Assorted notes concerning generating functions[1]

Shagnik Das

Introduction

We live in the twenty-first[2] century, a fact that can be both a blessing and a curse in many ways. For instance, while we have airplanes, we also have airline food. While we have televisions, we also have reality TV. One could go on in this manner, but I shall not, and will instead get straight to the point.

One of the great advantages of being a twenty-first centurion is the fact that the human race has made a great deal of progress in the millenia before our own existence, with our ancestors doing much of the hard work for us. They moved from trees to caves to high-rise apartment buildings, learned to walk upright and ride Segways, developed languages and social networks, and, perhaps most crucially[3], discovered Mathematics. Somehow, while achieving all these amazing feats, they also managed to leave us timeless words of wisdom to guide us in our times of need.

For instance, whenever one gets comfortable and complacent in life, one would do well to recall an Ancient Greek saying:

> "The unexamined life is not worth living." — Socrates.

From time to time, according to Socrates, one ought to pause and reflect upon the course one is following, and the person one has become. When doing so, the following rule of thumb is useful:

> "Be the person your dog thinks you are." — Person on the Internet.[4]

It is therefore with some shame that, having looked back over the last week, I must admit my dog would not have been very impressed with my teaching.[5] There are a number of things that I did not find time to fit into lecture, or did not explain to my satisfaction. I have thus prepared these notes to go over some of the material in more depth, and hope that this will help clear up any confusion I may have caused.

Partial fraction decompositions

To motivate the study of generating functions, we showed how one could use them to solve constant-coefficient linear homogeneous recurrence relations. Using the recurrence

[1] Or: Things I wish I had said in lecture but didn't

[2] I suppose there is a chance of this document surviving beyond the dawn of the twenty-second century, but I do not foresee any interest in reading it lasting longer than a few months, and so I shall permit myself to make a statement of limited veracity.

[3] I imagine there are people who might disagree with this statement, but I also imagine they would not be reading this in the first place.

[4] Who, it transpires, may have been quoting J. W. Stephens.

[5] To be fair, though, I do not think he has ever thought much of my teaching. In all these years, I have only managed to teach him how to shake paws. My attempts to teach him new tricks only lead to my rolling around on the carpet while he looks at me, wondering what he did to deserve such an idiotic human.

relation, we found that the generating function for such a sequence must be a rational function[6]. We then used the partial fraction decomposition to simplify the expression, which enabled us to extract the coefficients of the generating function and find a closed formula for the terms of the sequence.

The idea behind the partial fraction decomposition is to take a rational function, which could involve rather complicated polynomials in the numerator and denominator, and express it as a sum of much simpler functions.[7] One typically first meets partial fractions when learning integration, as one can then integrate any rational function by simply antidifferentiating the simpler partial fraction summands.

However, if my own experience is typical, one is usually just taught an algorithm for determining what the coefficients of the partial fraction decomposition are, all the while implicitly assuming that such a decomposition exists. Some of you were asking for a proof[8] that one can always decompose a rational function into partial fractions. Here we shall prove the existence of such a decomposition over the complex numbers.[11]

Proposition 1. *Let $P(x)$ and $Q(x)$ be polynomials in $\mathbb{C}[x]$ such that $\deg(P) < \deg(Q)$ and $P(x)$ and $Q(x)$ have no common roots. If $Q(x)$ has distinct roots λ_i, $1 \leq i \leq r$, with respectively multiplicities $m_i \geq 1$, then there are constants $\{\alpha_{i,j} : 1 \leq i \leq r, 1 \leq j \leq m_i\}$ such that*

$$\frac{P(x)}{Q(x)} = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \frac{\alpha_{i,j}}{(x-\lambda_i)^j}.$$

Note that the conditions placed on the polynomials $P(x)$ and $Q(x)$ do not cause any loss of generality. Indeed, if $\deg(P) \geq \deg(Q)$, then one can perform polynomial division to write $P(x) = A(x)Q(x) + R(x)$, where $A(x)$ and $R(x)$ are polynomials such that $\deg(A) = \deg(P) - \deg(Q)$ and $\deg(R) < \deg(Q)$. We then have $\frac{P(x)}{Q(x)} = A(x) + \frac{R(x)}{Q(x)}$, and can apply Proposition 1 to $\frac{R(x)}{Q(x)}$. On the other hand, if $P(x)$ and $Q(x)$ share a common root $\lambda \in \mathbb{C}$, then $P(x) = (x-\lambda)\hat{P}(x)$ and $Q(x) = (x-\lambda)\hat{Q}(x)$ for some polynomials $\hat{P}(x)$ and $\hat{Q}(x)$, and so one can apply the proposition to $\frac{\hat{P}(x)}{\hat{Q}(x)}$ instead.

[6]That is, one polynomial divided by another.

[7]This is clearly a sensible thing to want to do, and it is also an ancient pursuit. In the Middle Kingdom of Egypt, for instance, the system of Egyptian fractions was developed, where a rational number is written as the sum of reciprocals of distinct natural numbers. It turns out that every positive rational number can be written in this way; to convince you of this fact, I shall provide you with a random example: $\frac{7}{23} = 4^{-1} + 19^{-1} + 583^{-1} + 1019084^{-1}$. Despite having been around for thousands of years, plenty of open problems about Egyptian fractions remain. For example, the Erdős–Straus conjecture from 1948 postulates that for every $n \geq 2$, there are $a, b, c \in \mathbb{N}$ such that $\frac{4}{n} = a^{-1} + b^{-1} + c^{-1}$. While this is known to be true for all $n \leq 10^{17}$, a general solution has evaded us thus far.

[8]I must say, this insistence on rigour is the kind of thing that restores one's faith[9] in the new generation. One should always be wary of unproven statements.[10]

[9]Faith that may have been lost after seeing what Cartoon Network has become, or after witnessing the rise of T20 cricket at the expense of Test cricket.

[10]To which there is the excellent response: "Why?"

[11]Things are simpler in this setting, since every polynomial can be split into linear factors. In the more general case, a partial fraction decomposition still exists, but one needs the notion of irreducible polynomials.

To prove Proposition 1, we will make repeated use of the following lemma.

Lemma 2. *Suppose $\lambda \in \mathbb{C}$ and $m \in \mathbb{N}$, and let $N(x)$ and $D(x)$ be polynomials in $\mathbb{C}[x]$ such that $D(\lambda) \neq 0$ and $\deg(N) < \deg(D) + m$. There are $\alpha \in \mathbb{C}$ and $\tilde{N}(x) \in \mathbb{C}[x]$ such that $\deg(\tilde{N}) < \deg(D) + m - 1$ and*

$$\frac{N(x)}{(x-\lambda)^m D(x)} = \frac{\tilde{N}(x)}{(x-\lambda)^{m-1} D(x)} + \frac{\alpha}{(x-\lambda)^m}.$$

Proof. We first perform a bit of polynomial division, and find polynomials $\hat{N}(x)$ and $\hat{D}(x)$ and constants $\beta, \gamma \in \mathbb{C}$ such that

$$N(x) = (x-\lambda)\hat{N}(x) + \beta \quad \text{and} \quad D(x) = (x-\lambda)\hat{D}(x) + \gamma.$$

Note that $\gamma = D(\lambda) \neq 0$ and $\beta = N(\lambda)$, $\deg(\hat{N}) = \deg(N) - 1$ and $\deg(\hat{D}) = \deg(D) - 1$.

We can then write

$$\frac{N(x)}{(x-\lambda)^m D(x)} = \frac{(x-\lambda)\hat{N}(x) + \beta}{(x-\lambda)^m D(x)} = \frac{\hat{N}(x)}{(x-\lambda)^{m-1} D(x)} + \frac{\beta}{(x-\lambda)^m D(x)}.$$

This is almost what we want, except the denominator in the second term involves $D(x)$. Multiplying our equation for $D(x)$ by β/γ, we find $\beta = \frac{\beta}{\gamma} D(x) - \frac{\beta}{\gamma}(x-\lambda)\hat{D}(x)$, and so

$$\frac{N(x)}{(x-\lambda)^m D(x)} = \frac{\hat{N}(x)}{(x-\lambda)^{m-1} D(x)} + \frac{\beta/\gamma}{(x-\lambda)^m} - \frac{\beta \hat{D}(x)/\gamma}{(x-\lambda)^{m-1} D(x)}.$$

Setting $\alpha = \frac{\beta}{\gamma} = \frac{N(\lambda)}{D(\lambda)}$ and rearranging, we obtain

$$\frac{N(x)}{(x-\lambda)^m D(x)} = \frac{\tilde{N}(x)}{(x-\lambda)^{m-1} D(x)} + \frac{\alpha}{(x-\lambda)^m},$$

where $\tilde{N}(x) = \hat{N}(x) - \alpha \hat{D}(x)$, and hence

$$\deg(\tilde{N}) = \max(\deg(\hat{N}), \deg(\hat{D})) = \max(\deg(N), \deg(D)) - 1 < \deg(D) + m - 1. \quad \square$$

We can now prove the main result.

Proof of Proposition 1. We prove the proposition by induction on $\deg(Q)$.

For the base case, suppose $\deg(Q) = 1$, so $Q(x) = x - \lambda_1$ for some $\lambda_1 \in \mathbb{C}$. Since $\deg(P) < \deg(Q)$, we must have $\deg(P) = 0$, and so $P(x) = \alpha_{1,1}$ for some $\alpha_{1,1} \in \mathbb{C}$. Thus $\frac{P(x)}{Q(x)} = \frac{\alpha_{1,1}}{x - \lambda_1}$, as required.

For the induction step, we may assume $\deg(Q) \geq 2$. By multiplying numerator and denimator by a constant, we may further assume Q is monic[12]. Hence $Q(x) = \prod_{i=1}^{r} (x - \lambda_i)^{m_i}$. Let $N(x) = P(x)$, and set $D(x) = \prod_{i=1}^{r-1} (x - \lambda_i)^{m_i}$. We thus have

[12] That is, its leading coefficient is 1.

$\frac{P(x)}{Q(x)} = \frac{N(x)}{(x-\lambda_r)^{m_r} D(x)}$, with $D(\lambda_r) \neq 0$. Applying Lemma 2, we find some α_{r,m_r} and polynomial $\tilde{N}(x)$ such that

$$\frac{P(x)}{Q(x)} = \frac{\tilde{N}(x)}{(x-\lambda_r)^{m_r-1} D(x)} + \frac{\alpha_{r,m_r}}{(x-\lambda_r)^{m_r}}.$$

Cancel any common factors between $\tilde{N}(x)$ and $(x-\lambda_r)^{m_r} D(x)$, so that we have

$$\frac{\tilde{N}(x)}{(x-\lambda_r)^{m_r-1} D(x)} = \frac{\tilde{P}(x)}{\tilde{Q}(x)},$$

where $\tilde{P}(x)$ and $\tilde{Q}(x)$ have no common roots. Note that the roots of $\tilde{Q}(x)$ are roots of $Q(x)$ as well, perhaps with reduced multiplicity. We thus have $\deg(\tilde{Q}) < \deg Q$, and so, by the induction hypothesis, we find constants $\alpha_{i,j}$ such that

$$\frac{\tilde{P}(x)}{\tilde{Q}(x)} = \sum_{i=1}^{r-1} \sum_{j=1}^{m_i} \frac{\alpha_{i,j}}{(x-\lambda_i)^j} + \sum_{j=1}^{m_r-1} \frac{\alpha_{r,j}}{(x-\lambda_r)^j}.$$

Hence we obtain the partial fraction decomposition

$$\frac{P(x)}{Q(x)} = \frac{\tilde{P}(x)}{\tilde{Q}(x)} + \frac{\alpha_{r,m_r}}{(x-\lambda_r)^{m_r}} = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \frac{\alpha_{i,j}}{(x-\lambda_i)^j}. \qquad \square$$

Note that the proof of Proposition 1 is algorithmic; that is, it tells you how to compute the coefficients $\alpha_{i,j}$. However, this computation requires several polynomial divisions, which can be a bit time-consuming[13]. In practice, now that we know the decomposition exists, it is easier to find the coefficients by multiplying through by $Q(x)$ and then obtaining a linear system of equations to solve. That is,

$$P(x) = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \frac{\alpha_{i,j}}{(x-\lambda_i)^j} Q(x) = \sum_{i=1}^{r} \sum_{j=1}^{m_i} \left(\alpha_{i,j} (x-\lambda_i)^{m_i-j} \prod_{k \neq i} (x-\lambda_k)^{m_k} \right).$$

The left-hand side is a polynomial of degree at most $\deg(Q)-1$, while the right-hand side is a polynomial of degree $\deg(Q)-1$. In order to have equality, the coefficients of x^n must be equal on both sides for all $0 \leq n \leq \deg(Q)-1$. This gives a linear system of $\deg(Q)$ equations in $\deg(Q)$ variables,[14] which we can then solve.[15]

Writing down what these equations are involves multiplying out the polynomials on the right-hand side, but that is simpler than polynomial division. Another simplification that can be made is to plug in $\deg(Q)$ distinct values for x into the equality above, which will lead to $\deg(Q)$ linear equations in the $\alpha_{i,j}$ that can also be solved.

[13] And, unless you are an excellent calculator, error-strewn.

[14] Since $\deg(Q) = \sum_{i=1}^{r} m_i$, which is the number of $\alpha_{i,j}$ variables we have.

[15] We know there is a solution, because we have proven that such a decomposition exists.

The average number of cycles in permutations

Recall that we defined the Stirling number of the first kind,[16] $s_{n,k}$, as the number of permutations $\pi \in S_n$ composed of exactly k cycles. In the homework you were asked to compute the average number of cycles contained in permutations in S_n. This could be done through a double-counting argument.

Proposition 3. *The average number of cycles in permutations in S_n is*

$$H_n = \sum_{k=1}^{n} \frac{1}{k}.$$

Proof. The average number of cycles in permutations can be computed by summing up over every permutation the number of cycles C it contains, then dividing by the total number of permutations. That is,

$$\frac{1}{n!} \sum_{\pi \in S_n} |\{\text{cycle } C \subseteq \pi\}| = \frac{1}{n!} \sum_{\pi \in S_n} \sum_{C \subseteq \pi} 1.$$

We now exchange the order of summation, and define, for some cycle C and permutation π,

$$\mathbf{1}_{\{C \subseteq \pi\}} = \begin{cases} 1 & \text{if } C \subseteq \pi \\ 0 & \text{otherwise} \end{cases}.$$

Thus

$$\frac{1}{n!} \sum_{\pi \in S_n} \sum_{C \subseteq \pi} 1 = \frac{1}{n!} \sum_{C} \sum_{\pi \in S_n} \mathbf{1}_{\{C \subseteq \pi\}},$$

where the first sum on the right-hand side is over all possible cycles appearing in permutations in S_n. Note that if we fix the length of the cycle to be k, for some $1 \leq k \leq n$, there are $\binom{n}{k}$ ways to choose the elements in the cycle, and $(k-1)!$ distinct ways to order them. Moreover, each cycle of length k is contained in $(n-k)!$ permutations in S_n, since the remaining $n-k$ elements can be permuted arbitrarily. Hence the average number of cycles is equal to

$$\frac{1}{n!} \sum_{k=1}^{n} \binom{n}{k} (k-1)!(n-k)! = \frac{1}{n!} \sum_{k=1}^{n} \frac{n!}{k!(n-k)!} (k-1)!(n-k)! = \sum_{k=1}^{n} \frac{1}{k} = H_n,$$

as claimed. □

While this is certainly a very nice proof — the calculation seems hopeless before one is struck by the idea to double-count — there is another delightful proof that uses generating functions. In order to use this, we need to recall one fact from lectures, which we had proven by induction.

[16] These are sometimes more descriptively called the Stirling cycle numbers, while the Stirling numbers of the second kind are called the Stirling set numbers.

Proposition 4. *For all $n \geq 0$,*

$$\sum_{k=0}^{n}(-1)^{n-k}s_{n,k}x^k = x^{\underline{n}} = \prod_{i=0}^{n-1}(x-i).$$

Given this, one can compute the average number of cycles in permutations as follows.

Proof of Proposition 3. Grouping permutations by the number of cycles they contain, we can write the average number of cycles in permutations $\pi \in S_n$ as

$$\frac{1}{n!}\sum_{\pi\in S_n}|\{\text{cycle } C \subseteq \pi\}| = \frac{1}{n!}\sum_{k=0}^{n} s_{n,k}\cdot k = \frac{1}{n!}\sum_{k\geq 1} s_{n,k}\cdot k.$$

Now define the generating function $S_n(x) = \sum_{k\geq 0} s_{n,k}x^k$.[17] This looks similar to the sum above, except we are missing the factor k. This can be introduced by taking a derivative, since

$$S_n'(x) = \frac{d}{dx}S_n(x) = \sum_{k\geq 0}\frac{d}{dx}s_{n,k}x^k = \sum_{k\geq 1} s_{n,k}\cdot kx^{k-1}.$$

To remove the x^{k-1} factor, we simply substitute $x = 1$.[18] Hence the average number of cycles in permutations in S_n is given by $\frac{1}{n!}S_n'(1)$.

In order to evaluate $S_n'(1)$, we must find a closed-form expression for $S_n(x)$. By Proposition 4, we have

$$\begin{aligned} S_n(-x) &= \sum_{k\geq 0}(-1)^k s_{n,k}x^k = (-1)^n\sum_{k=0}^{n}(-1)^{n-k}s_{n,k}x^k = (-1)^n x^{\underline{n}} \\ &= (-1)^n\prod_{i=0}^{n-1}(x-i) = \prod_{i=0}^{n-1}(-x+i) = (-x+n-1)^{\underline{n}}, \end{aligned}$$

and so $S_n(x) = (x+n-1)^{\underline{n}} = \prod_{i=0}^{n-1}(x+i)$.

Using the product rule for derivatives,

$$S_n'(x) = \sum_{i=0}^{n-1}\prod_{0\leq j\leq n-1, j\neq i}(x+j) = S_n(x)\sum_{i=0}^{n-1}\frac{1}{x+i}.$$

Hence the average number of cycles in permutations in S_n is

$$\frac{1}{n!}S_n'(1) = \frac{1}{n!}(1+n-1)^{\underline{n}}\sum_{i=0}^{n-1}\frac{1}{1+i} = \frac{n^{\underline{n}}}{n!}\sum_{i=1}^{n}\frac{1}{i} = \frac{n!}{n!}\sum_{i=1}^{n}\frac{1}{i} = H_n. \qquad \square$$

In general, if we have some finite set $\mathcal{A}$ of objects, and let $(a_n)_{n\geq 0}$ be the sequence counting the number of objects of size[19] n, then this proof shows that the average size of the objects in the set is given by $\frac{A'(1)}{A(1)}$, where $A(x) = \sum_{n\geq 0} a_n x^n$.[20]

[17] Note that n is fixed, while k is the index of the sequence and generating function.

[18] Since $s_{n,k} = 0$ for all $k > n$, this generating function is given by a finite sum, and hence converges for all $x \in \mathbb{C}$. We may thus meaningfully substitute any value for x.

[19] For some abstract notion of "size".

[20] We are using the fact that $A(1) = \sum_{n\geq 0} a_n$ counts the total number of objects in the set.

Multiplying generating functions and convolutions of sequences

We have already seen that generating functions provide a very convenient framework for handling sequences, since natural combinatorial operations on sequences have convenient analytic counterparts in the world of generating functions. One of the most useful such connections concerns the multiplication of generating functions, and in this section we will explore what this does combinatorially to the sequences.

Many of the sequences we deal with enumerate some graded set, a term we now explain. Suppose we have some set $\mathcal{A}$ of objects, together with some (abstract) notion of size, $|\cdot| : \mathcal{A} \to \mathbb{N} \cup \{0\}$. We can then define a counting sequence $(a_n)_{n\geq 0}$, where a_n is the number of objects $\alpha \in \mathcal{A}$ with $|\alpha| = n$.

For example, the all-1 sequence $a_n \equiv 1$ is the counting sequence for the graded set $\mathbb{N} \cup \{0\}$, where the size is given by the identity function; $|n| = n$. For some $r \in \mathbb{N}$, the binomial sequence $a_n = \binom{r}{n}$ is the counting sequence for $2^{[r]}$, the set of all subsets of the r-element set $[r]$, with the natural notion of size. Even the Fibonacci sequence can be thought of as such a counting sequence: imagine we took a photo of all the rabbits every month. The set $\mathcal{A}$ would then be the set of photographs, with the "size" of the photograph being the month it was taken.[21]

If a sequence $(a_n)_{n\geq 0}$ is the counting sequence for some set $\mathcal{A}$ with size function $|\cdot|$, its generating function then has the nice representation

$$A(x) = \sum_{n\geq 0} a_n x^n = \sum_{\alpha\in\mathcal{A}} x^{|\alpha|}.$$

Many operations on generating functions have natural interpretations if we think of their sequences as counting sequences. For instance, suppose $\mathcal{A}$ and $\mathcal{B}$ are two disjoint sets with size functions $|\cdot|_{\mathcal{A}}$ and $|\cdot|_{\mathcal{B}}$ respectively. Let $A(x)$ and $B(x)$ be the generating functions for their counting sequences. The generating function of the counting sequence of the union $\mathcal{C} = \mathcal{A} \sqcup \mathcal{B}$, with size function

$$|\gamma|_{\mathcal{C}} = \begin{cases} |\alpha|_{\mathcal{A}} & \text{if } \gamma = \alpha \in \mathcal{A} \\ |\beta|_{\mathcal{B}} & \text{if } \gamma = \beta \in \mathcal{B} \end{cases},$$

is then $C(x) = \sum_{\gamma\in\mathcal{C}} x^{|\gamma|_{\mathcal{C}}} = \sum_{\alpha\in\mathcal{A}} x^{|\alpha|_{\mathcal{A}}} + \sum_{\beta\in\mathcal{B}} x^{|\beta|_{\mathcal{B}}} = A(x) + B(x)$.

Multiplication of generating functions also has a natural interpretation of this kind, but this time as the counting sequence of the Cartesian product of the underlying sets. As before, let $\mathcal{A}$ and $\mathcal{B}$ be two sets, not necessarily disjoint, with size functions $|\cdot|_{\mathcal{A}}$ and $|\cdot|_{\mathcal{B}}$. This time we take the Cartesian product $\mathcal{C} = \mathcal{A} \times \mathcal{B} = \{(\alpha, \beta) : \alpha \in \mathcal{A}, \beta \in \mathcal{B}\}$, which is the set of all pairs of an object from $\mathcal{A}$ and an object from $\mathcal{B}$. We equip $\mathcal{C}$ with the size function $|\cdot|_{\mathcal{C}}$ given by

$$|(\alpha, \beta)|_{\mathcal{C}} = |\alpha|_{\mathcal{A}} + |\beta|_{\mathcal{B}},$$

[21] Indeed, since the set $\mathcal{A}$ and the size function $|\cdot|$ can be defined abstractly, every sequence $(a_n)_{n\geq 0}$ is a counting sequence. Simply add a_n objects to $\mathcal{A}$ for every n, whose size we define to be n.

so the size of a pair is the sum of the sizes of its components, a natural definition. If $C(x)$ is the generating function for $\mathcal{C}$, we have

$$\begin{aligned} C(x) &= \sum_{\gamma \in \mathcal{C}} x^{|\gamma|_{\mathcal{C}}} = \sum_{(\alpha,\beta) \in \mathcal{A}\times\mathcal{B}} x^{|(\alpha,\beta)|_{\mathcal{C}}} \\ &= \sum_{\alpha\in\mathcal{A}} \sum_{\beta\in\mathcal{B}} x^{|\alpha|_{\mathcal{A}}+|\beta|_{\mathcal{B}}} = \left(\sum_{\alpha\in\mathcal{A}} x^{|\alpha|_{\mathcal{A}}}\right)\left(\sum_{\beta\in\mathcal{B}} x^{|\beta|_{\mathcal{B}}}\right) = A(x)B(x). \end{aligned}$$

This gives some combinatorial meaning to the analytic operation of multiplying two generating functions. However, it remains to determine what the sequence $(c_n)_{n\geq 0}$ corresponding to $C(x)$ is in terms of $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$.

Claim 5. *If $A(x) = \sum_{n\geq 0} a_n x^n$ and $B(x) = \sum_{n\geq 0} b_n x^n$, and $C(x) = A(x)B(x) = \sum_{n\geq 0} c_n x^n$, then for all $n \geq 0$, we have*

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

Proof. Let $(a_n)_{n\geq 0}$ be the counting sequence[22] for the set $\mathcal{A}$ and $(b_n)_{n\geq 0}$ the counting sequence of $\mathcal{B}$. We know $(c_n)_{n\geq 0}$ is the counting sequence for the set $\mathcal{A}\times\mathcal{B}$, and so c_n is the number of pairs (α,β) with total size n.

We can first choose how much of the size comes from the α component. This can be any integer k with $0 \leq k \leq n$. The β component must then have size $n-k$.

There are, by definition, a_k elements $\alpha \in \mathcal{A}$ of size k. These can each be paired with any of b_{n-k} elements $\beta \in \mathcal{B}$ of size $n-k$. This gives a total of $a_k b_{n-k}$ pairs $(\alpha,\beta) \in \mathcal{A}\times\mathcal{B}$ where the first component has size k and the second has size $n-k$. Summing over the different choices for k, we find there are $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ pairs $(\alpha,\beta) \in \mathcal{A}\times\mathcal{B}$ with combined size n, as claimed. □

This formula, $c_n = \sum_{k=0}^{n} a_k b_{n-k}$, is called the *convolution* of the sequences $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$, and is an important operation in several areas of mathematics. It can also be derived directly from the product of the two generating functions. When we multiply two generating functions, we multiply every pair of terms, and then group together terms with the same power of x.[23] We obtain an x^n term in $A(x)B(x)$ by pairing the x^k term from $A(x)$ with the x^{n-k} term from $B(x)$, which has coefficient $a_k b_{n-k}$. Summing these up for all choices of $0 \leq k \leq n$ gives $c_n = \sum_{k=0}^{n} a_k b_{n-k}$.[24]

We close this section by giving a few examples to show when it is appropriate to take the product of generating functions.

[22] By Footnote 21, every sequence can be thought of as a counting sequence.

[23] One needs to be a little careful, since the coefficients should always require only a finite computation. However, since the x^n term can only arise from terms of degree at most n, we do not require any infinite sums.

[24] For an alternative exposition, consider $A(x)B(y) = \left(\sum_{k\geq 0} a_k x^k\right)\left(\sum_{\ell\geq 0} b_\ell y^\ell\right) = \sum_{k,\ell\geq 0} a_k b_\ell x^k y^\ell$. If we then make the substitution $y = x$, we get $C(x) = A(x)B(x) = \sum_{k,\ell\geq 0} a_k b_\ell x^{k+\ell}$. We can now change variables by defining $n = k+\ell$, which gives $C(x) = \sum_{k,n\geq 0} a_k b_{n-k} x^n = \sum_{n\geq 0}\left(\sum_{k=0}^{n} a_k b_{n-k}\right) x^n$. If we write $C(x) = \sum_{n\geq 0} c_n x^n$, it follows that $c_n = \sum_{k=0}^{n} a_k b_{n-k}$.

Example 1: A pizza party Let us start with everyone's favourite mathematical construct: pizza.[25] Suppose you wish to celebrate some occasion[26] with a pizza party. You thus have to order pizza and get some drinks.

Suppose a plain large pizza costs €5, with €2 for every topping. There are r distinct toppings you can choose from. You will of course only serve tea with your pizza, and every tea bag costs €1. How many ways are there to spend €n on the party?

To solve this problem, we note that the party budget consists of two components — the pizza and the tea. Hence, if $\mathcal{P}$ is the set of possible pizzas, and $\mathcal{T}$ is the set of possible tea orders, the set of possible pizza parties is given by $\mathcal{P} \times \mathcal{T}$. We can equip all of these sets with the "size" measure given by how much things cost. The generating function for parties costing €n is then the product of the corresponding generating functions for pizzas and for tea.

For every n, we know there are $\binom{r}{n}$ pizzas costing €$(2n+5)$, as we can choose any n of the r toppings. Thus if p_n is the sequence counting the number of pizzas costing €n, the sequence is obtained by taking the sequence $(\binom{r}{n})_{n\geq 0}$, spacing it out by a factor of 2 $[n \mapsto 2n]$, and shifting everything 5 places to the right $[2n \mapsto 2n+5]$. Since the sequence $(\binom{r}{n})_{n\geq 0}$ has generating function $(1+x)^r$, it follows that $\mathcal{P}$ has $P(x) = x^5(1+x^2)^r$ as its generating function.

Now we consider the set $\mathcal{T}$. There is exactly one way to spend €n on tea: buy n tea bags. Hence the counting sequence is $t_n \equiv 1$, and hence we have the generating function $T(x) = (1-x)^{-1}$.

Thus the generating function $C(x) = P(x)T(x) = x^5(1+x^2)^r(1-x)^{-1} = \sum_{n\geq 0} c_n x^n$ describes the number of ways of spending €n on the party.

If you are fortunate enough to have friends, pizza and tea are all you need for a pizza party. However, there are those of us who lack such social support, and hence also have to pay people to attend our pizza parties.[27] If $G(x)$ is the generating function whose coefficients, g_n, describe how many ways there are of hiring guests for €n, then $P(x)T(x)G(x)$ is the generating function describing the total number of ways to spend €n on the party.

In general, if for $1 \leq i \leq m$ we have graded sets $\mathcal{A}_i$ with some notion of size, with corresponding generating functions $A_i(x) = \sum_{n\geq 0} a_{i,n} x^n$, then $A(x) = \prod_i A_i(x)$ is the generating function for

$$\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_m = \{(\alpha_1, \alpha_2, \ldots, \alpha_m) : \alpha_i \in \mathcal{A}_i \text{ for all } 1 \leq i \leq m\}.$$

The coefficients of $A(x) = \sum_{n\geq 0} a_n x^n$ are given by $a_n = \sum_{(k_1,\ldots,k_m):\sum_i k_i = n} \prod_{i=1}^m a_{i,k_i}$. Here a_n counts the number of m-tuples $(\alpha_1, \alpha_2, \ldots, \alpha_m) \in \mathcal{A}$ with total size n.

[25] Best theorem of all time: the volume of a pizza with radius z and thickness a is $(pi)zza$.

[26] It could be a birthday, a wedding, or the completion of your Discrete Maths I homework — the reason for ordering pizza does not have much bearing on the mathematics.

[27] One of the great (football) World Cup stories: when North Korea took part in the 2010 World Cup in South Africa, her citizens were not allowed to travel to watch their team in action. Instead, North Korea paid a small group of Chinese actors to support the team.[28]

[28] I think I read this somewhere, but am not certain it is true. At any rate, one should not let the truth get in the way of a good story.

Example 2: The Binomial Theorem In the previous example, we used the Binomial Theorem, in the sense that $B(x) = \sum_{n\geq 0} \binom{r}{n} x^n = (1+x)^r$ is the generating function for the sequence $(\binom{r}{n})_{n\geq 0}$. This generating function is a product, and indeed one can derive it using this general framework as well.

Consider the set $[r] = \{1, 2, \ldots, r\}$, which is a set of r elements. The binomial coefficient $\binom{r}{n}$ counts the number of subsets of $[r]$ of size n. This sequence is thus the counting sequence for the power set of $[r]$; that is, the set of all subsets of $[r]$, $\mathcal{S} = \{S : S \subseteq [r]\} = 2^{[r]}$. In this formulation, $\mathcal{S}$ does not look like a Cartesian product of sets.

However, consider how we form a subset $S \subseteq [r]$. For each element i, we must decide whether $S \cap \{i\} = \emptyset$ or $\{i\}$. $2^{[r]}$ is thus the Cartesian product of the smaller power sets $2^{\{i\}}$. Each of these smaller power sets consists of two elements objects, the empty set $\emptyset$ of size 0, and the singleton $\{i\}$ of size 1. Hence, for each i, $2^{\{i\}}$ has the generating function $(1+x)$.

As $\mathcal{S}$ is the Cartesian product of r such sets, it has generating function $(1+x)^r$. Thus we have recovered the Binomial Theorem.

Example 3: Building structures on intervals This last example is not an example per se, but translates our counting sequence framework into the setting through which multiplication is defined in some other sources.

Suppose we have two different types of "structures" we can build on intervals.[29] Let $(a_n)_{n\geq 0}$ be the sequence denoting the number of structures of Type I that one can build on $[n]$, and let $(b_n)_{n\geq 0}$ denote the number of structures of Type II that can be built on $[n]$. Let $A(x)$ and $B(x)$ denote the corresponding generating functions.

If $C(x) = A(x)B(x) = \sum_{n\geq 0} c_n x^n$, then c_n denotes the number of ways of partitioning the interval $[n]$ into two disjoint (but possibly empty) subintervals $[k]$ and $[n] \setminus [k] \cong [n-k]$, and then building a structure of Type I on the first subinterval and a structure of Type II on the second subinterval.

To see why this is the case, let $\mathcal{A}$ be the set of all structures of Type I, where the size of a structure is the length of the interval it is built on. Let $\mathcal{B}$ be the set of all structures of Type II. The generating function $C(x)$ then corresponds to the counting function for the set $\mathcal{A} \times \mathcal{B}$, which consists of a structure of Type I followed by a structure of Type II. Furthermore, c_n counts the number of pairs of these structures with total length n.

Multiplication of generating functions thus has a very interesting and natural combinatorial interpretation, and is thus a powerful tool. Once you understand what the sequence corresponding to the product of two generating functions represents, it allows you to quickly build generating functions for more and more complicated sequences. We shall see some further examples of this in the following sections.

[29] As with the reason for the pizza party, the exact details of what the structures are is unimportant. The structures could be the choice of one element from the interval, any subset of the interval, an ordering of the elements in the interval, or anything of that sort.

The Catalan numbers

Recall from lectures that the Catalan numbers[30] $(c_n)_{n\geq 0}$ denote the number of Dyck paths of length $2n$ — diagonal lattice paths from $(0,0)$ to $(2n,0)$ that do not drop below the x-axis. We have $c_0 = c_1 = 1$, and derived the recurrence relation[31]

$$c_n = \sum_{k=0}^{n-1} c_k c_{n-1-k} \quad \text{for all } n \geq 1.$$

Using this recurrence relation, we can find a closed form for the generating function $C(x) = \sum_{n\geq 0} c_n x^n$. Indeed, we have

$$\begin{aligned} C(x) &= \sum_{n\geq 0} c_n x^n = c_0 + \sum_{n\geq 1} c_n x^n = 1 + \sum_{n\geq 1} \left(\sum_{k=0}^{n-1} c_k c_{n-1-k} \right) x^n \\ &= 1 + x \sum_{n\geq 1} \left(\sum_{k=0}^{n-1} c_k c_{n-1-k} \right) x^{n-1} = 1 + x \sum_{n\geq 0} \left(\sum_{k=0}^{n} c_k c_{n-k} \right) x^n \end{aligned}$$

However, by the convolution formula, we know

$$C(x)^2 = C(x)C(x) = \sum_{n\geq 0} \left(\sum_{k=0}^{n} c_k c_{n-k} \right) x^n.$$

Hence $C(x) = 1 + xC(x)^2$, or $xC(x)^2 - C(x) + 1 = 0$. To solve for $C(x)$, we use the quadratic formula.[32] This gives two possible solutions,

$$C_+(x) = \frac{1 + \sqrt{1-4x}}{2x} \quad \text{and} \quad C_-(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

To determine which one is correct, recall that $C(x) = \sum_{n\geq 0} c_n x^n$, and so $C(0) = \sum_{n\geq 0} c_n 0^n = c_0 = 1$. However, if we take the limit as $x \searrow 0$, we observe that $C_+(x) \to \infty$, as the numerator tends to 2 but the denominator tends to 0. Hence $C_+(x)$ cannot be the generating function we are looking for. Thus

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

[30] In the literature, the Catalan numbers are usually denoted C_n, but I am trying to stick to my convention of using lower-case letters for sequences and upper-case letters for generating functions.

[31] This recurrence relation comes from the fact that a Dyck path can be decomposed into two smaller Dyck paths (depending on the first return to the x-axis). Any structure which can be built out of similar smaller structures is likely to obey a similar recurrence relation, which explains the ubiquity of the Catalan numbers in combinatorics.

[32] "But wait," you ask, your concern about the rigour of this proof evident in your tone, "does the quadratic formula apply to power series?" If $C(x)$ describes an analytic function, then this is certainly sensible when x is in the radius of convergence, since $C(x)$ simply denotes some complex value (as does x itself). When we proceed with the calculation, we will arrive at an answer[33] that can be shown to be smaller than 4^n, which implies that the generating function is indeed analytic when $|x| < \frac{1}{4}$.

[33] We will arrive at an answer, but that must be the only answer, since the recurrence relation, together with the initial conditions, uniquely determines the entire sequence.

Having determined the generating function, we still have to find a formula for the Catalan numbers. Fortunately, though, this generating function is built out of simpler functions that are familiar to us, so we can use our sequence–function "dictionary" to solve for c_n.

We know the function $(1+x)^{\frac{1}{2}}$ corresponds to the sequence $(\binom{\frac{1}{2}}{n})_{n\geq 0}$. By definition, $\binom{\frac{1}{2}}{0} = 1$, and for $n \geq 1$,

$$\binom{\frac{1}{2}}{n} = \frac{1}{n!}\prod_{i=0}^{n-1}\left(\frac{1}{2}-i\right) = \frac{1}{2n!}\prod_{i=1}^{n-1}\left(\frac{1}{2}-i\right) = \frac{1}{2n!}\prod_{i=0}^{n-2}\left(\frac{-2i-1}{2}\right) = \frac{(-1)^{n-1}(2n-3)!!}{2^n n!},$$

where $(-1)!! = 1$.

Hence $(1-4x)^{\frac{1}{2}}$ is obtained by multiplying the nth term in the above sequence by $(-4)^n$, giving 1 when $n = 0$ and

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

The generating function $\frac{1-(1-4x)^{\frac{1}{2}}}{2}$ corresponds to subtracting this sequence from the sequence $(1,0,0,\ldots)$ and then dividing by 2. The $n=0$ term is then 0, while for $n \geq 1$ we get $\frac{2^{n-1}(2n-3)!!}{n!}$.

Finally, the Catalan generating function $C(x) = \frac{1-(1-4x)^{\frac{1}{2}}}{2x}$ corresponds to shifting this function to the left, and so for all $n \geq 0$, we have

$$c_n = \frac{2^n(2n-1)!!}{(n+1)!}.$$

While this is certainly a closed-form expression for c_n, it can be made a little nicer with a bit of arithmetic manipulation. Multiplying by $\frac{n!}{n!}$ and observing that $2^n n! = (2n)!!$, we have

$$c_n = \frac{2^n n!(2n-1)!!}{(n+1)!n!} = \frac{(2n)!!(2n-1)!!}{(n+1)!n!}.$$

Now $(2n)!!$ is the product of all even numbers between 2 and $2n$, while $(2n-1)!!$ is the product of all odd numbers between 1 and $2n-1$. Between them, we have the product of all integers between 1 and $2n$. Hence

$$c_n = \frac{(2n)!}{(n+1)n!n!} = \frac{1}{n+1}\binom{2n}{n},$$

giving a more compact expression for c_n.[34] Since $\binom{2n}{n} = \frac{n+1}{2n+1}\binom{2n+1}{n}$, this is equivalent to $c_n = \frac{1}{2n+1}\binom{2n+1}{n}$. In either form, this is a beautiful combinatorial formula, and in your homework you are asked to find a direct counting proof.

[34]Note that $\binom{2n}{n}$ is the number of subsets of $[2n]$ of size n, and hence is at most the total number of subsets of $[2n]$, which is $2^{2n} = 4^n$. Hence this sequence is indeed exponentially bounded, which confirms that its generating function gives an analytic function on the domain $|x| < \frac{1}{4}$ (see Footnote 32).

Generating functions for number partitions

We now turn to the partition function, adored by combinators and number theorists alike. Recall that $\vec{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k)$ is a partition of n if $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ and $|\vec{\lambda}| = \sum_{i=1}^{k} \lambda_i = n$, and $p(n)$[35] denotes the total number of partitions of n. While there is no known closed formula for $p(n)$, a lot is known about this function, and much of this knowledge comes via its generating function.

How, then, do we compute the generating function for the partition function? It turns out that it can naturally be written as a product of simpler generating functions, which implies that the set of partitions should be a Cartesian product of smaller sets. While we do write our partitions as vectors, this representation cannot arise from a Cartesian product. Indeed, the vectors are of variable lengths, depending on the number of parts, and the coordinates are not independent, since we require the parts to be ordered in non-increasing fashion.

Instead, we realise the Cartesian product representation by grouping the parts of the same size together. Given a partition $\vec{\lambda}$, let $\vec{\lambda}|_j$ be the subpartition consisting of parts of size j. For example, given $\vec{\lambda} = (7, 5, 5, 4, 3, 3, 3, 2, 1, 1) \vdash 34$, we have $\vec{\lambda}|_1 = (1, 1)$, $\vec{\lambda}|_2 = (2)$, $\vec{\lambda}|_3 = (3, 3, 3)$, $\vec{\lambda}|_4 = (4)$, $\vec{\lambda}|_5 = (5, 5)$, $\vec{\lambda}|_7 = (7)$, and $\vec{\lambda}|_j = \emptyset$ for $j = 6$ or $j \geq 8$.

Moreover, we can write $\vec{\lambda} = (\vec{\lambda}|_1, \vec{\lambda}|_2, \vec{\lambda}|_3, \ldots)$, and this representation can be realised as a Cartesian product. Indeed, there is one (possibly empty) component for each part size j, and the components are independent of each other. Hence, if $\mathcal{P}$ is the set of all number partitions, we have $\mathcal{P} = \mathcal{P}|_1 \times \mathcal{P}|_2 \times \mathcal{P}|_3 \times \ldots = \times_{j \in \mathbb{N}} \mathcal{P}|_j$, where $\mathcal{P}|_j$ is the set of number partitions only consisting of parts of size j.

Hence, in order to determine $P(x)$, the generating function for $\mathcal{P}$, it should suffice to find $P|_j(x)$, the generating function for $\mathcal{P}|_j$. Let $p|_j(n)$ denote the number of partitions of n with parts of size j. We trivially have

$$p|_j(n) = \begin{cases} 1 & \text{if } j|n \\ 0 & \text{otherwise} \end{cases}.$$

Indeed, if j does not divide n, then there is no way to write n as a sum of parts of size j, and so $p|_j(n) = 0$. On the other hand, if $j|n$, then there is no choice — we must write n as a sum of $\frac{n}{j}$ parts of size j. Thus there is a unique partition of n in $\mathcal{P}|_j$. Thus the sequence takes the form

$$(p|_j(n))_{n \geq 0} = (1, \underbrace{0, \ldots, 0, 1}_{j}, \underbrace{0, \ldots, 0, 1}_{j}, \ldots),$$

which is the all-1 sequence spaced out by a factor of j. The all-1 sequence has the generating function $(1 - x)^{-1}$, and so spacing it out by a factor of j results in $P|_j(x) = (1 - x^j)^{-1}$.

[35] If we were to follow our convention, the sequence would be denoted p_n. However, we had earlier used the standard notation $p(n)$, and shall continue to do so here. We will write $P(x)$ for the generating function.

Since the size of a partition $\vec{\lambda}$ is the sum of all its parts, it is also the sum of the sizes of its subpartitions $\vec{\lambda}|_j$, which is precisely how we want the sizes to behave under Cartesian products. Hence, by our results on the products of generating functions, we should have $P(x) = \sum_{n\geq 0} p(n)x^n = \prod_{j\in\mathbb{N}} P|_j(x)$.

However, this is an infinite product, and so we should take a little care to ensure it is well-defined. Each coefficient $p(n)$ of x^n should be determined by a finite computation. However, observe that if $j > n$, then the $P|_j(x)$ factor can only contribute a factor of 1 to the x^n term, since all its positive powers of x have larger degree. Hence the coefficient of x^n in $P(x)$ is the same as in $\prod_{j=1}^{n} P|_j(x)$,[36] which is indeed a finite and valid computation. Thus

$$P(x) = \sum_{n\geq 0} p(n)x^n = \prod_{j\in\mathbb{N}} P|_j(x) = \prod_{j\in\mathbb{N}} \frac{1}{1-x^j}.$$

For an alternative explanation of why this is the generating function for $p(n)$, we can use the geometric series to expand each factor. We have $(1-x^j)^{-1} = \sum_{a_j=0}^{\infty} x^{a_j\cdot j}$. Any term in the product $\prod_{j\in\mathbb{N}}(1-x^j)^{-1}$ thus corresponds to choosing some $a_j \in \{0, 1, 2, \ldots\}$ for each $j \in \mathbb{N}$.[37] The terms contributing to the coefficient for x^n are given by those choices for which $\sum_{j\in\mathbb{N}} a_j \cdot j = n$. In terms of the partitions themselves, this choice corresponds to the partition with a_j parts of size j. As there is a one-to-one correspondence between these choices $(a_j)_{j\geq 1}$ and partitions of n, the coefficient of x^n in the product is precisely the number of partitions of n.

Restricted number partitions One of the wonderful features of this generating function is that it allows one to study restricted classes of number partitions. The generating function can be easily modified to handle special classes of partitions, which can then lead to unexpected identities, or, through analytic means, asymptotic results.

For instance, if $S \subseteq \mathbb{N}$ is some subset of positive integers, then

$$P|_S(x) = \prod_{j\in S} P|_j(x) = \prod_{j\in S} \frac{1}{1-x^j}$$

is the generating function where the coefficient of x^n is the number of partitions of n into parts whose size belongs to S.

For example, suppose we wish to know how many ways there are of writing 10 as a sum of prime numbers. We may take S to be the set of primes up to 10; that is, $S = \{2, 3, 5, 7\}$. Our answer is then the coefficient of x^{10} in

$$\begin{aligned} P|_S(x) &= (1-x^2)^{-1}(1-x^3)^{-1}(1-x^5)^{-1}(1-x^7)^{-1} \\ &= (1+x^2+x^4+x^6+x^8+x^{10}+\ldots)(1+x^3+x^6+\ldots)(1+x^5+x^{10}+\ldots)(1+x^7+\ldots). \end{aligned}$$

It is then a simple task to multiply out and find that the desired coefficient is 5.[38]

[36] In other words, partitions of n can only have parts of sizes between 1 and n.

[37] Observe that the only choices that appear in the formal power series from the infinite product are those for which only finitely many a_j are positive.

[38] In fact, the multiplication even tells us what the partitions are: $x^{5\cdot 2}, x^{2\cdot 2+2\cdot 3}, x^{1\cdot 2+1\cdot 3+1\cdot 5}, x^{1\cdot 3+1\cdot 7}$ and $x^{2\cdot 5}$.

Rather than just placing restrictions on the sizes of the parts, restrictions can also be placed on the number of parts of a given size. This can be achieved by modifying the factor $P|_j(x)$. In its original form, this factor is $\sum_{a_j=0}^{\infty} x^{a_j \cdot j}$. However, by restricting the values of a_j allowed, we place conditions on the number of parts of size a_j.

For instance, suppose we want to count the number of partitions with an even number of parts of any given size. If $Q(x)$ is the generating function, we have

$$Q(x) = \prod_{j \in \mathbb{N}} \sum_{2 | a_j} x^{a_j \cdot j} = \prod_{j \in \mathbb{N}} \sum_{\ell=0}^{\infty} x^{(2\ell) \cdot j} = \prod_{j \in \mathbb{N}} \sum_{\ell=0}^{\infty} x^{\ell \cdot (2j)} = P|_S(x),$$

where $S = \{2, 4, 6, \ldots\} = 2\mathbb{N}$ is the set of even natural numbers. This proves that the number of partitions of n with an even number of parts of any given size is equal to the number of partitions of n into parts of even size.[39]

In closing, let us remember the existence of conjugate partitions, $\vec{\lambda}^*$, obtaining by exchanging rows and columns in Ferrers diagrams. As we had observed in lecture, conjugation provides a bijection between partitions of n with exactly k parts and partitions of n with largest part of size k. With these generating functions, it is not clear how to count partitions with exactly k parts,[41] but it is straightforward to count partitions with largest part k.[42] Conjugation can therefore be used together with these generating functions to obtain a wider class of identities.

Multivariate generating functions

In these notes, we have seen how generating functions can be used to count elements of some set $\mathcal{A}$, when sorted by some parameter we called the "size". However, in many instances, when we are enumerating some set there are several statistics on its elements that we would like to study. For example, in the previous section we studied number partitions, and were interested in the size of the partition. However, when we first encountered number partitions, we saw that it was also natural to consider the number of parts in the partition. This led to the refined number $p(n, k)$, which was the number of partitions of size n with k parts.

Generating functions can also be used to study several statistics simultaneously. To do so, we simply introduce a new variable for every statistic of interest. We saw that when we have a size function $|\cdot| : \mathcal{A} \to \mathbb{N} \cup \{0\}$, the generating function is given by $A(x) = \sum_{\alpha \in \mathcal{A}} x^{|\alpha|}$. Suppose now we also have some other statistic $\sigma : \mathcal{A} \to \mathbb{N} \cup \{0\}$. To keep track of this statistic as well, we define $A(x, y) = \sum_{\alpha \in \mathcal{A}} x^{|\alpha|} y^{\sigma(\alpha)}$. The coefficient of $x^n y^k$ in $A(x, y)$ then counts the number of elements in $\mathcal{A}$ with $|\alpha| = n$ and $\sigma(\alpha) = k$.

More generally, suppose we have some set $\mathcal{A}$ of elements we wish to enumerate, and some m statistics (including, perhaps, the size) $\sigma_i : \mathcal{A} \to \mathbb{N} \cup \{0\}$, $1 \leq i \leq m$. We can

[39] There is also a simple bijective proof of this result, which is left as an exercise for the reader.[40]

[40] Math-speak for "I do not feel like typing it out."

[41] At least, this is not clear yet, but the next section may help clarify things.

[42] Exercise: this generating function is $x^k P|_{[k]}(x)$.

define the multivariate generating function

$$A(x_1, x_2, \ldots, x_m) = \sum_{\alpha \in \mathcal{A}} \prod_{i=1}^{m} x_i^{\sigma_i(\alpha)}.$$

In this generating function, the coefficient of $\prod_{i=1}^{m} x_i^{n_i}$ counts the number of elements of $\mathcal{A}$ with $\sigma_i(\alpha) = n_i$ for all $1 \leq i \leq m$.

Example 1: The UEFA Champions League final As you are most probably aware,[43] on the 28th of May, Real and Atlético Madrid will face off against each other for the second time in three years to decide which football club is the best in Europe.[44] If you are a normal person, then you will spend the next few weeks excitedly and endlessly discussing the important footballing questions.[45] However, if you are a UEFA executive, you instead only have one question in mind: how much money can we make? For the purposes of this exercise, suppose you are a UEFA executive.

The final will be played at Milan's historic Stadio Giuseppe Meazza, more commonly known as the San Siro, which has a capacity of 81277. Your task is to fill this stadium, and make as much money as you can in the process. There are certain restrictions: you must reserve a number of seats for your corporate sponsors, you must ensure that each team has roughly equal support, and, of course, you cannot invite more people than the stadium can seat. What you would like to do is to enumerate all possible ways to distribute the tickets.

For each possible attendee, there are a certain number of statistics we must consider: how many seats will they occupy? Are they a corporate sponsor? Are they a Real fan? Are they an Atlético fan? How much will they pay for their ticket? This leads us to a generation function of the form $A(x_1, x_2, x_3, x_4, x_5)$, where x_1 counts the total number of people, x_2 the number of corporate sponsors, x_3 the number of Real fans, x_4 the number of Atlético fans, and x_5 the amount of money earned.

The set of stadium audiences is a Cartesian product over all possible attendees — for each person, they can either come to the stadium or not. If they do not come, they contribute nothing to the statistics, so we get a factor of 1. If they do come, they will add to the corresponding statistics, and so we get a monomial with the appropriate powers of each of the variables. For instance, a Real fan who would spend €1200 for a ticket gives the term $x_1 x_3 x_5^{1200}$ when attending, leading to a factor of $(1 + x_1 x_3 x_5^{1200})$ in the generating function. On the other hand, a corporate sponsor who gets in for free gives a factor of $(1 + x_1 x_2)$.

For a simplified example, suppose there are three types of potential attendees: 2000 corporate sponsors who get in for free, 100000 Real fans who would each pay €1200 for a ticket, and 78000 Atlético fans who would pay €900 for their tickets. The generating

[43] Unless you are not a big fan of football, are living under a rock, or are reading this in the future.

[44] Fans of Bayern Munich/Barcelona/Paris Saint-Germain/(other defeated club) will moan about how unlucky they were/lucky the Madrid teams were, but they can [censored].

[45] Who will prevail, Real's attack or Atlético's defence?[46] Will Torres score the winner to cap the unlikeliest of comeback seasons? Who will fake an injury most convincingly?

[46] In other words, what happens when an unstoppable force meets an immovable object?

functions for possible audiences is then given by

$$A(x_1,x_2,x_3,x_4,x_5) = \underbrace{(1+x_1x_2)^{2000}}_{\text{corporate}}\underbrace{(1+x_1x_3x_5^{1200})^{100000}}_{\text{Real fans}}\underbrace{(1+x_1x_4x_5^{900})^{78000}}_{\text{Atlético fans}}.$$

If we wanted to determine the largest amount of money we could make by filling the stadium with 1277 corporate sponsors and 40000 supporters of each club, we should determine the largest n such that the coefficient of $x_1^{81277}x_2^{1277}x_3^{40000}x_4^{40000}x_5^n$ is positive.

Example 2: Refined partitions Suppose we wish to build a refined generating function for number partitions that also keeps track of the total number of parts. We can introduce a new variable y to count the number of parts.

Hence the $P|_j(x) = \sum_{a_j=0}^{\infty} x^{a_j \cdot j}$ factor, which counted the parts of size j, would be refined to a $P|_j(x,y) = \sum_{a_j=0}^{\infty} x^{a_j \cdot j}y^{a_j} = \sum_{a_j=0}^{\infty}(x^j y)^{a_j}$ factor. Here the x term counts the total size of the parts of size j, while the y term counts how many there are.

By the geometric series formula, $P|_j(x,y) = \sum_{a_j=0}^{\infty}(x^j y)^{a_j} = (1-x^j y)^{-1}$. This leads to the overall formula $P(x,y) = \prod_{j\in\mathbb{N}}(1-x^j y)^{-1}$. The coefficient of $x^n y^k$ counts the number of partitions of size n with exactly k parts, and so we also have

$$P(x,y) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} p(n,k)x^n y^k = \prod_{j\in\mathbb{N}}(1-x^j y)^{-1}.$$

Manipulating multivariate generating functions By introducing new variables to keep track of several statistics, we of course gain more information about the objects we are enumerating. We can then manipulate this information by substituting different values for these variables.

For example, if we decide that we no longer wish to keep track of a certain statistic, we can set the corresponding variable to be 1. Since $1^n = 1$ for all $n \geq 0$, this means that the value of this statistic will no longer have any effect on the generating function, and is thus irrelevant. For example, if in the above example we make the number of parts irrelevant, we recover our original generating function for number partitions:

$$P(x,1) = \prod_{j\in\mathbb{N}}(1-x^j)^{-1} = P(x).$$

Another useful value to use is 0, as we have $0^0 = 1$ and $0^n = 0$ for all $n \geq 1$. This means that if we substitute 0 for one of the variables, we are only left with those objects where the corresponding statistic is equal to 0. For instance, in Example 1 above, $A(x_1,0,x_3,x_4,x_5) = (1+x_1x_3x_5^{1200})^{100000}(1+x_1x_4x_5^{900})^{78000}$ is the generating function for the stadium audiences without any corporate sponsors at all.

Finally, some results can also be obtained by setting variables equal to -1, as this separates even values of the statistic from odd values. For example, let q_n be the number of partitions of n with an even number of parts. We claim that its generating function is given by

$$Q(x) = \frac{1}{2}\left(P(x,1)+P(x,-1)\right) = \frac{1}{2}\left(\prod_{j\in\mathbb{N}}(1-x^j)^{-1} + \prod_{j\in\mathbb{N}}(1+x^j)^{-1}\right).$$

Indeed, we have

$$P(x,1) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} p(n,k)x^n 1^k = \sum_{n=0}^{\infty}\sum_{k \text{ even}} p(n,k)x^n + \sum_{n=0}^{\infty}\sum_{k \text{ odd}} p(n,k)x^n,$$

and

$$P(x,-1) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} p(n,k)x^n (-1)^k = \sum_{n=0}^{\infty}\sum_{k \text{ even}} p(n,k)x^n - \sum_{n=0}^{\infty}\sum_{k \text{ odd}} p(n,k)x^n.$$

Hence when we add the two together, the terms corresponding to partitions with an odd number of parts cancel out, while the partitions with an even number of parts are counted twice. The factor of $\frac{1}{2}$ corrects the overcount.

Note that, by conjugation, the number of partitions of n with an even number of parts is equal to the number of partitions of n whose largest part has even size. By the exercise in Footnote 42,[47] the generating function for partitions with largest part having size k is given by $x^k \prod_{j\in[k]}(1-x^j)^k$. Summing up over all even k gives us $Q(x)$, the generating function for partitions with an even number of parts. This results in the non-trivial[48] identity

$$\frac{1}{2}\left(\prod_{j\in\mathbb{N}}(1-x^j)^{-1} + \prod_{j\in\mathbb{N}}(1+x^j)^{-1}\right) = Q(x) = \sum_{k\in 2\mathbb{N}} x^k \prod_{j\in[k]}(1-x^j)^{-1}.$$

For those interested in further reading, the Rogers–Ramanujan[51] identities are more serious examples of identities that can be found through the use of generating functions.

Conclusion

This brings us to the end of this note, in which we surveyed some further applications of generating functions. That the product of generating functions corresponds to the combinatorially useful and interesting operation of convolution further illustrates the power of the generating function approach, which provides a uniform framework to handle problems in enumerative combinatorics. In the next few lectures, we will study some other aspects of generating functions, but we will still just be scratching the surface, and so the interested student is recommended to consult the suggested texts for further information.

[47] In retrospect, this should perhaps not have been a footnote, but what's done is done.

[48] Non-trivial in the sense that if we only consider the two power series, without thinking of what kind of number partitions they represent, it is not obvious[49] that they are equal. The complexity comes from the use of conjugation, which doesn't have a direct counterpart in the world of generating functions.

[49] I know this, because I asked my clever combinatorial friend Tuan, and he thought[50] about it for five minutes, but didn't find a direct proof.

[50] I even let him write down his thoughts.

[51] The observant reader will notice that these names are not ordered alphabetically. They are instead listed historically, as these identities were first discovered by Rogers, and later rediscovered by Ramanujan. They then published a joint paper where they gave new proofs of these identities, and are indeed listed alphabetically as authors of the paper.