Chapter 19

Generating Functions and Convolutions

A common complaints in mathematics is the ubiquitous: *I can follow the proof when you do it line by line, but how could anyone ever think of doing this!* Of all the areas in probability, one of the most appropriate for sounding this complaint is in generating functions. At first glance, it seems like it's making our lives needlessly complex; however, at the end of the chapter you'll have learned how many different problems generating functions solve. Further, the time you spend learning these techniques will continue to pay dividends as you continue your studies, as these are used not just in probability, but throughout mathematical physics.

The reason for their helpfulness is that they allow us to package neatly a lot of information about a problem. You should be skeptical as to whether or not this is worthwhile; however, we'll see that time and time again that this new viewpoint simplifies the algebra we need to do. We'll give several motivating examples from previous courses of how a change in viewpoint can save you hours of labor, and then describe many of the properties and applications of generating functions. While there are many problems where it's quite difficult to find and use the correct generating function, a lot of useful problems can be handled with a small bag of tricks. Thus, be patient as you read on – the time you spend mastering this material will help you for years to come.

In probability, the most important use of generating functions is to understand moments of random variables. As we know, the moments tell us about the shape of the distribution. A very powerful application of this is in proving the Central Limit Theorem, which tells us that, in many cases, the sum of independent random variables tends towards a Gaussian as the number of summands grows. We'll devote Chapter 20 to this theorem (which shouldn't be surprising – anything given the name 'Central' should be expected to play a prominent role in a course).

19.1 Motivation

Frequently in mathematics we encounter complex data sets, and then do operations on it to make it even more complex! For example, imagine the first data set is the probabilities that the random variable X_1 takes on given values, and the second set

is the probabilities of another random variable X_2 taking on given values. From these we can, painfully through brute force, determine the probabilities of $X_1 + X_2$ equaling anything; however, if at all possible we would like to avoid these tedious computations. Below we'll study this problem in great detail in the special case that our two random variables have Poisson distributions (see §12.6 for properties of Poisson random variables). We'll solve the problem completely in this case, but the solution will be unsatisfying. The problem is we need to have some moments of divine inspiration in how to handle the algebra. The purpose of this example is to set the stage: we will introduce generating functions to automate the algebra.

Let's consider the case when X_1 has the Poisson distribution with parameter 5 and X_2 is a Poisson with parameter 7. This means

$$\begin{aligned} \text{Prob}(X_1 = m) &= 5^m e^{-5}/m! \\ \text{Prob}(X_2 = n) &= 7^n e^{-7}/n!, \end{aligned}$$

where m and n range over the non-negative integers. If k is a non-negative integer, then the probability that $X_1 + X_2 = k$ can be found by looking at all the different ways two non-negative integers can add to k. Clearly X_1 must take on a value between 0 and k; if it's ℓ then we must have X_2 equaling $k-\ell$. As our random variables are independent, the probability this happens is just the product of the probability that X_1 is ℓ and the probability that X_2 is $k - \ell$. If we now sum over ℓ we get the probability that $X_1 + X_2$ is k:

$$\begin{aligned} \text{Prob}(X_1 + X_2 = k) &= \sum_{\ell=0}^{k} \text{Prob}(X_1 = \ell)\text{Prob}(X_2 = k - \ell) \\ &= \sum_{\ell=0}^{k} \frac{5^\ell e^{-5}}{\ell!} \cdot \frac{7^{k-\ell} e^{-7}}{(k-\ell)!}. \end{aligned}$$

For general sums of random variables, it would be hard to write this in a more illuminating manner; however, we're lucky for sums of Poisson random variables *if we happen to think of the following sequence of simplifications!*

1. First, note that we have a factor of $1/\ell!(k-\ell)!$. This is almost $\binom{k}{\ell}$, which is $k!/\ell!(k-\ell)!$. We do one of the most useful tricks in mathematics, we **multiply cleverly by 1** (see §A.12 for more examples), where we write 1 as $k!/k!$. Thus this factor becomes $\binom{k}{\ell}/k!$. As our sum is over ℓ, we may pull the $1/k!$ outside the ℓ-sum.

2. The e^{-5} and e^{-7} inside the sum don't depend on ℓ, so we may pull them out, giving us an e^{-12}.

3. We now have $\frac{e^{-12}}{k!} \sum_{\ell=0}^{k} \binom{k}{\ell} 5^\ell 7^{k-\ell}$. Recalling the Binomial Theorem (Theorem A.2.2), we see the ℓ-sum is just $(5+7)^k$, or just 12^k.

Putting all the pieces together, we find

$$\text{Prob}(X_1 + X_2 = k) = \frac{12^k e^{-12}}{k!};$$

note this is the probability density for a Poisson random variable with parameter 12 (and $12 = 5 + 7$). There's nothing special about 5 and 7 in the argument above. Working more generally, we see the sum of two Poisson random variables with parameters λ_1 and λ_2 is a Poisson random variable with parameter $\lambda_1 + \lambda_2$.

This argument can be generalized. Using induction (or cleverly group parentheses), we find

Sums of Poisson random variables. The sum of n independent Poisson random variables with parameters $\lambda_1, \dots, \lambda_n$ is a Poisson random variable with parameter $\lambda_1 + \cdots + \lambda_n$.

We were fortunate in this case in that we found a 'natural' way to manipulate the algebra so that we could recognize the answer. What would happen if we considered other sums of random variables? We want a procedure that will work in general, which will *not* require us to see these clever algebra tricks.

Fortunately, there is such an approach. It's the theory of generating functions. We'll first describe what generating functions are (there are several variants; depending on what you are studying, some versions are more useful than others), and then show some applications.

19.2 Definition

We now define the generating function of a sequence. Though the most common applications are when the terms in the sequence are probabilities of different events or moments of distributions, a generating function can be defined for any sequence. In this section we'll define generating functions and give an example of their utility. Later on we'll apply what we learn to probability by either (1) taking the a_n's below to be the probability that a discrete random variable taking only non-negative integer values is n, or (2) taking the a_n's to be the moments of a random variable.

Definition 19.2.1 (Generating Function) *Given a sequence $\{a_n\}_{n=0}^{\infty}$, we define its generating function by*

$$G_a(s) \;=\; \sum_{n=0}^{\infty} a_n s^n$$

for all s where the sum converges.

The standard convention is to use the letter s for the variable; however, it's just a dummy variable and we could use any letter: s, x or even a ☺. Just looking at this definition, there's no reason to believe that we've made any progress in studying anything. We want to understand a sequence $\{a_n\}_{n=0}^{\infty}$ – how can it possibly help to make an infinite series out of these! The reason is that frequently there's a simple, closed form expression for $G_a(s)$, and from this simple expression we can derive many properties of the a_n's with ease!

Let's do an example. This example is long, but it's worth the time as it highlights many of the points of generating functions, and why they're so useful. Almost everyone has seen the **Fibonacci numbers**, defined by $F_0 = 0$, $F_1 = 1$ and in general $F_n = F_{n-1} + F_{n-2}$. The first few terms are 0, 1, 1, 2, 3, 5, 8, 13, ... These numbers have many wonderful properties. They occur throughout nature, from pine cones to branchings in trees (and of course to counting rabbits). They have applications in computer science, and generalizations arise in gambling theory (we'll discuss that application in Chapter 23). In principle, there are no mysteries about the Fibonacci numbers, as we have an explicit formula that allows us to compute any term in the sequence; in practice, this formula is clearly not useful for large n. While we can compute $F_{10} = 55$, it would be tedious to find $F_{100} = 354{,}224{,}848{,}179{,}261{,}915{,}075$, while computing F_{2011} with pen and paper is cause for alarm, as there are over 400 digits!

We now show how generating functions allow us to determine *any* Fibonacci number without having to compute *any* of the previous terms! The generating function is

$$G_F(s) \;=\; \sum_{n=0}^{\infty} F_n s^n.$$

We isolate the $n = 0$ and $n = 1$ terms, and for $n \geq 2$ we use the defining recurrence $F_n = F_{n-1} + F_{n-2}$ and find

$$\begin{aligned} G_F(s) &= F_0 + F_1 s + \sum_{n=2}^{\infty} (F_{n-1} + F_{n-2}) s^n \\ &= 0 + s + \sum_{n=2}^{\infty} F_{n-1} s^n + \sum_{n=2}^{\infty} F_{n-2} s^n. \end{aligned}$$

Notice the last two sums are almost our original generating function – they differ in having the wrong power of s, and the sums don't start at $n = 0$. We can fix this by pulling out some powers of s and then relabeling the summation; this is the hardest part of the argument, but after many examples it does eventually start to appear as a natural thing to do:

$$\begin{aligned} G_F(s) &= s + s \sum_{n=2}^{\infty} F_{n-1} s^{n-1} + s^2 \sum_{n=2}^{\infty} F_{n-2} s^{n-2} \\ &= s + s \sum_{m=1}^{\infty} F_m s^m + s^2 \sum_{m=0}^{\infty} F_m s^m. \end{aligned}$$

As $F_0 = 0$, we may extend the first sum to also be from $m = 0$. The two sums above are just $G_F(s)$, and thus we find

$$G_F(s) \;=\; s + s G_F(s) + s^2 G_F(s).$$

We now use the quadratic formula, and find

$$G_F(s) \;=\; \frac{s}{1 - s - s^2}. \tag{19.1}$$

Great – we've determined the generating function for the Fibonacci numbers: *How does this help us?* The reason we've made so much progress, though it doesn't

appear as if we have, is that the left hand side and right hand side of (19.1) are both functions of s. On the left hand side, the coefficient of s^n is just F_n; thus the coefficient of s^n on the right hand side must also be F_n. That said, it's not at all clear what the coefficient of s^n is on the right hand side. One natural idea is to try and expand using the geometric series:

$$\frac{1}{1-(s+s^2)} \;=\; \sum_{k=0}^{\infty}(s+s^2)^k \;=\; \sum_{k=0}^{\infty}\sum_{\ell=0}^{k}\binom{k}{\ell}s^{\ell}(s^2)^{k-\ell},$$

which gives

$$\frac{s}{1-s-s^2} \;=\; \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\binom{k}{\ell}s^{2k-\ell+1};$$

it's not easy to look at this and collect powers of s (but it's a nice exercise and leads to an interesting formula for the Fibonacci numbers)!

Fortunately there's a better way of looking at the right hand side. It goes back to one of the most disliked integration methods from calculus: **partial fractions**. Not surprisingly, there are good reasons your calculus professors taught this; in addition to being useful here, partial fractions also arise in solving certain differential equations. We factor $1-s-s^2$ as $(1-As)(1-Bs) = 1-(A+B)s+ABs^2$, and then write

$$\frac{s}{1-s-s^2} \;=\; \frac{a}{1-As}+\frac{b}{1-Bs},$$

and then use the geometric series to expand each fraction. It's because we want to use the geometric series formula that we write it as $(1-As)(1-Bs)$ and not $-(s-C)(s-D)$; for the geometric series formula we want the denominator to look like 1 minus something small.

A little algebra (or the quadratic formula) gives the values for A and B. We have $A+B=1$ and $AB=-1$. Thus $B=-1/A$ and $A-1/A=1$, or $A^2-A-1=0$. Therefore $A = \frac{1\pm\sqrt{5}}{2}$. We take the positive sign, and simple algebra then gives $B = \frac{1-\sqrt{5}}{2}$ (if we had taken the minus sign, the roles of A and B would just be reversed).

We now find a and b:

$$\frac{s}{1-s-s^2} \;=\; \frac{a}{1-As}+\frac{b}{1-Bs} \;=\; \frac{a+b-(aB+bA)s}{(1-As)(1-Bs)}.$$

Note the above is an equality, and it must hold for all values of s. As the denominators are the same, the only way this can happen is if the two numerators are equal. Each numerator is a polynomial in s; there's only one way these two polynomials can be equal for every choice of s – they must be the same polynomial, which means they must have the same coefficients.

Looking at the constant term, we find $a+b=0$, so $b=-a$. We now consider the coefficients of the s term. We now need $-(aB+bA)$ to equal 1. Using our values for A and B and the fact that $b=-a$ gives

$$-a\frac{1-\sqrt{5}}{2}+a\frac{1+\sqrt{5}}{2} \;=\; 1,$$

or $a = 1/\sqrt{5}$ and thus $b = -1/\sqrt{5}$. We've proved

$$G_F(s) \;=\; \frac{s}{1-s-s^2} \;=\; \frac{1}{\sqrt{5}}\frac{1}{1-As} - \frac{1}{\sqrt{5}}\frac{1}{1-Bs}.$$

We now expand with the geometric series, and see

$$\begin{aligned} G_F(s) &= \frac{1}{\sqrt{5}}\sum_{n=0}^{\infty} A^n s^n - \frac{1}{\sqrt{5}}\sum_{n=0}^{\infty} B^n s^n \\ &= \sum_{n=0}^{\infty}\left[\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n\right] s^n. \end{aligned}$$

We've found and proved the desired formula for the n^{th} Fibonacci number.

Binet's formula. Let $\{F_n\}_{n=0}^{\infty}$ denote the Fibonacci series, with $F_0 = 0$, $F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$. Then

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

Binet's formula is spectacular. We can now jump to *any* term in the sequence without calculating all the previous terms! I've always been amazed by it. The Fibonacci numbers are integers, yet this expression involves division and square-roots, yet somehow it all works out to be an integer.

After such a long argument, it's a good idea to go back and see what we've done. We started with a relation for the Fibonacci numbers. While we could use it to find any term, it would be time consuming. We bundled the Fibonacci numbers into a generating function $G_F(s)$. The miracle is that there's a nice closed form expression for $G_F(s)$, and from that we can deduce a nice formula for the Fibonacci numbers.

It's worth emphasizing the miracle that occurred, namely that $G_F(s)$ is nice. If we were to take a random sequence of numbers for the a_n's, this would not happen. Fortunately in many problems of interest, when the a_n's are related to probabilistic items we care about, there will be a nice form for the generating function.

The rest of this section may be safely skipped; however, as miracles are rare, it's worth trying to understand why one just happened. We're trying to answer why it's worth constructing a generating function. After all, if it's just equivalent to our original sequence of data, what have we gained? Were we just really lucky with the Fibonacci numbers, or do we expect this to happen again? Their most important advantage is that generating functions help simplify the algebra we'll encounter in probability calculations. We can't stress too strongly how useful it is in life to minimize the algebra you need to do. In addition to being a frequent source for errors, the more elaborate an expression is, the harder it is to see patterns and connections. Simplifying algebra is a great aid in illuminating connections, and often leads to enormous computational savings.

We give two examples to remind you how useful it can be to simplify algebra. The first is from calculus, and involves telescoping series.

Consider the following addition problem: evaluate

$$\begin{array}{rrrr} & 12 & - & 7 \\ + & 45 & - & 12 \\ + & 231 & - & 45 \\ + & 7981 & - & 231 \\ + & 9812 & - & 7981. \end{array}$$

The 'natural' way to do this is to do evaluate each line and then add; if we do this we get

$$5 + 33 + 186 + 7750 + 1831 \ = \ 9805$$

(or at least that's what we got on our calculators). A much faster way to do this is to regroup (see §A.3 for additional instances of **proofs by grouping**); we have a $+12$ and a -12, and so these terms cancel. Similarly we have a $+45$ and a -45, so these terms cancel. In the end we're left with

$$9812 - 7 \ = \ 9805,$$

a much simpler problem! One of the most important applications of telescoping series is in the proof of the fundamental theorem of calculus, where they're used to show the area under the curve $y = f(x)$ from $x = a$ to b is given by $F(b) - F(a)$, where F is any anti-derivative of f.

We turn to linear algebra for our second example; if you haven't seen eigenvalues and eigenvectors don't worry, as we won't use this later in the book but merely provide it as another illustration of the utility of simplifying algebra. Consider the matrix

$$A \ = \ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix};$$

what is A^{100}? If your probability (or linear algebra) grade depended on you getting this right, you would be in good shape. So long as you don't make any algebra errors, after a lot of brute force computations (namely 99 matrix multiplications!) you'll find

$$A^{100} \ = \ \begin{pmatrix} 218922995834555169026 & 354224848179261915075 \\ 354224848179261915075 & 573147844013817084101 \end{pmatrix}.$$

We can find this answer much faster if we diagonalize A. The eigenvalues of A are $\varphi = \frac{1+\sqrt{5}}{2}$ and $-1/\varphi$, with corresponding eigenvectors

$$\overrightarrow{v}_1 \ = \ \begin{pmatrix} -1+\varphi \\ 1 \end{pmatrix} \ \text{ and } \ \overrightarrow{v}_2 \ = \ \begin{pmatrix} -1-1/\varphi \\ 1 \end{pmatrix}$$

(remember $\overrightarrow{v}$ is an **eigenvector** of the matrix A with **eigenvalue** λ if $A\overrightarrow{v} = \lambda\overrightarrow{v}$; in other words, applying A to $\overrightarrow{v}$ doesn't change the direction – it just rescales its

length). Letting $S = (\vec{v}_1\ \vec{v}_2)$ and $\Lambda = \begin{pmatrix} \varphi & 0 \\ 0 & -1/\varphi \end{pmatrix}$, we see $A = S\Lambda S^{-1}$. The key observation is that $S^{-1}S = I$, the 2×2 identity matrix. Thus

$$A^2 \ = \ (S\Lambda S^{-1})(S\Lambda S^{-1}) \ = \ S\Lambda(S^{-1}S)\Lambda S^{-1} \ = \ S\Lambda^2 S^{-1};$$

more generally,

$$A^n \ = \ S\Lambda^n S^{-1}.$$

If we only care about finding A^2, this is significantly more work; however, there's a lot of savings if n is large. Note how similar this is to the telescoping example, with all the $S^{-1}S$ terms canceling.

As you might have guessed, this is not a randomly chosen matrix! This matrix arises in another approach to solving the Fibonacci relation $F_{n+1} = F_n + F_{n-1}$ (with $F_0 = 0$, $F_1 = 1$). If we let

$$\vec{v}_0 \ = \ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \vec{v}_n \ = \ \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix},$$

then $\vec{v}_n \ = \ A^n \vec{v}_0$. Thus, if we know A^n, we can quickly compute any Fibonacci number without having to determine its predecessors. This gives an alternative derivation of **Binet's formula**.

19.3 Uniqueness and Convergence of Generating Functions

Depending on the sequence $\{a_n\}_{n=0}^{\infty}$, it's possible for the generating function $G_a(s)$ to exist for all s, for only some s, or sadly only $s = 0$ (as $G_s(0) = a_0$, this isn't really saying much!).

Consider the following examples.

1. The simplest case is when $a_0 = 1$ and all other $a_n = 0$, which leads to $G_a(s) = 1$. More generally, if a_n is zero except for finitely many n then $G_a(s)$ is a polynomial.

2. If $a_n = 1$ for all n then $G_a(s) = \sum_{n=0}^{\infty} s^n = \frac{1}{1-s}$ by the geometric series formula. Of course, we need $|s| < 1$ in order to use the geometric series formula; for larger s, the series doesn't converge.

3. If $a_n = 1/n!$, then $G_a(s) = \sum_{n=0}^{\infty} s^n/n!$. This is the definition of e^s, and hence $G_a(s)$ exists for all s.

4. If $a_n = 2^n$, then $G_a(s) = \sum_{n=0}^{\infty} 2^n s^n = \sum_{n=0}^{\infty} (2s)^n$. This is a geometric series with ratio $2s$; the series converges for $|2s| < 1$ and diverges if $|2s| > 1$. Thus $G_a(s) = (1-2s)^{-1}$ if $|s| < 1/2$.

5. If $a_n = n!$, a little inspection shows $G_a(s)$ diverges for any $|s| > 0$. Probably the easiest way to see that this series diverges is to note that for any fixed $s \neq 0$, for all n sufficiently large we have $n!|s| > 1$; as the terms in the series don't

tend to zero, the series can't converge. Using Stirling's formula (see Chapter 18) we can get a good estimate on how large n must be for $n!|s| > 1$. Stirling's formula states that $n! \sim (n/e)^n\sqrt{2\pi n}$, so $n!|s|^n > (n|s|/e)^n$, which doesn't go to zero as whenever $n > e/|s|$ we have $|n!s^n| > 1$.

If we're given a sequence $\{a_n\}_{n=0}^{\infty}$, then clearly we know its generating function (it may not be *easy* to write down a closed form expression for $G_a(s)$, but we do have a formula for it). The converse is also true: if we know a generating function $G_a(s)$ (which converges for $|s| < \delta$ for some r), then we can recover the original sequence. This is easy if we can differentiate $G_a(s)$ arbitrarily many times, as then $a_n = \frac{1}{n!}\frac{d^n G_a(s)}{ds^n}$. This result is extremely important; as we'll use it frequently later, it's worth isolating as a theorem.

Theorem 19.3.1 (Uniqueness of generating functions of sequences) *Let $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ be two sequences of numbers with generating functions $G_a(s)$ and $G_b(s)$ which converge for $|s| < \delta$. Then the two sequences are equal (i.e., $a_i = b_i$ for all i) if and only if $G_a(s) = G_b(s)$ for all $|s| < \delta$. We may recover the sequence from the generating function by differentiating: $a_n = \frac{1}{n!}\frac{d^n G_a(s)}{ds^n}$.*

Proof: Clearly if $a_i = b_i$ then $G_a(s) = G_b(s)$. For the other direction, if we can differentiate arbitrarily many times, we find $a_i = \frac{1}{i!}\frac{d^i G_a(s)}{ds^i}$ and $b_i = \frac{1}{i!}\frac{d^i G_b(s)}{ds^i}$; as $G_a(s) = G_b(s)$, their derivatives are equal and thus $a_i = b_i$. □

Remark 19.3.2 *The division by $n!$ is a little annoying; later we'll see a related generating function that doesn't have this factor. If we don't want to differentiate, we can still determine the coefficients from the generating function. Clearly we can get a_0 by setting $s = 0$. We can then find a_1 by looking at $(G_a(s) - a_0)/s$ and sending s to zero in this expression; continuing in this manner we can find any a_m. Note, of course, how similar this is to differentiating!*

We end with a quick caveat to the reader: just because we've written down the generating function, it doesn't mean that it makes sense! Unfortunately it's possible that the resulting sum doesn't converge for any value of s (other than $s = 0$, of course, which trivially converges). Fortunately the generating functions that arise in probability frequently (but not always) converge, at least for some s; we'll discuss this in much greater detail later. There are many tests to determine whether or not a series converges or diverges, and we summarize four of the more popular and powerful (ratio, root, comparison and integral) in Appendix B.3.

In the next section we show how generating functions behave nicely with convolution, and from this we'll finally get some examples of why generating functions are so useful in probability.

19.4 Convolutions I: Discrete random variables

Above we introduced generating functions. We gave a few examples, we talked about how to see where it converges and diverges; however, we haven't seen why they're such a powerful tool in probability. We correct that now. After defining some notation, we'll return to the problem from the motivation section, namely determining the density of the sum of two random variables. The main result is that generating functions allow us to readily determine probability densities.

First, however, we need some notation.

Definition 19.4.1 (Convolution of sequences) *If we have two sequences $\{a_m\}_{m=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$, we define their convolution to be the new sequence $\{c_k\}_{k=0}^{\infty}$ given by*

$$c_k \;=\; a_0b_k + a_1b_{k-1} + \cdots + a_{k-1}b_1 + a_kb_0 \;=\; \sum_{\ell=0}^{k} a_\ell b_{k-\ell}.$$

We frequently write this as $c = a * b$.

This definition arises from multiplying polynomials; if $f(x) = \sum_{m=0}^{\infty} a_m x^m$ and $g(x) = \sum_{n=0}^{\infty} b_n x^n$, then assuming everything converges we have

$$h(x) \;=\; f(x)g(x) \;=\; \sum_{k=0}^{\infty} c_k x^k,$$

with $c = a * b$. For example, if $f(x) = 2 + 3x - 4x^2$ and $g(x) = 5 - x + x^3$, then $f(x)g(x) = 10 + 13x - 23x^2 + 6x^3 + 3x^4 - 4x^5$. According to our definition, c_2 should equal

$$a_0b_2 + a_1b_1 + a_2b_0 \;=\; 2\cdot 0 + 3\cdot(-1) + (-4)\cdot 5 \;=\; -23,$$

which is exactly what we get from multiplying $f(x)$ and $g(x)$.

Lemma 19.4.2 *Let $G_a(s)$ be the generating function for $\{a_m\}_{m=0}^{\infty}$ and $G_b(s)$ the generating function for $\{b_n\}_{n=0}^{\infty}$. Then the generating function of $c = a * b$ is $G_c(s) = G_a(s)G_b(s)$.*

We can now give a nice application of how generating functions can simplify algebra: *What is $\sum_{m=0}^{n} \binom{n}{m}^2$?* If we evaluate this sum for small values of n we find that when $n = 1$ the sum is 1, when $n = 2$ it's 6, when $n = 3$ it is 20, then 70 and then 252. We might realize that the answer seems to be $\binom{2n}{n}$, but even if we notice this, how would we prove it? A natural idea is to try induction. We could write $\binom{n}{m}^2$ as $\left(\binom{n-1}{m-1} + \binom{n-1}{m}\right)^2$ (noting that we have to be careful when $m = 0$). If we expand the square we get two sums similar to the initial sum but with an $n-1$

instead of an n, which we would know by induction; the difficulty is that we have the cross term $\binom{n-1}{m-1}\binom{n-1}{m}$ to evaluate, which requires some effort to get this to look like something nice times something like $\binom{n-1}{\ell}^2$.

Using generating functions, the answer just pops out. Let $a = \{a_m\}_{m=0}^n$, where $a_m = \binom{n}{m}$. Thus

$$G_a(s) \;=\; \sum_{m=0}^{n} \binom{n}{m} s^m \;=\; \sum_{m=0}^{n} \binom{n}{m} s^m 1^{n-m} \;=\; (1+s)^n$$

(when we have binomial sums such as this, it's *very* useful to introduce factors such as 1^{n-m}, which facilitates using the Binomial Theorem, Theorem A.2.2).

Let $c = a * a$, so by Lemma 19.4.2 we have $G_c(s) = G_a(s)G_a(s) = G_a(s)^2$. At first this doesn't seem too useful, until we note that

$$c_n \;=\; \sum_{\ell=0}^{n} a_\ell a_{n-\ell} \;=\; \sum_{\ell=0}^{n} \binom{n}{\ell}\binom{n}{n-\ell} \;=\; \sum_{\ell=0}^{n} \binom{n}{\ell}^2$$

as $\binom{n}{n-\ell} = \binom{n}{\ell}$. Thus the answer to our problem is c_n. We don't know c_n, but we *do* know its generating function, *and the entire point of this exercise is to show that sometimes it's more useful to know one and deduce the other.* We have

$$\sum_{k=0}^{2n} c_k s^k \;=\; G_c(s) \;=\; G_a(s)^2 \;=\; (1+s)^n \cdot (1+s)^n \;=\; (1+s)^{2n} \;=\; \sum_{k=0}^{2n} \binom{2n}{k} s^k,$$

where the last equality is just the Binomial Theorem. Thus $c_n = \binom{2n}{n}$ as claimed.

While we've found an example where it's easier to study the problem through generating functions, some things are unsatisfying about this example. The first is we still needed to have some combinatorial expertise, noting $\binom{n}{\ell} = \binom{n}{n-\ell}$. This is minor for two reasons. First, this is one of the most important properties of binomial coefficients (the number of ways of choosing ℓ people from n people when order doesn't matter is the same as the number of ways of excluding $n-\ell$). The second is more severe: *why would one ever consider convolving our sequence a with itself to solve this problem!*

The answer to the second objection is that convolutions arise all the time in probability, and thus it's natural to study any process which is nice with respect to convolution. To see this, we define

Definition 19.4.3 (Probability generating function) *Let X be a discrete random variable taking on values in the integers. Let $G_X(s)$ be the generating function to $\{a_m\}_{m=-\infty}^{\infty}$ with $a_m = \text{Prob}(X = m)$. Then $G_X(s)$ is called the probability generating function. If X is only non-zero at the integers, a very useful way of computing $G_X(s)$ is to note that*

$$G_X(s) = \mathbb{E}[s^X] = \sum_{m=-\infty}^{\infty} s^m \text{Prob}(X = m).$$

More generally, if the probabilities are non-zero on an at most countable set $\{x_m\}$, then

$$G_X(s) = E[s^X] = \sum_m s^{x_m} \text{Prob}(X = x_m).$$

The function $G_X(s)$ can be a bit more complicated than the other generating functions we've seen if X takes on negative values; if this is the case, we're no longer guaranteed that $G_X(0)$ makes sense! One way we can get around this problem is by restricting to s with $0 < \alpha < |s| < \beta$ for some α, β; another is to restrict ourselves to random variables that are never negative, and thus this issue can't arise! We concentrate on the latter. While this does restrict the distributions we may study a bit, so many of the common, important probability distributions (Bernoulli, geometric, Poisson, negative binomial, ...) of Chapter 12 take on non-negative integer values that we have a wealth of examples and applications.

We can now state one of the most important results for probability generating functions.

Theorem 19.4.4 *Let $X_1, \dots, X_n$ be* independent *discrete random variables taking on non-negative integer values, with corresponding probability generating functions $G_{X_1}(s), \dots, G_{X_n}(s)$. Then*

$$G_{X_1+\cdots+X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s).$$

Proof: This is one of the cornerstone results in the subject; you should keep reading the proof until it completely sinks in. We'll do the case when $n = 2$ in full detail, and leave arbitrary n for you.

Basically, all we need to do is unwind the definitions. We have

$$\text{Prob}(X_1 + X_2 = k) = \sum_{\ell=0}^{\infty} \text{Prob}(X_1 = \ell)\text{Prob}(X_2 = k - \ell).$$

If we let $a_m = \text{Prob}(X_1 = m)$, $b_n = \text{Prob}(X_2 = n)$ and $c_k = \text{Prob}(X_1 + X_2 = k)$, we see that $c = a * b$. Thus $G_c(s) = G_a(s)G_b(s)$, or equivalently, $G_{X_1+X_2}(s) = G_{X_1}(s)G_{X_2}(s)$.

What if now $n = 3$? It's another **proof by grouping** (see §A.3): write $X_1 + X_2 + X_3$ as $(X_1 + X_2) + X_3$. Using the $n = 2$ result *twice* we get

$$\begin{aligned} G_{X_1+X_2+X_3}(s) &= G_{(X_1+X_2)+X_3}(s) \\ &= G_{X_1+X_2}(s)G_{X_3}(s) = G_{X_1}(s)G_{X_2}(s)G_{X_3}(s). \end{aligned}$$

A similar idea works for all n. □

Whenever you see a theorem, you should remove a hypothesis and ask if it's still true. Usually the answer is a resounding *NO!* (or, if true, the proof is usually significantly harder). In the theorem above, how important is it for the random variables to be independent? As an extreme example consider what would happen if $X_2 = -X_1$. Then $X_1 + X_2$ is identically zero, but $G_{X_1+X_2}(s) \neq G_{X_1}(s)G_{-X_1}(s)$.

The above shows why generating functions play such a central role in probability.

The density of the sum of independent discrete random variables is the convolution of their probabilities!

We can begin to see why generating functions are so useful. From Theorem 19.3.1 we know the generating function is unique, and from Theorem 19.4.4 we know that the generating function of the sum of random variables is the product of the generating functions. If we happen to recognize the resulting product, we can immediately glean the density function of the sum!

Let's return to the problem from the motivation section, §19.1. We have two independent Poisson random variables, X_1 with parameter 5 and X_2 with parameter 7, and we want to understand $X_1 + X_2$. From Definition 19.5.1, the generating function of a Poisson random variable X with parameter λ is just

$$\begin{aligned} G_X(s) &= \sum_{n=0}^{\infty} \text{Prob}(X = n)s^n \\ &= \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} s^n \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda s)^n}{n!} \\ &= e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}, \end{aligned}$$

where we used the exponential function's series expansion: $e^u = \sum_{n=0}^{\infty} u^n/n!$. Thus

$$G_{X_1} = e^{5(s-1)}, \quad G_{X_2} = e^{7(s-1)}.$$

From Theorem 19.4.4 we have

$$\begin{aligned} G_{X_1+X_2}(s) &= G_{X_1}(s)G_{X_2}(s) \\ &= e^{5(s-1)} \cdot e^{7(s-1)} \\ &= e^{12(s-1)}; \end{aligned}$$

however, note that $e^{12(s-1)}$ is just the generating function of a Poisson random variable with parameter 12. As Theorem 19.3.1 tells us generating functions are unique,

we can now deduce that $X_1 + X_2$ is a Poisson random variable with parameter 12.

In the above example, note how much easier it was to understand $X_1 + X_2$ by using properties of generating functions than from doing the algebra directly. We tackled the algebra in §19.1; while we solved the problem, we had to make several clever choices in the analysis. The arguments are far more straightforward when we use generating functions. We'll do more examples of this later, and even study cousins of generating functions that makes the algebra even easier, namely the moment generating functions and the characteristic functions.

19.5 Convolutions II: Continuous random variables

Fortunately the same arguments that analyzed the discrete case can be easily adapted to handle continuous random variables. Essentially the only difference is writing integrals rather than sums. (There's a few subtle, technical difficulties with integration, which we'll briefly mention.) While a general random variable need not be purely discrete or continuous, for most problems our random variables are one or the other. Frequently books adopt the convention that a sum could also mean an integral, or an integral could mean a sum. This allows them greater flexibility in writing as one notation can refer to either case.

Let's now adjust our notation and study the case of generating functions for continuous random variables.

Definition 19.5.1 (Probability generating function) *Let X be a continuous random variable with density f. Then*

$$G_X(s) \;=\; \int_{-\infty}^{\infty} s^x f(x)dx$$

is the probability generating function of X.

Let's compute some generating functions of continuous random variables. If we let X be an exponential with parameter λ, we have its density is

$$f(x) \;=\; \begin{cases} \frac{1}{\lambda}\exp(-x/\lambda) & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(Note that there's unfortunately a difference in opinion among authors as to what the exponential density should be; some books use this notation while others would use $\lambda\exp(-\lambda x)$; I prefer the first choice as this way an exponential random variable with parameter λ has mean λ and not mean $1/\lambda$.) The generating function is thus

$$G_X(s) \;=\; \int_0^{\infty} s^x \frac{1}{\lambda}\exp(-x/\lambda)dx \;=\; \frac{1}{\lambda}\int_0^{\infty} \exp(x\log s)\exp(-x/\lambda)dx.$$

Notice we rewrote s^x as $\exp(x\log s)$. While we can see these two expressions are the same by taking logarithms, why did we do this? Remember s is fixed and x is the integration variable. If instead of s^x we had e^x then we could just combine the