close relationship between the notion of eigenvalue here and the notion of eigenvalue as applied to the left-shift operator.) Now what we want to do is this: given a combinatorial problem, convert it into the problem of counting walks on some graph. Rather than state a theorem, I'd like to illustrate this point with a few examples.

Example The characteristic polynomial of the adjacency matrix of the Fibonacci graph is $x^2 - x - 1 = 0$, as can be verified by direct calculation, and in total agreement with our other discussions. In fact, from the path-walking interpretation of the Fibonacci numbers alone we can re-derive the identity

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^n = \begin{bmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{bmatrix}$$

that we proved using generating functions, which gives, upon taking determinants, the neat identity $F_{n+1}F_{n-1} - F_n^2 = (-1)^n$. Can you find a combinatorial proof of this fact?

Example USAMO 1996 #4: An n-term sequence $(x_1, x_2, \ldots, x_n)$ in which each term is either 0 or 1 is called a binary sequence of length n. Let a_n be the number of binary sequences of length n containing no three consecutive terms equal to 0, 1, 0 in that order. Let b_n be the number of binary sequences of length n that contain no four consecutive terms equal to 0, 0, 1, 1 or 1, 1, 0, 0 in that order. Prove that $b_{n+1} = 2a_n$ for all positive integers n.

Proof. There's a cute bijective solution (think finite differences mod2), but we're just going to compute a_n and b_n instead for illustrative purposes. Algorithmically, in making strings that satisfy the condition that they avoid 010 we need to keep track of how close we are to not satisfying the condition. There are three states we could be in: we may have traced out two of the three characters necessary to spell 010, one of the characters, or none of the characters. For example, if our string ends in 11 then we haven't traced out any of the characters necessary. In other words, our state is determined by "where we are" in the forbidden word 010, so we'll name the vertices of our graph as follows:

1. **A** means our string ends in 0, but not in 010, so in two more steps we can spell 010.
2. **B** means our string ends in 01, so in one more step we can spell 010.
3. **Z** means our string ends in 1, but not in 01, so in three more steps we can spell 010.

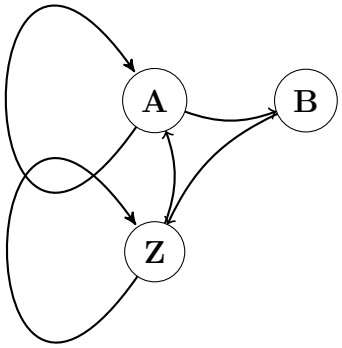
The three vertices will also lead a dual life: when talking about walks on this graph, we'll identify **A** with 0 and **B**, **Z** with 1, for the obvious reasons. (It's messier to talk about identifying edges, rather than vertices, with letters because of the need to have a start vertex.) This means that a valid walk of length l on this graph, when translated using these identifications, describes a binary sequence of length $l + 1$ without 010 in it, i.e. something counted by a_{l+1} (well, unless it starts at **B**). The edges between the vertices, which represent possible moves, are as follows:

1. **A** points to **B** and itself. If we plop down a 1, we're at **B**. If we plop down a 0, we're still at **A**.

2. **B** points to **Z**. Since we can't spell 010, the only thing we can do is plop down a 1, and that takes us "back to the beginning," i.e. we haven't spelled out any of 010.

3. **Z** points to **A** and itself. If we plop down a 0, we're at **A**; otherwise, we're still "at the beginning."

This gives the following graph:



What we've done is essentially performed the *Knuth-Morris-Pratt* algorithm, which is an algorithm that searches for a string by constructing a finite state machine that measures how close a given text is to spelling it out at any point. This state machine is very nearly the directed graph we want to construct to *avoid* this string; the above example should be enough to get the idea across in lieu of an actual description of the algorithm.

Now, the adjacency matrix of this graph is therefore

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

which has characteristic polynomial $P_a(\lambda) = \lambda^3 - 2\lambda^2 + \lambda - 1$. It's easy to verify that P_a has no rational roots, but the point is that a_n is the number of walks starting from either **Z** or **A** of length $n-1$, so we can write down the recursion

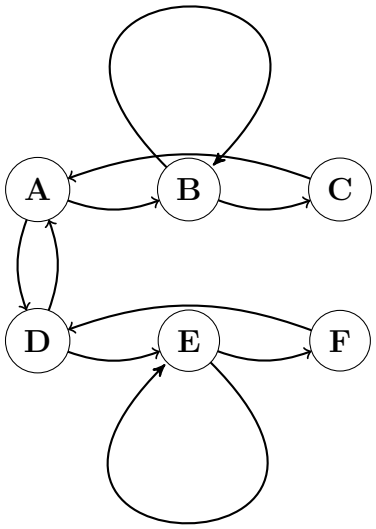
$$a_{n+3} = 2a_{n+2} - a_{n+1} + a_n$$

and even compute initial values fairly quickly. What I'd like to emphasize is that the above procedure is totally algorithmic and hence very general. While it might be possible to cleverly figure out the above recursion for this particular problem, the KMP algorithm allows us to solve any problem of the type "count the number of strings on some alphabet that avoid a particular substring."

The KMP algorithm doesn't apply to computing b_n because there are now two substrings we want to avoid, but it is not hard to modify: the generalization is called the *Aho-Corasick* algorithm, but again rather than describe the algorithm we'll perform it for b_n. The vertices and edges are as follows, with similar reasoning as the reasoning for a_n:

1. **A** means the string ends in 0, but not in 110. **A** points to **D** and **B**.
2. **B** means the string ends in 00, but not in 1100. **B** points to **C** and itself.
3. **C** means the string ends in 001. **C** points to **A**.
4. **D** means the string ends in 1, but not in 001. **D** points to **E** and **A**.
5. **E** means the string ends in 11, but not in 0011. **E** points to **F** and itself.
6. **F** means the string ends in 110. **F** points to **D**.

The graph is as follows:



As before, b_n is the number of walks of length $n-1$ starting from **A** or **D**. The adjacency

matrix of the graph is

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{N} & \mathbf{M} \end{bmatrix}$$

where $\mathbf{M}, \mathbf{N}$ are 3×3 matrices. Now, it's not hard to see that we can block-diagonalize as follows:

$$\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{N} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{M}+\mathbf{N} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}-\mathbf{N} \end{bmatrix}$$

where $\mathbf{I}$ is the 3×3 identity. This block-diagonalization takes advantage of a particular symmetry of the problem, but the point stands: Aho-Corasick is a totally algorithmic way of solving problems like this in general. But let's finish. Since b_n is the number of walks starting from either $\mathbf{A}$ or $\mathbf{D}$, it was (before we changed bases) generated by the vector with entries $1, 0, 0, 1, 0, 0$ - but this is precisely the first column of the matrix we changed bases by! It follows that the vector we want in the new basis is just $1, 0, 0, 0, 0, 0$; that is, b_n is the first entry in

$$\begin{bmatrix} \mathbf{M}+\mathbf{N} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}-\mathbf{N} \end{bmatrix}^{n-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and hence it is totally determined by the behavior of

$$\mathbf{M}+\mathbf{N} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

But this is just a permutation of the adjacency matrix of the graph of a_n! So in fact a_n and b_n satisfy the same recursion, and the rest can be proven by checking initial conditions or by writing down exactly how a_n and b_n are related by permuting the above matrix appropriately. We've even been naturally led, in some sense, to the bijective proof: what the above tells us is that certain pairs of points in the graph of b_n get sent to certain other pairs in the same way that points of the graph of a_n get sent to other points. $\square$

5.2 Exercises

1. IMO 1974 #3: Prove that for any natural number n, the number

$$\sum_{k=0}^{n} \binom{2n+1}{2k+1} 2^{3k}$$

   is not divisible by 5.

2. AIME II 2003 #13: A bug starts at a vertex of an equilateral triangle. On each move, it randomly selects one of the two vertices where it is not currently located, and crawls along a side of the triangle to that vertex. Given that the probability that the bug moves to its starting vertex on its tenth move is $\frac{m}{n}$ where m and n are relatively prime positive integers, find $m+n$.

3. Let c_n denote the number of binary sequences avoiding the substrings $0110, 1001$ and such that the total number of 0s is a multiple of 3. Construct a graph for c_n (as in the examples) and write down a recurrence that describes it. (The key point here is that you can no longer take c_n to be a valid walk and only fix its starting position: you'll also need to fix its ending position to get the number of 0s to come out right.)

4. IMC 2007 #5: Let n be a positive integer and $a_1, ...a_n$ be arbitrary integers. Suppose that a function $f : \mathbb{Z} \to \mathbb{R}$ satisfies

$$\sum_{i=1}^{n} f(k + a_i l) = 0$$

   whenever k and l are integers and $l \neq 0$. Prove that $f = 0$.

5. Let J_n denote the graph on n vertices $1, 2, ...n$ with the edges $1 \to 2, 2 \to 3, ...n-1 \to n$ and λ loops at each vertex. Count the number of paths from i to j of length l, hence compute the powers of $\mathbf{A}(J_n)$ (which is a Jordan block). Compare with the partial fraction decomposition of $\frac{1}{(1-\lambda x)^n}$.

6. Let T_n denote the number of lattice walks from $(0,0)$ to $(n,0)$ in steps of length 1 through k where the step of length i comes in c_i different colors, and c_i can be equal to zero. In the following problem it may be helpful to think of c_i as a symbolic weight rather than an integer.

   (a) Construct a graph G on k vertices such that T_n is the number of closed walks from a particular vertex to itself of length n. What is the adjacency matrix $\mathbf{A}(G)$ of G? What are its eigenvectors?

(b) Let G_1, G_2 be two graphs. The *Cartesian product* $G_1 \square G_2$ is the graph with vertex set the Cartesian product of the vertex sets of G_1, G_2 and an edge from (u, u') to (v, v') if and only if $u = v$ and there is an edge from u' to v' or $u' = v'$ and there is an edge from u to v. What is the relationship between $\mathbf{A}(G_1 \square G_2)$ and $\mathbf{A}(G_1), \mathbf{A}(G_2)$ (sometimes called the *Kronecker sum*)? Between their eigenvalues? What is the relationship between walks on $G_1 \square G_2$ and walks on G_1 and G_2? (You may want to look at the exercise on exponential generating functions if you get stuck.)

(c) The *tensor product* (also known as the direct product, categorical product, or *Kronecker product*) $G_1 \times G_2$ is the graph with vertex set the Cartesian product of the vertex sets of G_1, G_2 and an edge from (u, u') to (v, v') if and only if there is an edge from u to v and an edge from u' to v'. What is the relationship between $\mathbf{A}(G_1 \times G_2)$ and $\mathbf{A}(G_1), \mathbf{A}(G_2)$ (also called a Kronecker product)? Between their eigenvalues? What is the relationship between walks on $G_1 \times G_2$ and walks on G_1 and G_2?

(d) Let G be the Cayley graph of $\mathbb{Z}$ with $S = \{1, -1\}$. Why is $G \square G$ isomorphic to $G \times G$? Show that there are infinitely many finite graphs with this property.

(e) An *algebraic integer* is a complex number that is the root of a monic polynomial with integer coefficients. Show that the sum and product of two algebraic integers is algebraic.

6 The roots of unity filter

Suppose we have a polynomial or a power series

$$F(x) = \sum_{n \geq 0} f_n x^n$$

and we're interested in extracting only the terms with n even or n odd. It's not hard to see how to do this by noting that

$$F(-x) = \sum_{n \geq 0} f_n (-x)^n$$

which, combined with F, allows us to isolate the even and odd terms: in fact, we have

$$\sum_{n \geq 0} f_{2n} x^{2n} = \frac{F(x) + F(-x)}{2}$$

$$\sum_{n \geq 0} f_{2n+1} x^{2n+1} = \frac{F(x) - F(-x)}{2}.$$

Probably the most famous example of such an extraction is $F(x) = e^{ix}$. This also offers a pretty easy proof that

$$\sum_{k \geq 0} \binom{n}{2k} = \sum_{k \geq 0} \binom{n}{2k+1} = \frac{(1+1)^n \pm (1-1)^n}{2} = 2^{n-1},$$

although this is a pretty easy identity to prove combinatorially. Nevertheless, this idea is surprisingly useful.

Example Let $S_n = \{\frac{1}{2}, \frac{1}{3}, ...\frac{1}{n}\}$ and, given a subset S of S_n, let $P(S)$ denote the product of all of the elements of S. What is the sum of all such products over all subsets S with an even number of elements?

Proof. Remove the parity constraint. A moment's thought reveals that the answer is

$$\prod_{k=2}^{n} \left(1 + \frac{1}{k}\right) = \frac{n+1}{2}$$

since the expansion of this product contains every product $P(S)$ exactly once. This is just an expression of the fact that $(1+t_1)...(1+t_n)$ gives the elementary symmetric polynomials in $t_1, ...t_n$. Now, add the parity constraint back in. The sum over all products $P(S)$ with $|S|$ even minus the products with $|S|$ odd is then, after another moment's thought,

$$\prod_{k=2}^{n} \left(1 - \frac{1}{k}\right) = \frac{1}{n}.$$

Then the answer we want is just $\frac{\frac{n+1}{2} + \frac{1}{n}}{2} = \frac{n^2+n+2}{4n}$. □

Our goal is to generalize this technique to finding the terms with k in an arbitrary arithmetic progression; equivalently, with k in some residue class modulo n for some n. I could just write down what the generalization looks like, but instead I'll try to motivate its proof in as many ways as possible. To that end, fix a modulus n and define the generating functions

$$F_a(x) = \sum_{k \equiv a \bmod m} f_k x^k.$$

We've already seen that when $m = 2$ we have $F(x) = F_0(x) + F_1(x)$ and $F(-x) = F_0(x) - F_1(x)$. What's the natural generalization? If you guessed m^{th} roots of unity, you'd be right on! In fact, a good way to think about the functions F_a is that they are the decomposition of F into "symmetric" parts. Symmetry here means symmetry with respect to roots of unity.

For example, letting $\omega = e^{\frac{2\pi i}{3}}$ and $m = 3$ we have, in matrix notation,

$$\begin{bmatrix} F(x) \\ F(\omega x) \\ F(\omega^2 x) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix} \begin{bmatrix} F_0(x) \\ F_1(x) \\ F_2(x) \end{bmatrix}.$$

The question is how to invert this mysterious symmetric matrix. Instead of answering the question, I'll ask it in a different way. If we think of the operation that takes F to F_a as a sort of dot product (i.e. Hadamard product), then what we should instead be investigating is the functions $\frac{x^a}{1-x^n}$, since if there were such a dot product then

$$\left\langle F(x), \frac{x^a}{1-x^n} \right\rangle = F_a(x).$$

Of course, earlier I said that Hadamard products are hard to take. But there is "sort of" a way to do it. If you have two polynomials

$$P(x) = \sum_{i=0}^{n} p_i x^i, Q(x) = \sum_{i=0}^{n} q_i x^i$$

you can compute the Hadamard product of their coefficients by taking the constant term of the *Laurent polynomial*

$$P(x)Q\left(\frac{1}{x}\right) = \sum_{i=0}^{n} p_i q_i + \text{other terms}.$$

The problem with this approach is that we obviously can't just set $x = 0$ to extract the constant term. The bigger problem is that this operation is not very well-defined on power series: since each term will have an infinite number of summands, we can't multiply Laurent series formally unless they only have finitely many $\frac{1}{x^k}$ terms. What we could do is compute

$$P(x)Q\left(\frac{y}{x}\right) = \sum_{i=0}^{n} p_i q_i y^i + \text{other terms}$$

but it's still not clear how to extract the constant term here. If only there were some kind of linear operator that could conveniently remove the other terms...

Depending on how familiar you are with Fourier analysis, you may or may not find the next step motivated. We're going to substitute $x = e^{it}$ and integrate away all the terms we don't want. More formally, the following is true.

Theorem 6.1. *Let $A(t) = \sum_{n\in\mathbb{Z}} a_n e^{int}, B(x) = \sum_{n\in\mathbb{Z}} b_n e^{int}$ be two Fourier series. Then*

$$\sum a_n \overline{b_n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(t)\overline{B(t)}\,dt.$$

This is known as Parseval's theorem: it implies that the Fourier transform is unitary. If you haven't seen it before, just remember that the terms $e^{int}e^{-int}$ in the product $A(t)\overline{B(t)}$ integrate to 2π whereas the terms $e^{int}e^{-imt}, m \neq n$ integrate to zero. What Parseval's theorem tells us is that, letting $z = e^{it}, A(t) = F(rz)$ for some r which you can think of as being either formal or in the disc $|r| < 1$, and $B(t) = \frac{z^a}{1-z^n}$, we obtain

$$\begin{aligned} F_a(r) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{F(rz)z^{n-a}}{z^n-1}\,dt \\ &= \frac{1}{2\pi i}\int_{-\pi}^{\pi} \frac{F(rz)z^{n-a-1}}{z^n-1} iz\,dt \\ &= \frac{1}{2\pi i}\oint_{|z|=1} \frac{F(rz)z^{n-a-1}}{z^n-1}\,dz \end{aligned}$$

where we've rewritten the integral from $-\pi$ to π in x as a contour integral over the circle in z. Those familiar with complex analysis should recognize the Cauchy integral formula as being applicable, which says this:

Theorem 6.2. *Let D be a disc in $\mathbb{C}$, let C be the boundary of D, and let f be a holomorphic function on an open set containing D. Then for every p in the interior of D,*

$$f(p) = \frac{1}{2\pi i}\oint_C \frac{f(z)}{z-p}\,dz.$$

All we have to do now is figure out the partial fraction decomposition of $\frac{z^{n-a-1}}{z^n-1}$. But this is straightforward: Proposition 5.3 gives

$$\frac{z^{n-a-1}}{z^n-1} = \sum_{k=0}^{n-1} \frac{\zeta^{k(n-a-1)}}{n\zeta^{k(n-1)}(z-\zeta^k)} = \frac{1}{n}\sum_{k=0}^{n} \frac{\zeta^{-ka}}{z-\zeta^k}$$

where $\zeta = e^{\frac{2\pi i}{n}}$ is a primitive n^{th} root of unity. Now, of course, we have all the tools we

need because the Cauchy integral formula tells us that

$$\begin{aligned} F_a(r) &= \frac{1}{2\pi i}\oint_{|z|=1}\sum_{k=0}^{n-1}\frac{F(rz)\zeta^{-ka}}{z-\zeta^k}\,dz \\ &= \frac{1}{n}\sum_{k=0}^{n-1}F(r\zeta^k)\zeta^{-ka}. \end{aligned}$$

You may be familiar with the $a = 0$ case - it's the easiest one to convince yourself of - but the general case has a nice ring to it.

Example Let's compute the sum

$$\sum_{k\equiv a \bmod m}\binom{n}{k}.$$

For some reason, sums of this type appear on competitions a lot (such as Putnam 1974, which is $m = 3$, but it's not worth repeating when we can handle the general case immediately). If we take $F(x) = (1+x)^n$, then the above is just

$$F_a(1) = \frac{1}{m}\sum_{i=0}^{m-1}(1+\zeta^i)^n\zeta^{-ia}.$$

In particular,

$$\sum\binom{n}{3k} = \frac{2^n + (-\omega)^n + (-\omega^2)^n}{3}.$$

The $m = 3$ and $m = 4$ cases are somewhat special, since the terms $1+\zeta^i$ are themselves (at least multiples of) roots of unity; this stops being true for large m.

To see how this relates to inverting the matrix we wrote down earlier, think about Corollary 5.5, which says that if

$$\frac{P(x)}{Q(x)} = \sum_{i=0}^{n-1}\frac{y_i}{Q'(x_i)(x-x_i)}$$

then $P(x_i) = y_i \forall i$. If we write $P(x) = \sum_{i=0}^{n-1} p_i x^i$, then writing this condition out produces a system of linear equations whose matrix is the *Vandermonde matrix* for $x_0, ...x_{n-1}$. If we let then $x_i = \zeta^i$ and $y_i = \zeta^{(n-a-1)i}$, the interpolation is obvious: $P(x) = x^{n-a-1}$ (and $Q(x) = x^n - 1$). On the other hand, this is exactly the partial fraction decomposition we just found! And what we have to do to solve this interpolation problem is invert the Vandermonde matrix

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta & \zeta^2 & \dots & \zeta^{n-1} \\ 1 & \zeta^2 & \zeta^4 & \dots & \zeta^{2(n-1)} \\ 1 & \zeta^3 & \zeta^6 & \dots & \zeta^{3(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta^{n-1} & \zeta^{2(n-1)} & \dots & \zeta^{(n-1)(n-1)} \end{bmatrix}$$

which is precisely the matrix that relates $F_a(x)$ to $F(\zeta^k x)$. In fact, the result we just derived is equivalent to the rather remarkable statement that the above matrix is unitary (well, after multiplying by $\frac{1}{\sqrt{n}}$), i.e. its conjugate transpose is its own inverse. Actually, the connection goes even deeper: recall that the Fourier coefficients of a function on $[-\pi, \pi]$ are defined by

$$\hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx}\, dx$$

which gives

$$f(x) = \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{inx}$$

whereas here we have the relationship

$$F_a = \frac{1}{n} \sum_{k=0}^{n-1} F(r\zeta^k) \zeta^{-ka}$$

that was deduced from the relationship

$$F(x) = \sum_{k=0}^{n-1} F_a(x) \zeta^a.$$

Look familiar? If we identify n with 2π, the sum with the integral, and ζ with e^{ix} these formulas are identical! In fact, the relationship between F and F_a is *also that of a Fourier transform*. The unitary matrix that relates F and F_a is called the *discrete Fourier transform* matrix of order n, and these two cases are examples of a deeper phenomenon called *Pontryagin duality*. For a survey of the basic results I highly recommend Stein and Shakarchi [5].

As promised, I want to motivate what we're doing here in as many ways as possible. One way to think about what we did is the following: F_a is an n-periodic function in a. An n-periodic function satisfies the linear recurrence

$$s_{n+k} = s_k \forall k$$

so its characteristic polynomial is $x^n - 1$ and every such function can be written as a sum of n^{th} roots of unity raised to the appropriate power. If we've solved the problem for the periodic sequence $1, 0, 0, ...1, 0, 0, ...$ then we can just shift whatever coefficients we get in that case to get the coefficients for the other cases, and by linearity we're done. But observe that the claim that the DFT matrix is unitary is equivalent to the claim that the vectors (ζ^{ak}) form an *orthonormal basis* of the space of n-periodic functions (if we define the inner product correctly):

$$\left\langle \zeta^{ak}, \zeta^{bk} \right\rangle = \frac{1}{n} \sum_{k=0}^{n-1} \zeta^{ak} \overline{\zeta^{bk}} = \begin{cases} 1 \text{ if a = b} \\ 0 \text{ otherwise} \end{cases}.$$

Some authors don't put the n in the inner product and instead take the basis to be $\frac{\zeta^{ak}}{\sqrt{n}}$, but I think this is unnecessary. Orthogonality is a general property of Fourier series, and although it is very deep, it's also very easy to prove in this case: if $a \neq b$ then n can't divide $a-b$, so $\zeta^{(a-b)k}$ is some multiple of some sum of primitive m^{th} roots of unity for some $m|n$, and it's obvious that these will sum to zero. And once we have an orthonormal basis, we also know how to write any vector in that basis: the coefficient of ζ^{ak} is

$$\left\langle s, \zeta^{ak} \right\rangle = \frac{1}{n} \sum_{k=0}^{n-1} s_k \zeta^{-ak}.$$

And this is the bare-bones of how Fourier analysis on the cyclic group $\mathbb{Z}/n\mathbb{Z}$ works.

Here's another way to think about these results. If you think of the ζ^ks as being eigenvalues of some adjacency matrix, then the sums we've been looking at must count paths on some graph. And there is a very good candidate for such a graph if we think about what having a characteristic polynomial of $x^n = 1$ entails.

Definition Let G be a group and let S be a generating set for G. The Cayley graph of G with respect to S is the graph with vertex set G and edges (g, h) if and only if there is $s \in S$ such that $h = sg$. If $s \in G \Leftrightarrow s^{-1} \in G$, then the edges can be treated as undirected.

Note that if $G = \mathbb{Z}$ then we recover the one-dimensional walk with S the set of steps, and if $G = \mathbb{Z}^2$ then we recover the two-dimensional walk! Hence our discussion of lattice-path counting is really a special case of a more general phenomenon.

The graph we want is the Cayley graph C_n of $\mathbb{Z}/n\mathbb{Z}$ with $S = \{1\}$, otherwise known as a directed cycle. You can think of it as the n-gon that the n^{th} roots of unity trace out in the complex plane with $S = \{\zeta\}$. Unlike the other graphs we looked at, this one is easy to visualize. Counting paths on C_n is very easy. The number of paths of length l from 0 to a is 0 if $n \not| l - a$ and 1 if $n|l - a$. This sequence has generating function $\frac{x^a}{1-x^n}$; look familiar? The adjacency matrix of C_n is

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix},$$

which is a cyclic permutation matrix, which we'll denote $\mathbf{P}$. Powers of this matrix behave in the obvious way, and its eigenvalues are $1, \zeta, ...\zeta^{n-1}$, exactly as expected. What are its

eigenvectors? Well:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \zeta^a \\ \zeta^{2a} \\ \zeta^{3a} \\ \vdots \\ \zeta^{(n-1)a} \end{bmatrix} = \zeta^a \begin{bmatrix} 1 \\ \zeta^a \\ \zeta^{2a} \\ \zeta^{3a} \\ \vdots \\ \zeta^{(n-1)a} \end{bmatrix}$$

so we have the following beautiful result: the discrete Fourier transform matrix diagonalizes cyclic permutations! (One of the exercises considers a generalization of this result.) If you like, you can think of these eigenvectors as the "harmonics" of the n-gon. Now counting paths by diagonalizing gets us orthogonality, the same as before, as well as the partial fraction decomposition of $\frac{x^a}{1-x^n}$.

Actually, we don't even need to count paths: instead, we can take $S = \{1, -1\}$ and construct the undirected cycle UC_n, which is like the directed cycle but, well, undirected. The adjacency matrix of the undirected cycle is

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

which is $\mathbf{P} + \mathbf{P}^{n-1}$. Since $\mathbf{P}$ commutes with itself, it follows that the eigenvectors haven't changed and the eigenvalues are of the form $\zeta + \zeta^{n-1} = 2\cos\frac{2\pi i a}{n}$, which are real since the adjacency matrix is symmetric. But now we get orthogonality as a consequence of the spectral theorem.

Counting paths on UC_n is quite interesting: you can think of it as a finite approximation to counting paths on $\mathbb{Z}$, and so it becomes natural to think about, for example, the closed paths of length $2l$ from 0 to 0. (If n is even, there are no closed paths of odd length, so this question will be easier to answer.) As the number of vertices n gets large, if our intuition is correct we should expect this number to be asymptotic to $\binom{2l}{l}$, the number of ways to walk $2l$ steps in one dimension and get back to the origin. To show that this is the case, we'll use the following lemma.

Lemma 6.3. *Let* $\mathbf{A}$ *be an adjacency matrix of a graph. The total number of closed paths of length* l *is* $tr(\mathbf{A}^l)$.

This is by definition; that's exactly what the sum of the diagonal elements are. But we can take advantage of the fact that the trace of a matrix is also the sum of its eigenvalues, and since the graph we're looking at has this nice rotational symmetry, to find the number

of closed paths from a particular vertex to itself we just have to divide by n, so we want to compute

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1}\left(2\cos\frac{2\pi k}{n}\right)^{2l} &= \frac{1}{n}\sum_{k=0}^{n-1}(\zeta^k+\zeta^{-k})^{2l}\\ &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{i=0}^{2l}\binom{2l}{i}\zeta^{k(2i-2l)}\\ &= \sum_{i=0}^{2l}\binom{2l}{i}\frac{1}{n}\sum_{k=0}^{n-1}\zeta^{k(2i-2l)}.\end{aligned}$$

But we know how to compute sums like these; the sum over all k will just be zero unless $2i-2l\equiv 0 \bmod n$. Hence this sum will be equal to

$$\sum_{2i\equiv 2l \bmod n}\binom{2l}{i}.$$

This is definitely the correct answer from a combinatorial standpoint: we're just counting walks from $(0,0)$ to $(2l, 2i-l)$ for some i such that $2i\equiv 2l \bmod n$. Note that, exactly as expected, if $l<n$ then $\binom{2l}{l}$ is the lone term, since if we haven't walked more than l steps then our graph is indistinguishable from $\mathbb{Z}$. Here's the fun part: if we take the limit as $n\to\infty$, the above sum becomes a Riemann sum and hence an integral: in fact, we have

$$\int_0^1 (2\cos 2\pi x)^{2l}\,dx = \binom{2l}{l}.$$

Thus in some sense $2\cos 2\pi x, x\in[0,1)$ is an infinite set of eigenvalues, or spectrum, for the Cayley graph of $\mathbb{Z}$. We are also led to the philosophy that Fourier analysis on the circle is in some sense a "limit" of Fourier analysis on $\mathbb{Z}/n\mathbb{Z}$ as n is taken to infinity.

I'd like to end with a nontrivial application.

Example Suppose that the non-negative integers are written as a disjoint union of a finite number of arithmetic progressions with common differences $d_1\geq d_2\geq ...\geq d_n\geq 1$. Prove that $d_1=d_2$.

Proof. The generating functions proof is more or less obvious: this is possible if and only if

$$\frac{1}{1-x}=\sum_{k=1}^{n}\frac{x^{a_k}}{1-x^{d_k}}$$

where the arithmetic progression with common difference d_k has initial term a_k. If $d_1>d_2$, then the partial fraction decomposition of the RHS has terms involving d_1^{th} roots of unity that don't cancel with any of the other terms.

But there's a philosophy behind this solution that is worth thinking about. Subsets of the integers can be identified with their indicator functions, which is the function on $\mathbb{Z}$ that is 0 if an element isn't in the set and 1 if it is. And indicator functions have Fourier coefficients: in particular, large Fourier coefficients imply correlation with an arithmetic progression. This is a surprisingly powerful idea and has found some highly nontrivial applications in combinatorics, for example in the proof of *Roth's theorem*.

What it means for us is that there is a very nice interpretation of what we've done: let S be a disjoint union of a finite number of arithmetic differences. If $d_1 > d_2$, then the Fourier coefficients of period d_1 of the indicator function of S are nonzero. On the other hand, the indicator function of the non-negative integers is the constant function, so all of its Fourier coefficients are zero. $\square$

6.1 Exercises

1. Putnam 1983 A4: Let k be a positive integer and let $m = 6k - 1$. Let

$$S(m) = \sum_{j=1}^{2k-1} (-1)^{j+1} \binom{m}{3j-1}.$$

   Prove that $S(m)$ is never zero.

2. Putnam 1985 A6: If $p(x) = a_0 + a_1 x + ... + a_m x^m$ is a polynomial with real coefficients a_i, then set

$$\Gamma(p(x)) = a_0^2 + a_1^2 + ... + a_m^2.$$

   Let $F(x) = 3x^2 + 7x + 2$. Find, with proof, a polynomial $g(x)$ with real coefficients such that $g(0) = 1$ and $\Gamma(f(x)^n) = \Gamma(g(x)^n)$ for every integer $n \geq 1$.

3. Let $V_{n,m} = \text{span}\{(a_1 x_1 + a_2 x_2 + ... + a_n x_n)^m | (a_1, ...a_n) \in \mathbb{C}^n\}$ considered as a subspace of $\mathbb{C}[x_1, ...x_n]$. Show that it contains every monomial $x_1^{\lambda_1} x_2^{\lambda_2} ... x_n^{\lambda_n}$ where $\lambda_1 + ... + \lambda_n = m$, i.e. it is as large as it can be.

4. Let $U_n(y)$ denote the characteristic polynomial of the adjacency matrix of UC_n, with the conventions $U_0(y) = 1, U_1(y) = y$. Compute

$$\sum_{n=0}^{\infty} U_n(y) x^n.$$

   What are the roots of $U_n(y)$?

5. An $n \times n$ matrix is called *circulant* if its entries a_{ij} are a function of the value of $i - j \bmod n$.

   (a) Show that every circulant matrix is a sum of powers of the cyclic permutation matrix $\mathbf{P}$. Hence show that every circulant matrix is diagonalized by the DFT matrix. Conclude that the product of two circulant matrices is circulant and the circulant matrices form an algebra.

   (b) IMO 1979 #6: Let A and E be opposite vertices of an octagon. A frog starts at vertex A. From any vertex except E it jumps to one of the two adjacent vertices. When it reaches E it stops. Let a_n be the number of distinct paths of exactly n jumps ending at E. Prove that

$$a_{2n-1} = 0, a_{2n} = \frac{(2+\sqrt{2})^{n-1} - (2-\sqrt{2})^{n-1}}{\sqrt{2}}.$$

(c) Putnam 1985 B5: Let M_n be the $(2n+1) \times (2n+1)$ for which

$$(M_n)_{ij} = \begin{cases} 0 \text{ if } i = j \\ 1 \text{ if } i - j \equiv 1, 2, ...n \bmod (2n+1) \\ -1 \text{ if } i - j \equiv n+1, ...2n \bmod (2n+1). \end{cases}$$

Find the rank of M_n.

(d) Construct a graph on m vertices $0, 1, 2, ...m-1$ such that the number of walks of length l from 0 to a is

$$\sum_{k \equiv a \bmod m} \binom{l}{k}.$$

More generally, what kind of graphs have circulant adjacency matrices?

6. Let $G = (\mathbb{Z}/2\mathbb{Z})^n$ be the abelian group consisting of all n-tuples of 0s and 1s under componentwise addition mod2.

(a) Describe the Fourier transform on G: in other words, given a function $f : G \to \mathbb{R}$ (we don't need all of $\mathbb{C}$), decompose it into a sum of homomorphisms, or *characters*, of G, which are functions $\chi_u : G \to \mathbb{R}$ such that

$$\chi_u(ab) = \chi_u(a)\chi_u(b)$$

and

$$\frac{1}{2^n} \sum_{g \in G} \chi_u(g)\chi_v(g) = \begin{cases} 1 \text{ if } u = v \\ 0 \text{ otherwise} \end{cases}.$$

(b) Let S be the set of all elements of G with exactly one 1. What familiar geometric shape is the Cayley graph (G, S)? What are the eigenvalues and eigenvectors of its adjacency matrix?

(c) IMO 2008 #5: Let n and k be positive integers with $k \geq n$ and $k - n$ an even number. Let $2n$ lamps labelled 1, 2, ..., $2n$ be given, each of which can be either *on* or *off*. Initially all the lamps are off. We consider sequences of steps: at each step one of the lamps is switched (from on to off or from off to on).

Let N be the number of such sequences consisting of k steps and resulting in the state where lamps 1 through n are all on, and lamps $n+1$ through $2n$ are all off.

Let M be number of such sequences consisting of k steps, resulting in the state where lamps 1 through n are all on, and lamps $n+1$ through $2n$ are all off, but where none of the lamps $n+1$ through $2n$ is ever switched on.

Determine $\frac{N}{M}$.

7 Additional exercises

The following problems don't fit neatly into the above categories, but you should find yourself well-prepared to tackle them anyway!

1. Let $f(x) = a_0 + a_1 x + ...$ be a power series with integer coefficients, with $a_0 \neq 0$. Suppose that the power series expansion of $\frac{f'(x)}{f(x)}$ at $x = 0$ also has integer coefficients. Prove or disprove that $a_0 | a_n$ for all $n \geq 0$.

2. Putnam 1999 B3: Let $A = \{(x, y) : 0 \leq x, y < 1\}$. For $(x, y) \in A$, let

$$S(x, y) = \sum_{\frac{1}{2} \leq \frac{m}{n} \leq 2} x^m y^n$$

where m, n are positive integers. Evaluate

$$\lim_{(x,y)\to(1,1),(x,y)\in A} (1 - xy^2)(1 - x^2)S(x, y).$$

3. (a) Let p be a prime. Show that $(x + 1)^p \equiv x^p + 1 \bmod p$.

   (b) Let m, n be positive integers with base p representations

$$m = m_k p^k + m_{k-1} p^{k-1} + ... + m_0$$

$$n = n_k p^k + n_{k-1} p^{k-1} + ... + n_0.$$

   Prove *Lucas' theorem*:

$$\binom{m}{n} \equiv \prod_{i=0}^{k} \binom{m_i}{n_i} \bmod p.$$

   (c) Putnam 1983 A3: Let p be an odd prime and let

$$F(n) = 1 + 2n + 3n^2 + ... + (p - 1)n^{p-2}.$$

   Prove that if $a \not\equiv b \bmod p$ then $F(a) \not\equiv F(b) \bmod p$.

   (d) IMO Shortlist 1988 #2: Let n be a positive integer. Find the number of odd coefficients in

$$u_n(x) = (x^2 + x + 1)^n.$$

4. Let p be a prime and let $v_p(n)$ denote the greatest power of p that divides n. Let $s_p(n)$ denote the sum of the digits of n in base p.

   (a) Show that

$$v_p(n!) = \frac{n - s_p(n)}{p - 1}.$$

   Prove *Kummer's theorem*: $v_p\left(\binom{m}{n}\right)$ is the number of carries required to add n and $n - m$ in base p.

(b) Compute

$$\sum_{n\geq 0} s_p(n)x^n.$$

5. Let $T(n)$ denote the number of non-congruent non-degenerate triangles with integer side lengths adding up to n. For example, $T(3) = 1$. Show that

$$\sum_{n\geq 0} T(n)x^n = \frac{q^3}{(1-q^2)(1-q^3)(1-q^4)}.$$

6. The *exponential generating function* of a sequence a_n is given by

$$E_a(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}.$$

(a) Show that $E_a E_b = E_c$, where

$$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} = \sum_{S,T} a_{|S|} b_{|T|}$$

where S, T range over disjoint subsets of some n-element set.

(b) Suppose a_n, b_n are sequences such that $a_n = \sum_{k=0}^{n} \binom{n}{k} b_k$. Show that

$$b_n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} a_k.$$

Interpret this result in terms of Cartesian products of graphs.

(c) Show that the exponential generating function of a sequence satisfying a linear homogeneous recurrence satisfies a linear homogeneous ordinary differential equation.

(d) The Bell numbers B_n count the number of partitions of the set $\{1, 2, 3, ...n\}$, where neither the order of the partitions nor the order of the elements of the partitions matter. Show that

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

and hence compute $\sum_{n\geq 0} B_n \frac{x^n}{n!}$.

(e) Let I_n denote the number of permutations of n elements which have order 2. Show that

$$I_{n+1} = I_n + nI_{n-1}$$

and hence compute $\sum_{n\geq 0} I_n \frac{x^n}{n!}$.

(f) Prove the *fundamental theorem of exponential generating functions*: given a sequence $f(n)$, define $h(0) = 1$ and

$$h(|S|) = \sum_{k \geq 0} \sum_{B} f(|B_1|)...f(|B_k|)$$

where $B_1, ...B_k$ range over all partitions of S into disjoint subsets. Show that $E_h = \exp(E_f)$. Use this result to give a second proof of the result of the previous problem. What about the number of permutations with order 3?

(g) IMO 1987 #1: Let $p_n(k)$ be the number of permutations of $\{1, 2, ...n\}$ which have exactly k fixed points. Prove that

$$\sum_{k=1}^{n} k p_n(k) = n!.$$

(h) Putnam 2005 B6: Let S_n denote the set of all permutations of $\{1, 2, ...n\}$. For $\pi \in S_n$, let $\sigma(\pi) = 1$ if π is an even permutation and -1 if π is an odd permutation. Also, let $\nu(\pi)$ denote the number of fixed points of π. Show that

$$\sum_{\pi \in S_n} \frac{\sigma(\pi)}{\nu(\pi) + 1} = (-1)^{n+1} \frac{n}{n+1}.$$

(A permutation is called even or odd if it can be written as a product of, respectively, an even or odd number of transpositions. Equivalently, $\sigma(\pi)$ is the determinant of the permutation matrix associated to π.)

7. Given a function $a(n) : \mathbb{N} \to \mathbb{C}$, define its *Dirichlet generating function*

$$L(s, a) = \sum_{n \geq 1} \frac{a(n)}{n^s}.$$

where $s \in \mathbb{C}$.

(a) Suppose $a(n)$ is *multiplicative*, i.e. $a(mn) = a(m)a(n)$ if $\gcd(m, n) = 1$. Show that

$$L(s, a) = \prod_{p \text{ is prime}} \sum_{k \geq 0} \frac{a(p^k)}{p^{ks}},$$

i.e. L has an *Euler product.* Show a similar result in the case that $a(n)$ is *totally multiplicative* (the gcd assumption can be dropped), hence the case that $a(n) = 1$ (the Riemann zeta function).

(b) Show that the $a(n) = 1$ Euler product implies that

$$\sum_{p \text{ is prime}} \frac{1}{p}$$

diverges. Conclude that there are infinitely many primes and, moreover, that there does not exist a finite collection of polynomials of degree 2 or greater such that their values at the integers contain the primes.

(c) Show that $L(s, a)L(s, b) = L(s, c)$, where

$$c(n) = \prod_{d|n} a(n)b\left(\frac{n}{d}\right)$$

is the *Dirichlet convolution* of a and b, to be written $a * b$. Conclude that the Dirichlet convolution of two multiplicative functions is multiplicative.

(d) Show that the totient function $\varphi(n)$, the divisor function $d(n)$, and the Mobius function $\mu(n)$ (which is equal to 1 if n is a product of an even number of distinct primes, -1 if n is a product of an odd number of distinct primes, and 0 otherwise) are multiplicative. Compute their Dirichlet generating functions.

(e) Prove *Mobius inversion*: $b = a * 1$ if and only if $a = b * \mu$.

(f) Let a necklace of length l be a string of l beads that can be one of n colors, with the equivalence relation that two necklaces are the same if one can be obtained from the other by cyclic permutation. Compute the number of distinct necklaces of length l.

(g) For a given modulus m, define a *Dirichlet character* with respect to m to be a function $\chi : \mathbb{N} \to \mathbb{C}$ such that $\chi(n) = 0$ if $(n, m) > 1$, $\chi(n) = \chi(n + m)$, and χ is a homomorphism from $(\mathbb{Z}/m\mathbb{Z})^*$ to $\mathbb{C}$, i.e. it is a group character of $\mathbb{Z}/\varphi(m)/\mathbb{Z}$. Among other things, this implies that χ is totally multiplicative. For $m = 3, 4$, use the values of $L(s, \chi)$ as χ ranges over all characters to show that

$$\sum_{p \text{ is prime}, p \equiv a \bmod m} \frac{1}{p}$$

diverges, where $\gcd(a, m) = 1$. (The proof is substantially harder for general m; the result is known as Dirichlet's theorem.)

(h) Let $r_2(n)$ denote the number of ways n can be written as a sum of two squares, where order and sign are both taken into account, and let $\chi(n)$ be the Dirichlet character with respect to 4 given by $\chi(3) = -1$. Show that

$$r_2(n) = \sum_{d|n} \chi(d).$$

8. Given a semigroup G (a group which is not required to have inverses or an identity), its *semigroup algebra* $\mathbb{C}[G]$ is the algebra of formal linear combinations of elements of G with multiplication given by the semigroup operation, i.e. $g(ah_1 + bh_2) = a(gh_1) + b(gh_2)$ where $a, b \in \mathbb{C}$ and $g, h_1, h_2 \in \mathbb{C}$. If G is a group, $\mathbb{C}[G]$ is called the group algebra and is also known as the *regular representation* of G.

   (a) Show that the algebra of ordinary generating functions in n variables is (isomorphic to) the semigroup algebra of the free semigroup on n generators, i.e. for $n = 1$ the non-negative integers under addition. Also show that the algebra of Dirichlet generating functions is (isomorphic to) the semigroup algebra of the positive integers under multiplication.

   (b) Show that the algebra of $n \times n$ circulant matrices is (isomorphic to) the group algebra of $\mathbb{Z}/n\mathbb{Z}$, hence to $\mathbb{C}[x]/(x^n - 1)$. More generally, show how to get a Cayley graph out of an element of a group algebra. Hence rework the last exercise in the last section by figuring out an easy way to write down the group algebra of $(\mathbb{Z}/2\mathbb{Z})^n$.

   (c) Show that if G is a finite abelian group, $\mathbb{C}[G]$ has a basis in which all of its elements, considered as linear operators on $\mathbb{C}[G]$, are diagonal. Describe the Fourier transform on G.

References

[1] F. Bergeron, G. Labelle, P. Leroux. Combinatorial Species and Tree-like Structures. Cambridge University Press, 1998.

[2] G. C. Rota. Finite Operator Calculus. Academic Press, Inc., 1975.

[3] R. Stanley. Enumerative Combinatorics, vol. I. Monterey: Wadsworth and Brooks/Cole, 1987.

[4] R. Stanley. Enumerative Combinatorics, vol. II. Cambridge University Press, 2001.

[5] E. M. Stein, R. Shakarchi. Fourier Analysis: an Introduction. Princeton University Press, Princeton, NJ, 2003.

[6] H. Wilf. Generatingfunctionology. Academic Press, Inc., 1994. http://www.math.upenn.edu/ wilf/DownldGF.html.