100.10 Dice problems with generating function solutions

Introduction

Probability theory prescribes rules for computing likelihoods of complicated events. Unfortunately, these rules often result in exact formulas that are complicated beyond usefulness, like missing the forest for the trees. One handy tool to deal with complicated probabilities is the so-called *generating function*. Generating functions provide a bridge between the discrete world of combinatorics and the continuous world of calculus. The main idea is to convert a sequence of probabilities into a function (the generating function) and then, using techniques from algebra or calculus, to analyse the function. It turns out that the derivatives, singularities, etc., of the generating function often give telling information about the probabilities. The beauty of the method is that frequently only basic algebra and calculus skills are needed to reveal fundamental characteristics, including accurate estimates, of the underlying probabilities.

As an example of this method, we consider questions related to rolling dice in the game of Monopoly. In this game players roll a pair of dice to move a token around a game board. The most expensive property on the board is Mayfair (in the UK, or Boardwalk in the USA, or countless others in other countries), located thirty-nine sites from the start. Moving a token around the board with random rolls of a pair of dice, what is the probability of landing on Mayfair during the first circuit? In other words, what is the probability of exactly obtaining a sum of thirty-nine with rolls of dice? A closely related question deals with obtaining or exceeding a particular sum. In Monopoly, players collect a salary of £200 each time they pass Go, which is located forty spaces from the start. On average, how many rolls does it take to pass Go? In other words, what is the expected number of rolls of dice to obtain or exceed a sum of forty? Exact answers to these questions are complicated. However, using generating functions, we shall see that it is easy to obtain accurate estimates of these probabilities and expectations using only basic methods from elementary algebra and calculus.

Preliminaries

Generating functions for dice games are usually rational functions (ratios of polynomials). To deal with rational functions, only a handful of skills are necessary. The first is polynomial arithmetic. The addition of polynomials goes term by term so, for instance,

$$\left(4 + z^2\right) + \left(8 - 8z + 2z^2 - 2z^3\right) + \left(2z - 4z^2 + 2z^3\right) = 12 - 6z - z^2.$$

The multiplication of polynomials coincides with the so-called discrete convolution of sequences: the coefficient of each term in the product is the sum of the products of coefficients whose powers add up to the power of the given term. For instance,

$$(1 - z)^2 \times \left(4 + z^2\right) = 4 - 8z + 5z^2 - 2z^3 + z^4.$$

https://doi.org/10.1017/mag.2016.19 Published online by Cambridge University Press

Quotients of polynomials can be decomposed with partial fractions, one term for each prime power factor of the denominator. For instance,

$$\frac{12 - 6z - z^2}{4 - 8z + 5z^2 - 2z^3 + z^4} = \frac{1}{(1-z)^2} + \frac{2}{1-z} + \frac{2z}{4+z^2}.$$

Another tool for dealing with rational functions is the geometric series. A geometric series with an initial term of 1 and ratio z with $|z| < 1$ has a sum of $1 + z + z^2 + z^3 + \ldots = \dfrac{1}{1-z}$. Likewise, a geometric series with an initial term of 1 and a ratio of $\frac{1}{2}z$ with $|z| < 2$ has a sum of

$$1 + \frac{z}{2} + \frac{z^2}{4} + \frac{z^3}{8} + \ldots = \frac{1}{1-\frac{z}{2}}.$$

More generally, $\dfrac{1}{1-\frac{z}{r}} = 1 + \dfrac{z}{r} + \dfrac{z^2}{r^2} + \dfrac{z^3}{r^3} + \ldots$. One easy-but-useful observation about this geometric series is that if $|r| > 1$, then the coefficients of the terms in this series go to zero exponentially fast. For example, the rational function

$$\frac{4-z}{2-3z+z^2} = \frac{3}{1-z} - \frac{1}{1-\frac{1}{2}z}$$

has the series expansion

$$\sum_{n \geqslant 0} \left(3 - 2^{-n}\right) z^n,$$

but the coefficients of this series approach 3 since 2^{-n} goes to 0 exponentially fast. The main observation that we use in first-order approximations here is the following. If a root of the denominator of a term of the decomposition of a rational function in partial fractions is greater than 1 in absolute value, then we can ignore this term completely since, like 2^{-n} in the previous example, the coefficients go to 0 (see, for instance, Theorem IV.9 of [1] for an exact expansion of rational functions).

A series related to the geometric series is $\dfrac{1}{(1-z)^2}$. This series can be thought of as either the square of the geometric series (with the product following the discrete convolution) or as the derivative of the geometric series. In any case, the coefficient of the z^n term is $n + 1$, so that

$$\frac{1}{(1-z)^2} = \sum_{n \geqslant 0} (n+1) z^n, \qquad \text{for } |z| < 1.$$

Landing on Mayfair

Getting an exact expression for the probability of rolling dice and obtaining a particular sum is a thorny problem. Exact expressions for this probability exist, but they are so complicated that they cloud the big picture.

https://doi.org/10.1017/mag.2016.19 Published online by Cambridge University Press

However, with the aid of generating functions, it is easy to get accurate approximations that clearly show the general nature of the solutions. We consider two cases—rolling a single fair die or rolling a pair of fair dice. In each case we compute the probability of obtaining a particular sum. For instance, Mayfair is located on the thirty-ninth site from the start of the board in Monopoly.

Let us start by computing probabilities of several events related to repeated rolls of a single fair die. The probability of rolling value n is $\frac{1}{6}$ if $1 \leqslant n \leqslant 6$ and 0 otherwise. The generating function of this sequence is the polynomial

$$\mathrm{Roll}(z) = \tfrac{1}{6}z + \tfrac{1}{6}z^2 + \tfrac{1}{6}z^3 + \tfrac{1}{6}z^4 + \tfrac{1}{6}z^5 + \tfrac{1}{6}z^6.$$

The sum of the probabilities of all the individual values is one, so $\mathrm{Roll}(1) = 1$. Then, the quotient $\dfrac{1 - \mathrm{Roll}(z)}{1 - z}$ is a quintic polynomial

$$\frac{6}{6} + \frac{5}{6}z + \frac{4}{6}z^2 + \frac{3}{6}z^3 + \frac{2}{6}z^4 + \frac{1}{6}z^5,$$

which we denote by $Q(z)$. Now the probability of getting a sum of n with two rolls of a single die is a sum over the probabilities of obtaining a particular value m on the first roll times the probability of obtaining the difference $n - m$ on the second. This sum exactly corresponds to polynomial multiplication. So, the generating function of the sequence of probabilities of getting particular sums with two rolls is $(\mathrm{Roll}(z))^2$. Likewise, the generating function of the sequence of probabilities of getting particular sums with r rolls is $(\mathrm{Roll}(z))^r$. Finally, the probability of obtaining a particular value, regardless of the number of rolls, is just the sum of the probabilities of obtaining that value over all numbers of rolls. So, the generating function for the sequence of probabilities of obtaining a particular sum of rolls with a single die is

$$\big(\text{Sum of Rolls}\big)(z) = \sum_{r \geqslant 0} (\mathrm{Roll}(z))^r.$$

This is a geometric series of polynomials, which is

$$\big(\text{Sum of Rolls}\big)(z) = \frac{1}{1 - \mathrm{Roll}(z)}.$$

In order to find the probability of landing on Mayfair, all that is left to do is to extract the coefficient of z^{39} in this generating function. It has a partial fractions decomposition of the form

$$\begin{aligned}\big(\text{Sum of Rolls}\big)(z) &= \frac{1}{(1 - z)Q(z)} \\ &= \frac{\frac{2}{7}}{1 - z} + \frac{?}{Q(z)},\end{aligned}$$

where ? denotes a yet-to-be determined quartic polynomial. Granted, the

https://doi.org/10.1017/mag.2016.19 Published online by Cambridge University Press

roots of the quintic $Q(z)$ are difficult to obtain exactly, but they are easy to approximate with the following proposition, which is a corollary of the so-called Eneström-Kakeya theorem (see exercises 1 and 2 following theorem (30,2) in [2]).

Proposition: All the roots of a polynomial $a_0 + a_1z + \ldots + a_nz^n$, whose coefficients are real and positive, lie in the annulus with radii $\min \frac{a_k}{a_{k+1}}$ and $\max \frac{a_k}{a_{k+1}}$.

Proof: Consider a path in the complex plane consisting of the line segments joining successively a_0, a_1z, … , a_nz^n and suppose $\left|\frac{a_{k+1}z^{k+1}}{a_kz^k}\right| > 1$ for each k. Then the exterior angles at all vertices are congruent (to $\arg z$) and the path must spiral out. So, z cannot be a root. Therefore, if z is a root, this inequality must fail for some k, in which case $|z| \leqslant \frac{a_k}{a_{k+1}} \leqslant \max \frac{a_k}{a_{k+1}}$. Similarly, $\min \frac{a_k}{a_{k+1}} \leqslant |z|$ for any root z.

In other words, to obtain a bound for the roots of a polynomial with positive coefficients, just compute the ratios of consecutive coefficients and take their maximum and minimum. Computing the ratios of consecutive coefficients in $Q(z)$ shows that its roots must be between $\frac{6}{5}$ and 2 in absolute value. This is excellent news since, according to the example in the preliminaries, the coefficients of all the terms in the series expansion of the quintic part go to zero exponentially fast, and hence the unknown numerator in the partial fractions decomposition is irrelevant for a first-order approximation. So, the probability of rolling a single die and landing on Mayfair can be nicely approximated by the coefficient of z^{39} in the $\frac{2}{7}/(1 - z)$ term of $(\text{Sum of Rolls})(z)$, which is just $\frac{2}{7}$. This result makes intuitive sense since the probability coincides with the reciprocal of the average value of a roll (which also coincides with $\text{Roll}'(1)$). The exact probability of landing on Mayfair is $\frac{6365485061618916665100554384 17}{22279157564739556779731409960 96}$, which is less than 10^{-6} off the approximation.

Of course, most people play Monopoly with a pair of dice, rather than a single die. This has a minimal effect on our calculations. Instead of the generating function $\text{Roll}(z) = \frac{1}{6}z + \frac{1}{6}z^2 + \frac{1}{6}z^3 + \frac{1}{6}z^4 + \frac{1}{6}z^5 + \frac{1}{6}z^6$ for a single die, the generating function for a pair of dice is the same as the generating function for the sum of two dice $(\text{Roll}(z))^2$, as calculated above. Therefore, the probability of landing on Mayfair when rolling a pair of dice is the coefficient of z^{39} in $\frac{1}{1 - (\text{Roll}(z))^2}$. A partial fractions decomposition

https://doi.org/10.1017/mag.2016.19 Published online by Cambridge University Press

of this generating function has the form

$$\frac{\frac{1}{7}}{1 - z} + \frac{?}{Q(z)} + \frac{?}{1 + \operatorname{Roll}(z)}.$$

As before, the coefficients of the $Q(z)$ term go to zero exponentially fast. Likewise, by the proposition, the roots of $1 + \operatorname{Roll}(z)$ must be greater than 1 in absolute value, so the coefficients of the last term go to zero exponentially fast, as well. Therefore, an approximation of the probability of rolling a pair of dice and landing on Mayfair is $\frac{1}{7}$, a difference that turns out to be less than 10^{-4} off the exact value of

$$\frac{265132693875082928549304662 83}{185659646372829639831095083008}.$$

Passing Go

In the game of Monopoly, players collect a salary of £200 as they pass Go, which is forty sites from the start. A similar problem is calculating the expected number of rolls to pass Go and collect the salary. Let e_n denote the expected number of rolls needed to land on or first pass the nth site on the game board. This is equivalent to the expected number of rolls to obtain or first exceed sum n. As a first guess, one would suspect e_n would be close to n divided by the average value of a roll. The initial values are $e_n = 0$ for $n \leqslant 0$, and by taking conditions on the first roll

$$e_n = 1 + \sum_{m<n} e_m \, \mathrm{P}\big(\text{rolling } n - m\big), \qquad \text{for each } n \geqslant 1.$$

Let $E_1(z)$ denote the generating function of this sequence $\sum e_n z^n$ when rolling a single die. The generating function equivalent of this recurrence relation involves a geometric series with unit ratio and unit initial term, as well as the discrete convolution. So, $E_1(z) = \dfrac{z}{1 - z} + E_1(z) \operatorname{Roll}(z)$ and it follows that $E_1(z) = \dfrac{z}{(1 - z)(1 - \operatorname{Roll}(z))}$. Again, the generating function is rational, and it has a partial fractions decomposition of the form

$$E_1(z) = \frac{\frac{2}{7}}{(1 - z)^2} + \frac{\frac{4}{21}}{1 - z} + \frac{?}{Q(z)}.$$

The expected number of rolls to pass Go is the coefficient of z^{40} in $E_1(z)$. As noted in the preliminaries, the coefficient of the nth term of $\dfrac{1}{(1 - z)^2}$ is $n + 1$, the coefficient of the nth term of $\dfrac{1}{1 - z}$ is 1, and the coefficients of $\dfrac{?}{Q(z)}$ go to zero. Therefore, $e_n \approx \frac{2}{7}(n + 1) + \frac{4}{21}$. In other words, the expected number of rolls to reach sum n with a single die exceeds n divided by the average roll by about $\frac{10}{21}$. For passing Go with $n = 40$, the

https://doi.org/10.1017/mag.2016.19 Published online by Cambridge University Press

approximation of $\frac{250}{21}$ turns out to be within 10^{-6} of the exact expected number of rolls.

If a player rolls a pair of dice, the generating function for the sequence of expected number of rolls to pass a sum would be

$$E_2(z) = \frac{z}{(1 - z)\left(1 - (\mathrm{Roll}(z))^2\right)}.$$

This has a partial fractions decomposition of the form

$$E_2(z) = \frac{\frac{1}{7}}{(1 - z)^2} + \frac{\frac{29}{84}}{1 - z} + \frac{?}{Q(z)} + \frac{?}{1 + \mathrm{Roll}(z)}.$$

As before, we can ignore the terms with $Q(z)$ and $1 + \mathrm{Roll}(z)$ which go to zero, and so the expected number of rolls to reach sum n with a pair of dice is $e_n \approx \frac{1}{7}(n + 1) + \frac{29}{84}$. This time, the approximation of $\frac{521}{84}$ turns out to be within 10^{-4} of the exact expected number of rolls to pass Go with a pair of dice.

Further problems

There are many variations of these dice problems. For an example of a generalisation, you can show that if $\mathrm{Roll}(z)$ is the generating function for a *weighted* die (as opposed to a fair die), then the probability of obtaining the sum n is approximately $\dfrac{1}{\mathrm{Roll}'(1)}$. Actually, this approximation is a finite version of a theorem of Erdös, Feller and Pollard; see [3]. Likewise, the expected number of rolls needed to obtain or exceed sum n is approximately

$$\frac{n}{\mathrm{Roll}'(1)} + \frac{\mathrm{Roll}''(1)}{2(\mathrm{Roll}'(1))^2}.$$

In truth, the collection of dice problems is seemingly endless. (For an example of a large compilation with solutions, see [4].) Almost all of these problems can be solved, almost mechanically, with generating functions. If you find these generating functions intriguing, please read [5] and [6] for wonderful accounts.

References

1. P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press (2009).
2. M. Marden, *Geometry of polynomials*, Amer. Math. Soc. (1949).
3. P. Erdös, W. Feller and H. Pollard, A property of power series with positive coefficients, *Bull. of the Amer. Math. Soc.*, **55**(2) (1949) pp. 201-204.
4. Matthew M. Conroy, A collection of dice problems, available at: http://www.madandmoonly.com/doctormatt/mathematics/dice1.pdf

https://doi.org/10.1017/mag.2016.19 Published online by Cambridge University Press

5. S. K. Lando, *Lectures on generating functions*, Amer. Math. Soc. (2003).
6. H. S. Wilf, *Generatingfunctionology* (3rd edn.) (A. K. Peters Series, Taylor & Francis) (2005).

doi:10.1017/mag.2016.19

RUSSELL J. MAY
Department of Mathematics, 150 University Street, Morehead, KY 40351 USA
e-mail: *r.may@moreheadstate.edu*

100.11 A trinomial identity from both a tiling and an expansion

Introduction

It is well known that the number of ways of tiling a $1 \times n$ rectangle with 1×1 squares and 1×2 dominoes is given by F_{n+1}, where, for $n \geqslant 2$, the nth Fibonacci number F_n is defined via the recurrence relation $F_n = F_{n-1} + F_{n-2}$, with initial values $F_0 = 0$ and $F_1 = 1$ [1, 2]. We use this fact here in order to present a combinatorial proof of the following identity involving trinomial coefficients and Fibonacci numbers:

$$\sum \binom{n}{a,\, b,\, c} F_{m-2}^{a} F_{m-1}^{b} F_{m}^{c} F_{c} = F_{m}^{n} F_{2n}, \tag{1}$$

where the sum is be taken over all triplets of non-negative integers (a, b, c) such that $a + b + c = n$.

Admittedly, (1) might initially appear a somewhat intimidating identity to prove. However, we adapt here an idea given in [3] to provide a remarkably straightforward combinatorial proof of this result. Furthermore, we obtain an algebraic proof of (1) by utilising a suitable trinomial expansion.

A combinatorial proof

Let us consider the set $\mathcal{T}$ of tilings of a $1 \times (2n - 1)$ rectangle using 1×1 squares and 1×2 dominoes in which every element has the property that each of its first n tiles may take any one of F_m colours for some integer $m \geqslant 2$, while the remaining tiles are left uncoloured. Note here that each element of $\mathcal{T}$ will indeed always consist of at least n tiles. In order to prove (1) we will enumerate $\mathcal{T}$ in two different ways.

First, the factor F_{2n} on the right-hand side of (1) counts the number of possible uncoloured tilings of a $1 \times (2n - 1)$ rectangle using 1×1 squares and 1×2 dominoes. The factor F_m^n gives, for each element of $\mathcal{T}$, the number of ways in which the first n tiles may be coloured. It follows from this that the right-hand side of (1) enumerates $\mathcal{T}$.

Now consider the left-hand side of (1). We split the F_m colours into two classes, *light* and *dark* say, such that there are F_{m-2} colours of the former type and hence $F_m - F_{m-2} = F_{m-1}$ of the latter. Subject to the

https://doi.org/10.1017/mag.2016.19 Published online by Cambridge University Press