

Multivariate Generating Functions and Other Tidbits

Zachary R. Abel

1 Introduction

This article is devoted to some of the methods and applications of generating functions as used in Olympiad problem solving. Their basic properties are well known to many problem solvers; this article is intended to explore some more advanced applications and ideas, the most prominent of which is the use of multivariate generating functions.

2 What is a Generating Function?

In the words of Herbert Wilf, “A generating function is a clothesline on which we hang up a sequence of numbers for display.”¹ This placeholder or ‘clothesline’ role is what makes them so valuable. Generating functions allow us to exploit the algebra of polynomials to model many otherwise unrelated situations. Our first example shows how generating functions can easily store important information and illustrates the importance of exploiting properties of polynomials.

Problem 1 (Zeitiz) *A standard die is labeled 1, 2, 3, 4, 5, 6 (one integer per face). When you roll two standard dice, it is easy to compute the probability of the various sums. For example, the probability of rolling two dice and getting a sum of 2 is just 1/36, while the probability of getting a 7 is 1/6.*

Is it possible to construct a pair of “nonstandard” [six-sided] dice (possibly different from one another) with positive integer labels that nevertheless are indistinguishable from a pair of standard dice, if the sum of the dice is all that matters? For example, one of these nonstandard dice may have the label 8 on one of its faces, and two 3’s. But the probability of rolling the two and getting a sum of 2 is still 1/36, and the probability of getting a sum of 7 is still 1/6.

Proof. [Solution] We can encode the contents of a standard die as a generating function by letting the coefficient of x^k be the number of faces that contain the number k . Thus, we define

$$\mathcal{D}(x) = x^1 + x^2 + x^3 + x^4 + x^5 + x^6.$$

¹This is the opening line in Herbert Wilf’s book *generatingfunctionology*, a marvelous and well-written textbook available for download at <http://www.math.upenn.edu/~wilf/DownldGF.html>.

3 A MUST-KNOW TOOL

Here is why this encoding is useful: consider the product

$$\mathcal{D}(x) \cdot \mathcal{D}(x) = (x^1 + x^2 + x^3 + x^4 + x^5 + x^6)(x^1 + x^2 + x^3 + x^4 + x^5 + x^6).$$

To get a clear idea of what the expanded polynomial looks like, consider for example the coefficient of x^6 . The possible ways to get an x^6 term in the product are $(x^1)(x^5)$, $(x^2)(x^4)$, $(x^3)(x^3)$, $(x^4)(x^2)$, and $(x^5)(x^1)$, so the coefficient of x^6 is 5. Looking more closely, each of these terms in the expansion represents one of the 5 ways to roll a sum of 6 with two dice: $1 + 5$, $2 + 4$, $3 + 3$, $4 + 2$, and $5 + 1$. In essence, the polynomial multiplication does all dice combinations for us and stores the results in the coefficients. So by this reasoning,

$$\mathcal{D}(x) \cdot \mathcal{D}(x) = x^2 + 2x^3 + 3x^4 + 4x^5 + \dots + 2x^{11} + x^{12},$$

where the coefficient of x^k represents the number of ways to roll a sum of k with two dice. We apply the same method to the nonstandard dice. Let a_1, \dots, a_6 and b_1, \dots, b_6 be the numbers on the two dice, and define the generating functions $\mathcal{A}(x) = x^{a_1} + \dots + x^{a_6}$ and $\mathcal{B}(x) = x^{b_1} + \dots + x^{b_6}$ to describe these dice. By the same reasoning as above, the product $\mathcal{A}(x) \cdot \mathcal{B}(x)$ represents the generating function for the possible outcomes of the sum of the numbers on the two dice. Since these are supposed to be the same outcomes as those of two standard dice, we have $\mathcal{A}(x) \cdot \mathcal{B}(x) = \mathcal{D}(x) \cdot \mathcal{D}(x)$. So we factor:

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

It now remains to distribute these factors between \mathcal{A} and \mathcal{B} . Let $\mathcal{A}(x) = x^{e_1}(x + 1)^{e_2}(x^2 - x + 1)^{e_3}(x^2 + x + 1)^{e_4}$ and $\mathcal{B}(x) = x^{f_1}(x + 1)^{f_2}(x^2 - x + 1)^{f_3}(x^2 + x + 1)^{f_4}$ where the e_i and f_i are non-negative integers with $e_i + f_i = 2$ for $1 \leq i \leq 4$. Since $\mathcal{A}(0) = \mathcal{B}(0) = 0$, each polynomial is divisible by x , i.e. $e_1 = f_1 = 1$. Also, it is clear from their original definitions that $\mathcal{A}(1) = \mathcal{B}(1) = 6$, which means $2^{e_2}3^{e_4} = 2^{f_2}3^{f_4} = 6$, i.e. $e_2 = f_2 = e_4 = f_4 = 1$. If we also let $e_3 = f_3 = 1$ then $\mathcal{A}(x) = \mathcal{B}(x) = \mathcal{D}(x)$, which means both dice are standard. The only other option is to let $e_3 = 2$ and $f_3 = 0$ (or the other way around):

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

Now we can easily read the contents of the dice directly from the generating functions: 1, 3, 4, 5, 6, 8 and 1, 2, 2, 3, 3, 4. Since this factorization was forced, it also follows that this is the only pair of nonstandard dice that works. ■

3 A Must-Know Tool

The use of polynomials allows access to all the information stored in its coefficients through the manipulation of a single object. There are many ways

3 A MUST-KNOW TOOL

to utilize this relationship between the polynomial and its coefficients, and the next problem gives another example of this.

Problem 2 (Romanian MO '03) *How many n -digit numbers, whose digits are in the set $\{2, 3, 7, 9\}$, are divisible by 3?*

Proof. [Solution] Since a number is divisible by 3 if and only if the sum of its digits is divisible by 3, we'll find how many of those numbers have a digital sum divisible by 3. So let's try to make a generating function $\mathcal{F}(x) = f_0 + f_1x + f_2x^2 + \dots + f_{9n}x^{9n}$ where f_k is the number of n -digit numbers with digits in $\{2, 3, 7, 9\}$ and a digital sum of k . Then the answer we want is $A = f_0 + f_3 + f_6 + \dots + f_{9n}$. Each digit will add either 2, 3, 7, or 9 to the total digit sum, and so the generating function for each digit is $x^2 + x^3 + x^7 + x^9$. Thus, the generating function we're looking for is simply

$$\mathcal{F}(x) = (x^2 + x^3 + x^7 + x^9)^n.$$

But how do we extract every third coefficient of the polynomial? The ingenious answer comes from complex numbers. Define $\varepsilon = e^{2\pi i/3}$ as one of the cube roots of unity, i.e. one of the solutions to $x^3 = 1$. This number has the simple property that $1 + \varepsilon + \varepsilon^2 = \frac{1 - \varepsilon^3}{1 - \varepsilon} = 0$. This property allows us to single out every third coefficient of \mathcal{F} quite easily. We have

$$\begin{aligned} \mathcal{F}(1) &= f_0 + f_1 + f_2 + f_3 + f_4 + \dots \\ \mathcal{F}(\varepsilon) &= f_0 + f_1\varepsilon + f_2\varepsilon^2 + f_3 + f_4\varepsilon + \dots \\ \mathcal{F}(\varepsilon^2) &= f_0 + f_1\varepsilon^2 + f_2\varepsilon + f_3 + f_4\varepsilon^2 + \dots, \end{aligned}$$

and adding these equations gives

$$\begin{aligned} \mathcal{F}(1) + \mathcal{F}(\varepsilon) + \mathcal{F}(\varepsilon^2) &= 3f_0 + (1 + \varepsilon + \varepsilon^2)f_1 + (1 + \varepsilon + \varepsilon^2)f_2 + 3f_3 + (1 + \varepsilon + \varepsilon^2)f_4 \dots \\ &= 3(f_0 + f_3 + f_6 + \dots) \\ &= 3A. \end{aligned}$$

With this we can easily calculate the answer to the problem:

$$\begin{aligned} A &= \frac{1}{3}(\mathcal{F}(1) + \mathcal{F}(\varepsilon) + \mathcal{F}(\varepsilon^2)) \\ &= \frac{1}{3}((1 + 1 + 1 + 1)^n + (\varepsilon^2 + 1 + \varepsilon + 1)^n + (\varepsilon + 1 + \varepsilon^2 + 1)^n) \\ &= \frac{1}{3}(4^n + 2). \end{aligned}$$

■

The method used here is certainly not specific to 3, and it leads to an extremely important tool:

Theorem 1 (Root of Unity Filter) *Define $\varepsilon = e^{2\pi i/n}$ for a positive integer n . For any polynomial $\mathcal{F}(x) = f_0 + f_1x + f_2x^2 + \dots$ (where f_k is taken to be zero when $k > \deg \mathcal{F}$), the sum $f_0 + f_n + f_{2n} + \dots$ is given by*

$$f_0 + f_n + f_{2n} + \dots = \frac{1}{n} (\mathcal{F}(1) + \mathcal{F}(\varepsilon) + \mathcal{F}(\varepsilon^2) + \dots + \mathcal{F}(\varepsilon^{n-1})).$$

4 MULTIPLE VARIABLES

Proof. The proof is based on a property of the sum $s_k = 1 + \varepsilon^k + \dots + \varepsilon^{(n-1)k}$. If $n \mid k$ then $\varepsilon^k = 1$ and so $s_k = 1 + 1 + \dots + 1 = n$. Otherwise, $\varepsilon^k \neq 1$ and $s_k = \frac{1 - \varepsilon^{nk}}{1 - \varepsilon^k} = 0$. So

$$\mathcal{F}(1) + \mathcal{F}(\varepsilon) + \dots + \mathcal{F}(\varepsilon^{n-1}) = f_0 s_0 + f_1 s_1 + f_2 s_2 + \dots = n \cdot (f_0 + f_n + f_{2n} + \dots),$$

and the proof is complete after division by n . ■

4 Multiple Variables

Now that we're familiar with (or reminded of) some of the basics, we can get to the main focus of this article: multivariable generating functions.

As we've mentioned before, the importance of generating functions is their spectacular ability to store large amounts of information. If you come across a problem that requires a lot of information storage and that poor little x can't do it all, give it a buddy! Add other variables to help organize the information. We have chosen two difficult examples to illustrate the enormous power of these multivariate generating functions.

Problem 3 (IMO '95/6) *Let p be an odd prime number. Find the number of subsets A of the set $\{1, 2, \dots, 2p\}$ such that*

- (i) *A has exactly p elements, and*
- (ii) *the sum of all elements in A is divisible by p .*

Proof. [Solution] Of course, we'll use a generating function to store information about the subsets. But what information might we want to keep track of? The problem talks about the *size* and the *sum* of the subsets, so we probably want to have information on both of these. To this end, we'll design a generating function

$$\mathcal{G}(x, y) = \sum_{n, k \geq 0} g_{n, k} x^n y^k$$

such that $g_{n, k}$ is the number of k -element subsets of $\{1, 2, \dots, 2p\}$ with a sum of n . The answer to the problem is therefore $A = g_{p, p} + g_{2p, p} + g_{3p, p} + \dots$. The construction of the generating function \mathcal{G} is very straightforward. We ask ourselves, how could the number m affect a subset? If it is not in the subset, then it affects neither the sum nor the size. But if it is in the subset, then it increases the sum by m and the size by 1. So for each m we should have the term $(1 + x^m y)$ in \mathcal{G} . Therefore,

$$\mathcal{G}(x, y) = (1 + xy)(1 + x^2y)(1 + x^3y) \cdots (1 + x^{2p}y)$$

is the generating function we're looking for. To get to our answer A , we need to extract coefficients from two types of terms of \mathcal{G} : y^p , and powers of x^p . Since

4 MULTIPLE VARIABLES

we already have a tool to do the latter (the root of unity filter), we choose to do this step first. Define $\varepsilon = e^{2\pi i/p}$. The filter tells us that

$$\sum_{\substack{n,k \geq 0 \\ p|n}} g_{n,k} y^k = \frac{1}{p} (\mathcal{G}(1, y) + \mathcal{G}(\varepsilon, y) + \cdots + \mathcal{G}(\varepsilon^{p-1}, y)), \quad (1)$$

so we need to calculate $\mathcal{G}(\varepsilon^k, y)$ for $0 \leq k \leq p-1$. When $k=0$, $\mathcal{G}(1, y) = (1+y)^{2p}$. For $1 \leq k \leq p-1$, since $\gcd(p, k) = 1$, the numbers $\{k, 2k, \dots, pk\}$ form a complete list of residues modulo p (verify!), so

$$\begin{aligned} \mathcal{G}(\varepsilon^k, y) &= (1 + \varepsilon^k y)(1 + \varepsilon^{2k} y) \cdots (1 + \varepsilon^{2pk} y) \\ &= ((1 + \varepsilon^k y)(1 + \varepsilon^{2k} y) \cdots (1 + \varepsilon^{pk} y))^2 \\ &= ((1 + \varepsilon y)(1 + \varepsilon^2 y) \cdots (1 + \varepsilon^p y))^2 \\ &= (1 + y^p)^2. \end{aligned}$$

Putting these values back into equation (1) produces

$$\sum_{\substack{n,k \geq 0 \\ p|n}} g_{n,k} y^k = \frac{1}{p} ((1+y)^{2p} + (p-1)(1+y^p)^2).$$

Now that we have successfully extracted the coefficients of the powers of x^p , our second task of finding the coefficient of y^p is easy: this coefficient (and the answer to our problem) is

$$\frac{1}{p} \left(\binom{2p}{p} + 2(p-1) \right).$$

■

Our discussion of this next problem is written to offer motivation and insight into the solution, and as a result it is rather lengthy. For this reason we have broken it into two parts in order to highlight the two main ideas used in the solution. Let's get started.

Problem 4 (USAMO '86/5) *By a partition π of an integer $n \geq 1$, we mean here a representation of n as a sum of one or more positive integers where the summands must be put in nondecreasing order. (E.g., if $n = 4$, then the partitions π are $1 + 1 + 1 + 1$, $1 + 1 + 2$, $1 + 3$, $2 + 2$, and 4).*

For any partition π , define $A(\pi)$ to be the number of 1's which appear in π , and define $B(\pi)$ to be the number of distinct integers which appear in π . (E.g., if $n = 13$ and π is the partition $1 + 1 + 2 + 2 + 2 + 5$, then $A(\pi) = 2$ and $B(\pi) = 3$).

Prove that, for any fixed n , the sum of $A(\pi)$ over all partitions π of n is equal to the sum of $B(\pi)$ over all partitions π of n .

4 MULTIPLE VARIABLES

Proof. [Solution] The idea is to create two different generating functions $\mathcal{A}(x) = \sum a_n x^n$ and $\mathcal{B}(x) = \sum b_n x^n$, where a_n is the sum of $A(\pi)$ over all partitions π of n and b_n is (you guessed it!) the sum of $B(\pi)$ over all partitions π of n . If we succeed in finding $\mathcal{A}(x)$ and $\mathcal{B}(x)$, we can hopefully say that $\mathcal{A}(x) = \mathcal{B}(x)$, which would show that $a_n = b_n$ for all n , which would solve the problem. (The wonderful tactic of wishful thinking is hard at work.) So, let's create \mathcal{A} . We need information on two things: the sum of the partitions and the number of 1's in the partitions. So we make a generating function that gives us both pieces of that information:

$$\begin{aligned} \mathcal{F}(x, y) &= \sum_{n, k \geq 0} f_{n, k} x^n y^k \\ &= (1 + xy + x^2 y^2 + \cdots)(1 + x^2 + x^4 + \cdots)(1 + x^3 + x^6 + \cdots) \cdots \end{aligned}$$

We've used the variable x to keep track of the sum of each partition and the variable y to denote the number of 1's in that partition: the coefficient $f_{n, k}$ is, by design, the number of partitions of n that have k ones. We may rewrite the geometric series to obtain the relatively simple

$$\mathcal{F}(x, y) = \frac{1}{(1 - xy)(1 - x^2)(1 - x^3) \cdots}.$$

(For those readers who are writhing in pain because of the informal treatment of infinite products and sums, a peek at section 5 should calm you down.) Okay, so how do we pull \mathcal{A} out of that generating function? The coefficient a_n is the total number of 1's in all partitions of n . This is equal to the number of partitions of n with a single 1, plus twice the number of partitions of n with two 1's, plus 3 times the number of partitions of n with three 1's, etc. Hence,

$$a_n = f_{n, 1} + 2f_{n, 2} + 3f_{n, 3} + \cdots = \sum_{k \geq 0} k \cdot f_{n, k}.$$

Multiplying this by x^n and summing over $n \geq 0$ gives

$$\mathcal{A}(x) = \sum_{n \geq 0} a_n x^n = \sum_{n, k \geq 0} k \cdot f_{n, k} x^n. \quad (2)$$

Our goal now is to mold \mathcal{F} into this sum. The key to transforming \mathcal{F} into the generating function above is differentiation. If we take the partial derivative of \mathcal{F} with respect to y , we get

$$\frac{\partial \mathcal{F}}{\partial y} = \sum_{n, k \geq 0} k \cdot f_{n, k} x^n y^{k-1}.$$

Then, setting $y = 1$ produces the sum in equation (2):

$$\left. \frac{\partial \mathcal{F}}{\partial y} \right|_{y=1} = \sum_{n, k \geq 0} k \cdot f_{n, k} x^n = \mathcal{A}(x).$$

4 MULTIPLE VARIABLES

Now all that's left to do is to calculate. We find that

$$\frac{\partial \mathcal{F}}{\partial y} = \frac{\partial}{\partial y} \left(\frac{1}{1-xy} \right) \cdot \frac{1}{(1-x^2)(1-x^3)\cdots} = \frac{x}{(1-xy)^2} \cdot \frac{1}{(1-x^2)(1-x^3)\cdots},$$

and by putting in $y = 1$ we get

$$\mathcal{A}(x) = \frac{x}{(1-x)^2(1-x^2)(1-x^3)\cdots}.$$

Good! We're half done. Now we'll try to create $\mathcal{B}(x)$. We will proceed in a similar manner. We first create a generating function to keep track of the sum of the partitions (with the variable x) and the number of distinct parts in the partitions (with y):

$$\begin{aligned} \mathcal{G}(x, y) &= \sum_{n, k \geq 0} g_{n, k} x^n y^k \\ &= (1 + xy + x^2y + \cdots)(1 + x^2y + x^4y + \cdots)(1 + x^3y + x^6y + \cdots) \cdots \\ &= \left(1 + \frac{xy}{1-x}\right) \left(1 + \frac{x^2y}{1-x^2}\right) \left(1 + \frac{x^3y}{1-x^3}\right) \cdots \end{aligned}$$

The coefficient $g_{n, k}$ is the number of partitions of n with k distinct parts. Luckily, \mathcal{B} and \mathcal{G} have a familiar relationship. The number of distinct parts in all partitions of n is the number of partitions of n with only one distinct part, plus twice the number of partitions of n with two distinct parts, etc., which shows that

$$b_n = g_{n, 1} + 2g_{n, 2} + 3g_{n, 3} + \cdots = \sum_{k \geq 0} k \cdot g_{n, k}.$$

So once again we have $\mathcal{B}(x) = \sum_{n \geq 0} b_n x^n = \sum_{k \geq 0} k \cdot g_{n, k} x^n$, and by the same reasoning as above, $\mathcal{B}(x) = \frac{\partial \mathcal{G}}{\partial y} \Big|_{y=1}$. All that we need to do now is differentiate \mathcal{G} with respect to y and plug in $y = 1$! Well, not quite. Unfortunately, it isn't so easy to differentiate an infinite product of functions. The closest we can get is differentiating finite products. For example, if we define a function $\mathcal{H}(x) = h_1(x)h_2(x)h_3(x)$, then by repeated use of the product rule we obtain

$$\begin{aligned} \mathcal{H}' &= h_1' h_2 h_3 + h_1 h_2' h_3 + h_1 h_2 h_3' \\ &= \mathcal{H} \cdot \left(\frac{h_1'}{h_1} + \frac{h_2'}{h_2} + \frac{h_3'}{h_3} \right). \end{aligned}$$

This easily generalizes to the following:

Theorem 2 (Generalized Product Rule) *The derivative of the function $\mathcal{H}(x) = h_1(x)h_2(x)\cdots h_n(x)$, where each $h_k(x)$ is a differentiable function of x , is given by*

$$\mathcal{H}'(x) = \mathcal{H}(x) \cdot \sum_{k=1}^n \frac{h_k'(x)}{h_k(x)}.$$

5 A MORE RIGOROUS TREATMENT

The proof is an easy induction on n , which we have omitted. For the moment, to make sure we're on the right track, let's pretend that the generalized product rule holds for an infinite product of functions. Define $h_k(y) = 1 + \frac{x^k y}{1-x^k}$, so that $\mathcal{G}(x, y) = h_1(y)h_2(y)\cdots$. We have $h_k(1) = 1 + \frac{x^k}{1-x^k} = \frac{1}{1-x^k}$ and

$$\frac{h'_k(1)}{h_k(1)} = \frac{\frac{x^k}{1-x^k}}{\frac{1}{1-x^k}} = x^k,$$

and so by the dubious extension of the product rule,

$$\begin{aligned} \mathcal{B}(x, y) &= \left. \frac{\partial \mathcal{G}}{\partial y} \right|_{y=1} = \left(\prod_{k=1}^{\infty} h_k(1) \right) \cdot \left(\sum_{k=1}^{\infty} \frac{h'_k(1)}{h_k(1)} \right) \\ &= \left(\prod_{k=1}^{\infty} \frac{1}{1-x^k} \right) \cdot \left(\sum_{k=1}^{\infty} x^k \right) \\ &= \frac{1}{(1-x)(1-x^2)(1-x^3)\cdots} \cdot \frac{x}{1-x} \\ &= \frac{x}{(1-x)^2(1-x^2)(1-x^3)\cdots}, \end{aligned}$$

which is exactly what we want! Now that we're confident of our method, we have to figure out a way to make it rigorous. We will come back to this issue in a later section. ■

5 A More Rigorous Treatment

It is natural to feel somewhat skeptical about the validity of some of the manipulations we have been doing. And that skepticism has good reason!

Consider the generating function for the Fibonacci numbers

$$\mathcal{F}(x) = F_0 + F_1x + F_2x^2 + \cdots = \frac{x}{1-x-x^2}.$$

(Can you prove this?) We can substitute $x = \frac{1}{3}$ into this identity to prove

$$\frac{1}{3} + \frac{1}{9} + \frac{2}{27} + \cdots + \frac{F_n}{3^n} + \cdots = \mathcal{F}\left(\frac{1}{3}\right) = \frac{\frac{1}{3}}{1 - \frac{1}{3} - \frac{1}{9}} = \frac{3}{5},$$

which was a problem from the Round Four Individual Mandelbrot contest in February 2002. But on the same note, we can substitute $x = 1$ into the same identity to 'prove' that the sum of all the Fibonacci numbers is

$$F_0 + F_1 + F_2 + \cdots = \mathcal{F}(1) = \frac{1}{1-1-1} = -1,$$

which is absurd!

5 A MORE RIGOROUS TREATMENT

What went wrong? What makes this different from the Mandelbrot question? It seems that we need a more rigorous treatment of generating functions and a clear idea of what is and isn't legal. So here goes.

We begin by discussing the *formal* theory of generating functions. This theory deals with a generating function as a purely algebraic object, whose sole purpose is one of storage. For a sequence $\{a_n\}_{n \geq 0}$, called the *sequence of coefficients*, we define its *Ordinary Power Series Generating Function* $\mathcal{A}(x)$ by

$$\mathcal{A}(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots .$$

Ordinary generating functions are algebraic objects that form a ring whose laws of addition and multiplication are defined naturally:

$$\mathcal{A}(x) + \mathcal{B}(x) = \sum_{n=0}^{\infty} (a_n + b_n)x^n \quad \text{and} \quad \mathcal{A}(x) \cdot \mathcal{B}(x) = \sum_{n=0}^{\infty} \left(\sum_{i=0}^n a_i b_{n-i} \right) x^n .$$

(This multiplication rule is what makes ordinary generating functions applicable to so many combinatorial situations.) Other operations can also be defined in an intuitive way. For example, differentiation is defined by

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

But keep in mind that a formal generating function is nothing more than an algebraic object, and x is nothing more than a placeholder. It is *not* a variable, and the generating function $\mathcal{A}(x)$ is *not* a function of x . Therefore functional operations such as evaluating $\mathcal{A}(1)$ make absolutely no sense in the realm of formal generating functions.

However, it is often desirable to use generating functions for more than their place-holding capabilities. In that case, we turn to *analytic* generating functions. Basically, if a generating function $\mathcal{F}(x)$ actually converges to an analytic function on some nontrivial domain, then we may treat $\mathcal{F}(x)$ as an analytic function of x when x is in that domain. The power series $\mathcal{F}(x)$ still has all the properties of formal generating functions, but it has the added bonus of being an analytic function, which makes it very useful.

For example, with the Mandelbrot question, it is easy to determine that the radius of convergence of the power series $\mathcal{F}(x) = \sum_n F_n x^n$ is $1/\phi$ (ϕ is the golden ratio), and thus when α is inside that bound, saying that

$$\frac{\alpha}{1 - \alpha - \alpha^2} = \mathcal{F}(\alpha) = F_0 + F_1\alpha + F_2\alpha^2 + \cdots$$

is a valid assertion. Since $1/3$ meets this condition, our calculation of $\mathcal{F}(1/3)$ makes sense. But $x = 1$ lies outside this domain, and so the expression $\mathcal{F}(1)$ is meaningless.

In the solution to problem 3, we substituted values for x without checking that it was legal. But the function $\mathcal{G}(x, y)$ in that problem is merely a polynomial in x and y , and polynomials are certainly analytic functions. So we're okay there.

With problem 4 we were already having difficulties when we tried to differentiate the infinite product of functions. Well, now we have even more troubles: we need to be sure that $\mathcal{F}(x, y)$ and $\mathcal{G}(x, y)$ are analytic functions of x and y before we can start using them as such. And with the functions we have, it's very difficult to verify this, if it is even true! So we need a way to get around the challenges that infinity poses.

6 Finitization

The easiest way to sidestep the difficulties of infinity is to eliminate infinity from the problem and instead make everything finite. We illustrate this idea of *finitization* in the following completion of our solution to problem 4.

Proof. [Solution cont.] Previously we defined a generating function $\mathcal{G}(x, y)$ and said that the generating function $\mathcal{B}(x)$ that we were looking for is equal to $\frac{\partial \mathcal{G}}{\partial y}$ with y replaced by 1. But we don't know how to take the derivative of the infinite function, and we don't know if we can substitute $y = 1$ legally. Finitization will help us to get around that. Fix a positive integer m , and define

$$\mathcal{G}_m(x, y) = (1 + xy + \cdots + x^m y)(1 + x^2 y + \cdots + (x^2)^m y) \cdots (1 + x^m y + \cdots + (x^m)^m y).$$

We have simply removed terms from \mathcal{G} to make \mathcal{G}_m , but we have not changed any of the terms containing x^m or lower. So for the same reasons as before, g_m must be the coefficient of x^m in $\frac{\partial \mathcal{G}_m}{\partial y} \Big|_{y=1}$. (We indicate this with the notation $g_m = [x^m] \frac{\partial \mathcal{G}_m}{\partial y} \Big|_{y=1}$.) The only difference is that $\mathcal{G}_m(x, y)$ is a *finite polynomial*, and so we can calculate this derivative and substitute $y = 1$ without any problems. Define $h_k(y) = 1 + x^k y + \cdots + x^{mk} y$, so that $\mathcal{G}_m = h_1(y) \cdots h_m(y)$. We calculate

$$\frac{h'_k(1)}{h_k(1)} = \frac{x^k + \cdots + x^{mk}}{1 + x^k + \cdots + x^{mk}} = \frac{\frac{x^k(1-x^{mk})}{1-x^k}}{\frac{1-x^{(m+1)k}}{1-x^k}} = \frac{x^k(1-x^{mk})}{1-x^{(m+1)k}},$$

and so the generalized product rule (theorem 2) indicates that

$$\frac{\partial \mathcal{G}_m}{\partial y} \Big|_{y=1} = \left(\prod_{k=1}^m h_k(1) \right) \cdot \left(\sum_{k=1}^m \frac{h'_k(1)}{h_k(1)} \right) = \left(\prod_{k=1}^m \sum_{i=0}^m x^{ki} \right) \cdot \left(\sum_{k=1}^m \frac{x^k(1-x^{mk})}{1-x^{(m+1)k}} \right).$$

Remember that b_m is the coefficient of x^m in the analytic generating function above. But now we no longer care that this generating function is analytic; the only thing we care about it is that it has b_m as the coefficient of x^m . So we may instead treat this polynomial as a formal generating function. This updated interpretation of the polynomial allows us to do two things. First, we may modify it as much as we want as long as the coefficient of x^m is not changed. And second, we don't have to feel guilty about using infinite expressions that may or may not converge, because we don't care if they converge. So, now we will mold it *without changing the x^m term or any lower terms* in order to get a better view of b_m . First, the only term in

$$\frac{x^k(1-x^{mk})}{1-x^{(m+1)k}} = x^k(1-x^{mk})(1+x^{(m+1)k}+x^{2(m+1)k}+\cdots)$$

6 FINITIZATION

that we care to preserve is x^k ; we may do away with the rest. The other alterations made to the polynomial in the next few steps also do not modify the terms we care about:

$$\begin{aligned} b_m &= [x^m] \left(\prod_{k=1}^m \sum_{i=0}^m x^{ki} \right) \cdot \left(\sum_{k=1}^m \frac{x^k(1-x^{mk})}{1-x^{(m+1)k}} \right) \\ &= [x^m] \left(\prod_{k=1}^m \sum_{i=0}^{\infty} x^{ki} \right) \cdot \left(\sum_{k=1}^m x^k \right) \\ &= [x^m] \left(\prod_{k=1}^{\infty} \frac{1}{1-x^k} \right) \cdot \left(\sum_{k=1}^{\infty} x^k \right) \\ &= [x^m] \frac{x}{(1-x)^2(1-x^2)(1-x^3)\dots} \end{aligned}$$

(Go through these steps carefully and convince yourself that none of the important terms has been changed.) Since our m was arbitrary, we have rigorously proved that $\mathcal{B}(x) = \frac{x}{(1-x)^2(1-x^2)(1-x^3)\dots}$ is the generating function for the sequence $\{b_n\}$. The construction of \mathcal{A} that was given previously still contains the flaw of treating \mathcal{F} as an analytic function without knowing if the infinite product converges. But a similar application of finitization will sidestep this with ease. We went to an extreme in the finitization of \mathcal{G} for the sake of instruction, but other options are available that also work with less painful algebra. Here's an example of an easier finitization used to make our construction of \mathcal{A} rigorous. Define

$$\mathcal{F}_m(x, y) = (1 + xy + x^2y^2 + \dots)(1 + x^2 + x^4 + \dots) \cdots (1 + x^m + x^{2m} + \dots).$$

When $|x| < \frac{1}{2}$ and $|y| < 2$, each of the geometric series in \mathcal{F}_m converges, and since \mathcal{F}_m is a *finite* product of these, \mathcal{F}_m converges in that region. Furthermore, the coefficient of x^m in \mathcal{F}_m is the same as that in \mathcal{F} . So we have

$$\begin{aligned} a_m &= [x^m] \left. \frac{\partial \mathcal{F}_m}{\partial y} \right|_{y=1} \\ &= [x^m] \left. \frac{x}{(1-xy)^2(1-x^2)\cdots(1-x^m)} \right|_{y=1} \\ &= [x^m] \frac{x}{(1-x)^2(1-x^2)\cdots(1-x^m)} \\ &= [x^m] \frac{x}{(1-x)^2(1-x^2)(1-x^3)\dots} \end{aligned}$$

So we have now rigorously verified the identity of $\mathcal{A}(x)$. (Notice we were careful to ensure that $y = 1$ lies in the domain we chose. And did you catch the subtle change from analytic to formal generating functions in those last steps?) We have rigorously shown that \mathcal{A} and \mathcal{B} are identical, from which it follows that $a_n = b_n$ for all n , so the problem is *finally* solved. ■