1

The Number One

All for one and one for all.
— *The Three Musketeers* by Alexandre Dumas

Even if you are a minority of one, the truth is the truth.
— Mahatma Gandhi

The number one seems like such an innocuous value. What can you do with only one thing? But the simplicity of one can be couched in a positive light: *uniqueness*. We all know that being single has its virtues. In a mathematical world with so many apparent options, having exactly one possibility is a valued commodity. A search of mathematical research papers and books found that more than 2,700 had the word "unique" in their titles. Knowing that there is a unique solution to a problem can imply structure and strategies for solving. Some (but not all) of the sections in this chapter explore how uniqueness arises in diverse mathematical contexts. This brings new meaning to looking for "the one."

Sliced Origami

Origami traditionally requires that one start with a square piece of paper and attain the final form by only folding. When mathematicians

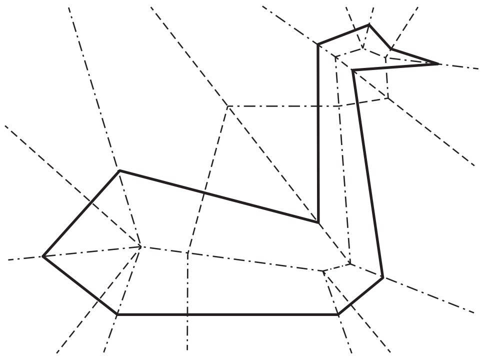

FIGURE 1.1: Origami fold pattern for a swan.

seriously entered the world of origami, they began to systematize constructions, including the use of computers to make precise fold patterns. Besides adding to the art, their contributions have also led to practical applications. For example, how can one transport a solar panel array into space? Mathematical origami has produced a design that compactly stores the whole assembly for transport. Once in space, the array is unfolded to its full size.

Children have long made beautiful paper snowflakes, but such constructions violate a fundamental origami tradition: no cutting, tearing, or gluing. But what if we were allowed a single cut? What new patterns could be attained? The surprising answer was found by Erik Demaine, a young Canadian professor at MIT, whose research intersects art, mathematics, and computer science. Demaine proved that *any* pattern whose boundary involves a finite number of straight line segments can be made by folding a paper appropriately and making a *single* cut! The possibilities include any polygon or multiple polygons. Figure 1.1 demonstrates the fold pattern needed to make a swan. After folding along each of the dashed and dashed-dot lines, the figure can be collapsed—with practice—so that a single cut along the bold line produces the swan.

Fibonacci Numbers and the Golden Ratio

The Fibonacci numbers are a sequence that has attracted attention from amateur investigators as well as seasoned mathematicians. As a reminder, the first two positive Fibonacci numbers are both 1, and each subsequent number in the sequence is formed by adding the previous two. This produces the sequence 1, 1, 2, 3, 5, 8, 13, 21, etc. Letting F_n denote the nth Fibonacci number, there is a tidy closed-form formula:

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

As n gets larger, the second term shrinks to zero, so F_n can be approximated as

$$F_n \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

This shrinkage implies that the ratio of successive terms satisfies

$$\frac{F_{n+1}}{F_n} \approx \frac{1+\sqrt{5}}{2} \tag{1.1}$$

The constant on the right—usually denoted by the Greek letter ϕ—is called the *Golden Ratio*. Connections of this number to art, architecture, and biological growth have long been studied, but the connection between ϕ and the number 1 is not as well known. Two beautiful formulas make the link apparent. The first is

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}} \tag{1.2}$$

An equivalent notation, which is more typographically cooperative, is

$$\phi = \frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\cdots$$

This formula takes the form of an infinite continued fraction. To understand this form, consider a finite version:

$$1+\cfrac{1}{1+\cfrac{1}{1+\frac{1}{1}}}=1+\cfrac{1}{1+\frac{1}{2}}=1+\frac{2}{3}=\frac{5}{3}$$

Note that in each step of simplification, the most "deeply buried" fraction—for example, 1/1 in the first expression, 1/2 in the second—is the ratio of two consecutive Fibonacci numbers. As the number of "plus ones" and fractions grows, the approximation (1.1) produces equation (1.2).

The second formula that connects ϕ and the number 1 involves nested square roots:

$$\phi=\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}} \tag{1.3}$$

Again, a finite counterpart works out to

$$\begin{aligned}\sqrt{1+\sqrt{1+\sqrt{1+1}}}&=\sqrt{1+\sqrt{1+\sqrt{2}}}\\&\approx\sqrt{1+\sqrt{2.414213562}}\\&\approx\sqrt{2.553773974}\\&\approx 1.598053182\end{aligned}$$

Unlike the situation with continued fractions, the numbers produced by adding 1 and taking square roots do not have a nice structure. To prove Equation 1.3, however, is not difficult. If we let $x=\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}}$, then

$$x^2=1+\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}}=1+x$$

and so $x^2 - x - 1$. Solving this quadratic equation—noting that $x > 0$—forces $x = \phi$.

Representing Numbers Uniquely

How many ways can a number be factored into a product of smaller numbers? Recall that a number is *prime* if it cannot be broken down. The first few primes are 2, 3, 5, 7, 11, and 13. The number 60 can be dissected in 10 different ways (where the factors are listed in nondecreasing order):

$$\begin{aligned} 2 \cdot 30 = 3 \cdot 20 = 4 \cdot 15 = 5 \cdot 12 = 6 \cdot 10 = 2 \cdot 2 \cdot 15 = 2 \cdot 3 \cdot 10 \\ = 2 \cdot 5 \cdot 6 = 3 \cdot 4 \cdot 5 = 2 \cdot 2 \cdot 3 \cdot 5 \end{aligned}$$

However, the last product is the sole way which involves only prime numbers. The Fundamental Theorem of Arithmetic claims that each number has a *unique* prime decomposition.

Factoring large numbers is a formidable problem; no efficient procedure for factoring is known. The most challenging numbers to factor are semiprimes, that is, numbers which are the product of two primes. Semiprimes play a huge role in cryptography, the science of making secret codes. This application of huge primes is important enough that the Electronic Frontier Foundation, concerned about Internet security, offers lucrative prize money for producing prime numbers with massive numbers of digits.

The Fundamental Theorem of Arithmetic no longer holds if we replace multiplication with addition. Even a lowly number like 16 cannot be written uniquely as the sum of two primes: $16 = 5 + 11 = 3 + 13$. What if we take only a *subset* of the natural numbers and insist that each number in this set is used at most once? The set of powers of two $\{1, 2, 4, 8, \ldots\}$ does the trick. For example, the number 45 can be written as $45 = 32 + 8 + 4 + 1 = 2^5 + 2^3 + 2^2 + 2^0$. This is equivalent to writing a number in base 2 since 45 in base 2 is simply 101101. Every number has a *unique* binary representation.

Another subset of the natural numbers that gives unique representations involves the Fibonacci numbers. Zeckendorf's Theorem

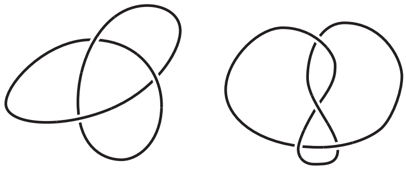

FIGURE 1.2: Trefoil knot (left) and figure-eight knot (right).

states that every positive integer can be represented uniquely as the sum of one or more distinct, nonconsecutive Fibonacci numbers. For example, $45 = 34 + 8 + 3 = F_9 + F_6 + F_4$. Note that we need to insist on having nonconsecutive Fibonacci numbers; otherwise, we could replace F_4 with $F_3 + F_2$ and have another way to represent 45. Although the Fibonacci numbers have been studied for roughly 800 years, Zeckendorf only discovered his result in 1939.

Factoring Knots

In the last section, we saw that every positive integer can be factored uniquely into a product of primes. This idea of decomposing objects into a set of fundamental pieces arises in some surprising contexts.

Knot theory aims to understand the structure of knots—think of lengths of string whose ends are tied together. How does one draw a three-dimensional knot in the plane? Imagine collapsing the knot onto a plane, being mindful of the over- and underpasses. The simplest configuration, a loop of string which makes a circle, isn't really a knot in the conventional sense, and is referred to as the *unknot*. The simplest nontrivial knots are the trefoil (the unique knot with three crossings) and the figure-eight knot (the unique knot with four crossings); see figure 1.2. The figure-eight knot is commonly used in both sailing and rock climbing.

Now comes the idea of decomposition. Imagine taking two nontrivial knots, making a cut in each one, then splicing the two knots together (figure 1.3). We call this a *composite knot*. One could take this process in the opposite direction as well; a knot could be "unspliced" or decomposed into two knots. We are not interested in the case where one (or both) of the new knots is the unknot; this is like saying that

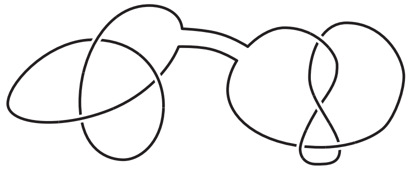

Figure 1.3: Adding the trefoil and the figure-eight knots.

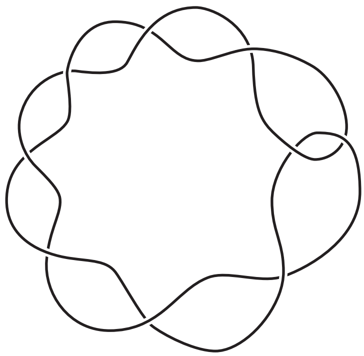

Figure 1.4: With one switch of a crossing, can you transform this into the unknot?

a number n can be written as $n \times 1$. If a knot cannot be unspliced, it is called a *prime knot*. The basic question then asks whether every knot can be unspliced into a set of prime knots, that is, does the Fundamental Theorem of Arithmetic extend to the knot setting? Yes! A theorem has shown that every knot has a *unique* prime decomposition. The order of the unsplicing also doesn't matter; regardless of how one unsplices, the end result is always the same set of prime knots.

There are other ways to ascribe complexity to a knot besides counting its crossings. Suppose we cut a knot to switch an overcrossing to an undercrossing (or vice versa). For a given knot, the minimum number of such switches needed to transform it into the unknot is its *unknotting number*. It may seem surprising to learn that there are knots with many crossings which have an unknotting number of 1. Figure 1.4 invites you to switch one crossing to transform this knot with nine crossings into the unknot. Such a situation is typically not recognized with only a casual once-over. Magicians can have fun by taking such knots, making the switch, and watching eyes bulge as a complex mass of crossings melts away to reveal a simple loop of rope. In general, it is relatively difficult to determine the unknotting number of a given knot. However, in 1985 it was shown that if the unknotting number of a knot is 1, then the knot is prime.

Counting and the Stern Sequence

The 19th century mathematician Georg Cantor shocked his contemporaries by developing a hierarchy of different kinds of infinity. Essentially, he came up with a new way to compare the sizes of two sets.

Let's start with a simple problem: How can you show that you have the same number of fingers as toes? Most people would argue, "I have 10 fingers and 10 toes, so they are the same number." This argument is fine, but it brings in an unnecessary concept; the actual *number* of toes and fingers. The question only asked to show that the two sets have the same size, not to actually count them. How else can one answer the question? By making a one-to-one correspondence between the toes and fingers, that is, pair each finger with exactly one toe. The left thumb could be paired with the large toe on the left foot, etc. With these pairings, we claim that the set of fingers has the same size—mathematicians say the same *cardinality*—as the set of toes. This idea of one-to-one correspondence—matching each member of one set to exactly one member of a second set—is what mathematicians formally use to claim that two sets have the same cardinality.

The one-to-one idea goes deeper when one encounters infinite sets. The set of positive integers has the same cardinality as the set of all nonzero integers. How is this possible since the first set fits into the second? Shouldn't the second set be twice as large as the first? Create a correspondence between the two sets:

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \cdots \\ -1 & 1 & -2 & 2 & -3 & 3 & \cdots \end{array}$$

Since every number in the first set matches with a number in the second set, the two sets have the same cardinality. More generally, any infinite

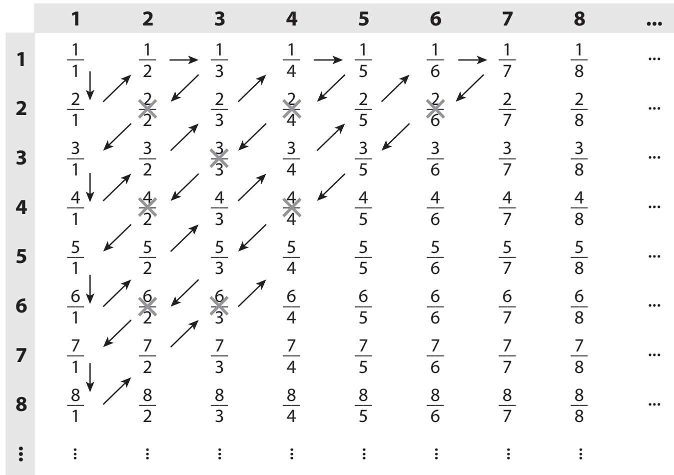

FIGURE 1.5: Counting the rational numbers.

set that can be "listed" has the same cardinality as the positive integers. Such sets are called countably infinite, or *countable* for short.

How about comparing the set of positive integers to the set of positive rational numbers? Cantor claimed that these two sets have the same size. Again, how can this be? There are infinitely many rational numbers between any two consecutive integers, making the claim seem ridiculous. The standard approach places the rationals on a grid (figure 1.5) and makes the correspondence by following the diagonal paths. The first few rationals are listed as $1, 2, \frac{1}{2}, \frac{1}{3}, 3, 4, \frac{3}{2}$, and $\frac{2}{3}$. Note from the figure that we have skipped over some numbers. For example, after the number 2/3 is encountered, it reappears as 4/6, 6/9, etc. Taking a first-come, first-served approach, we leapfrog over these latter incarnations of the same number. Each fraction will only be counted when it is in lowest terms.

How can one make the **one-to-one** correspondence *without* doing the skipping? One way involves what is called the *Stern sequence* of integers. This sequence is defined by $f(0) = 0$, $f(1) = 1$, and the two recurrence relations $f(2n) = f(n)$ and $f(2n+1) = f(n) + f(n+1)$. The first few terms are $0, 1, 1, 2, 1, 3, 2, 3, 1, 4, 3, 5, 2, 5, 3, 4$. It can be shown that any two neighboring numbers in this sequence are coprime, that is, they share no common factors. This observation eventually

Table 1.1

Stern Sequence.

n	0	1	2	3	4	5	6	7	8	9
$f(n)$	0	1	1	2	1	3	2	3	1	4
$f(n)/f(n+1)$	0	1	1/2	2	1/3	3/2	2/3	3	1/4	4/3

leads to the following remarkable theorem: the sequence of rationals generated by $f(n)/f(n+1)$ generates every positive rational number exactly **once**. We now have the desired correspondence between the positive integers and positive rationals. Table 1.1 shows the first few terms.

While Cantor's ideas about infinity are now considered part of the canon of mathematics, they were a great shock to his contemporaries. Poincaré considered Cantor's work a "grave disease" (Dauben, Georg Cantor, 1979, p. 266), and Kronecker asserted that Cantor was a "corrupter of youth" (Dauben, "Georg Cantor," 1977, p. 89). On the other hand, David Hilbert declared, "No one shall expel us from the Paradise that Cantor has created" (Hilbert, "Über das Unendliche," p. 170).

Fractals

The Ternary Cantor Set is one of the most studied oddball sets in mathematical analysis. To construct it, start by taking the interval $[0, 1]$ and removing its middle third, that is, the interval $[1/3, 2/3]$. Now remove the middle thirds of the remaining two intervals, specifically $[1/9, 2/9]$ and $[7/9, 8/9]$. Keep performing this removal process indefinitely (figure 1.6). It's reasonable to think that nothing will be left at the end of this process. Summing the lengths of the intervals removed, the geometric series helps us see that

$$\frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \cdots = 1$$

so the total length removed equals the total length of the interval. However, measuring intervals is not the same as measuring sets of

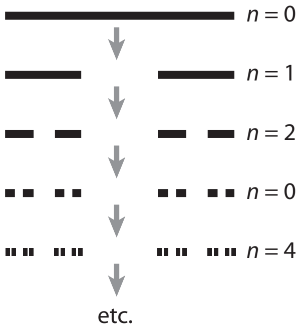

FIGURE 1.6: Approaching the Cantor Set.

points. There are actually infinitely many points left; this is the so-called *Cantor Set*. This set is so fine that it is sometimes called Cantor dust. Without diving into details, this set includes numbers that have an infinite base 3 expansion where none of the digits is a 1. An example is 7/10, whose base 3 expansion is $0.20022002200 \cdots$.

The Cantor Set has another interesting property. Make a copy of the set and scale it by a factor of 1/3. Make another copy, scale it by a factor of 1/3, then slide it to the right by 2/3. The union of these two shrunken sets is exactly the original Cantor Set. When a set can be written as the union of a finite number of shrunken copies of itself, we say it exhibits *self-similarity*. Could we start with another nonempty set, make two copies, scale and shift as before, and get the original set back? No. According to Hutchinson's Theorem, the set of transformations—the rules to shrink and slide copies of a set—*uniquely* determines a set that is self-similar under those transformations.

The process described in generating the Cantor Set can be generalized to build many other strange-looking sets. These sets are more striking when we consider sets of points in the plane. For example, imagine carving an equilateral triangle into four subtriangles and removing the middle one. Now take each of the remaining subtriangles, carve each of them into four pieces, and remove their middle triangles. Continuing this process indefinitely produces the Sierpiński Gasket (figure 1.7).

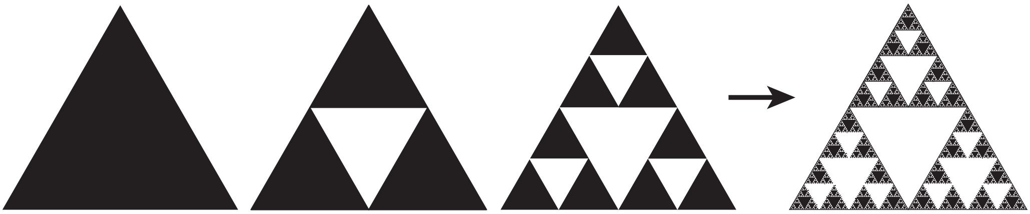

Figure 1.7: Approaching the Sierpiński Gasket.

Self-similarity is also evident with the Sierpiński Gasket: three shrunk copies can be patched together to produce the original figure. While programming the computer to draw a decent approximation of sets like the Sierpiński Gasket is not too taxing, a simpler way is to use the *chaos game*. Consider three possible rules that could be applied to a point in the plane:

1. change (x, y) to $(\frac{x}{2}, \frac{y}{2})$
2. change (x, y) to $(\frac{x}{2} + \frac{1}{2}, \frac{y}{2})$
3. change (x, y) to $(\frac{x}{2} + \frac{1}{4}, \frac{y}{2} + \frac{\sqrt{3}}{4})$

The first rule takes a point and moves it so that it is one half as far from the origin. The second rule does the same but also slides the point to the right by $1/2$. The third rule is also the same as the first rule but also slides the point to the right by $1/4$ and up by $\sqrt{3}/4$.

Each of these rules is inspired by the self-similarity of the set. How does the chaos game work? Pick any point in the plane and randomly apply one of the three rules. Now randomly pick a rule again and apply it to the new point. After repeating this process, say 100 times, start plotting the points (figure 1.8). The Sierpiński Gasket slowly emerges.

In general, suppose one has a finite number of transformations, each of which involves shrinking possibly followed by a shifting and rotation. By starting with any point and applying one of the rules (randomly chosen at each step) many times, Hutchinson's Theorem guarantees that a *unique* set called an *attractor* will be filled up. The structure of the

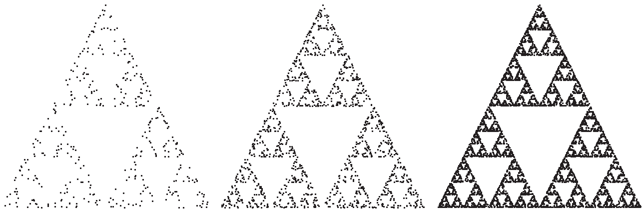

FIGURE 1.8: The chaos game produces the Sierpiński Gasket.

rules guarantees that the attractor will be self-similar. These self-similar sets are called *fractals*. Many intricate sets can be easily generated in this way, including the Barnsley fern and three-dimensional fractals like the Menger sponge (figure 1.9).

Gilbreath's Conjecture

While finding patterns in the primes is something of a Holy Grail, every attempt to find simple order in these numbers has led researchers back to square one—quite literally in the case of Gilbreath's Conjecture. Make a list of the first few primes, then take the absolute value of the differences between successive terms. Then do it again, and again, etc. Table 1.2 contains the first few rows.

Do you notice a pattern? Each row begins with the number 1. And this is not a coincidence because we've only done a few rows. A computer search has verified that the first entry in each row is a 1 for about 3.4×10^{11} rows. Gilbreath's Conjecture states that the first entry will always be a 1. Despite this problem's deceiving simplicity, a single-minded focus is required to crack this conundrum.

Benford's Law

Given a random positive integer, what is the probability that the first digit is a 1? One out of nine, of course. And there's nothing special about the number 1: each of the other digits also has a one out of nine chance of being encountered. What if we change the scenario,

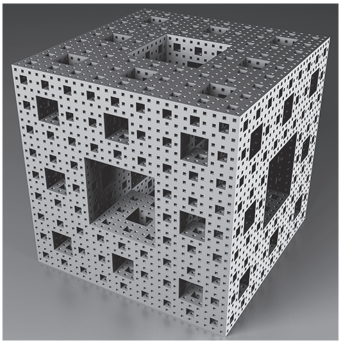

FIGURE 1.9: The Barnsley fern (left) and the Menger sponge (right).

Table 1.2
Gilbreath's Conjecture

2	3	5	7	11	13	17	19	23	29	31
1	2	2	4	2	4	2	4	6	2	6
1	0	2	2	2	2	2	2	4	4	2
1	2	0	0	0	0	0	2	0	2	0
1	2	0	0	0	0	2	2	2	2	0
1	2	0	0	0	2	0	0	0	2	0
1	2	0	0	2	2	0	0	2	2	2
1	2	0	2	0	2	0	2	0	0	0

however, so that we measure population sizes of cities and towns. The likelihood of having a 1 in the first digit is no longer $1/9 \approx 11\%$, but around 30%. And it doesn't apply only to city sizes; income

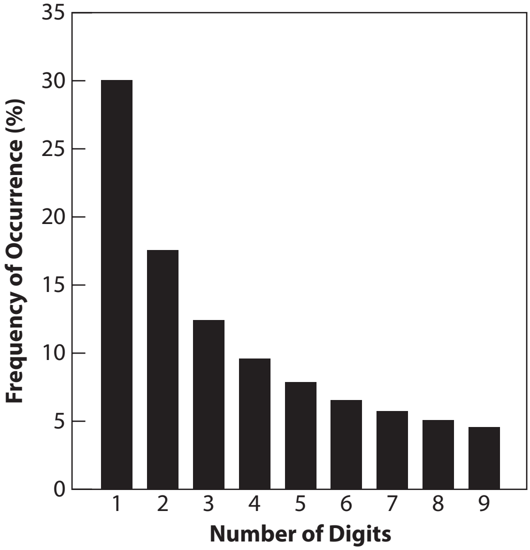

FIGURE 1.10: Distribution of first digits according to Benford's law.

taxes, street addresses, Fibonacci numbers, lengths of rivers, and many other phenomena betray the same bias. To be more specific, Benford's law states that the first digit in these phenomena takes the value n with probability $\log_{10}(1 + 1/n)$. Calculated percentages appear in figure 1.10.

It's easy to see that the sum S of these probabilities equals 1:

$$
\begin{aligned}
S = {} & \log_{10}\left(1+\frac{1}{1}\right) + \log_{10}\left(1+\frac{1}{2}\right) + \log_{10}\left(1+\frac{1}{3}\right) \\
& + \cdots + \log_{10}\left(1+\frac{1}{9}\right) \\
= {} & \log_{10}(2) + \log_{10}\left(\frac{3}{2}\right) + \log_{10}\left(\frac{4}{3}\right) + \cdots + \log_{10}\left(\frac{10}{9}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \log_{10}(2) + \log_{10}(3) - \log_{10}(2) + \log_{10}(4) - \log_{10}(3) \\
&\quad + \cdots + \log_{10}(10) - \log_{10}(9) \\
&= 1
\end{aligned}
$$

The first recorded observation related to this phenomenon was made by the U.S. astronomer Simon Newcomb in 1881. He noticed that the first pages of books of logarithms—such books were needed for calculations—were soiled much more than the later pages. Only in the 1930s did Frank Benford rediscover this observation, and he subsequently measured a large number of data sets, both artificially and naturally constructed. In general, it seems that phenomena that have some kind of power law growth are subject to Benford's law.

This nonintuitive bias of small digits has been used in the service of the law. In attempts to make falsified documents look authentic, tax cheats massage their numbers. In an effort to make their cooked data look real, they manufacture their numbers in a random way. Such violations of Benford's law have triggered audits.

The Brouwer Fixed-Point Theorem

Suppose you have two identical sheets of graph paper. Lay one sheet on a table, but take the other, crumple it up, and lay it on the first sheet so that none of it is hanging over the side. The claim is that there is at least one point on the crumpled paper that is fixed, that is, directly over its "twin" on the flat paper. This is a special instance of the Brouwer Fixed-Point Theorem. Unfortunately, the theorem is not constructive; we have no idea where the fixed point is! Fixed-point theorems like Brouwer's have been used in diverse settings, including mathematical economics. In contrast to the uniqueness theorems witnessed in some of the previous sections, fixed-point theorems are existence results; uniqueness theorems claim that *at most one* of something exists, while existence results assert that *at least one* exists.

A variation of the Brouwer Fixed-Point Theorem is called the Hairy Ball Theorem. Suppose each point on a ball has a short hair pointing outward and the direction of each hair changes in a continuous way. The theorem states that at least one hair must point straight up. You can picture this by trying to comb a coconut. As an aside, there is no such thing as a Hairy Donut Theorem; one could comb all the hairs on a donut flat and in the same direction so that no hair is sticking up.

Inverse Problems

The problem "Given x, find x^2" is a straightforward multiplication problem. Given the number 13, its square is 169. Turning this around, suppose one asks, "Given x, find y such that $x = y^2$." If $x = 169$, then $y = \pm 13$. A host of concerns arise with this second problem. As just witnessed, the answer may not be unique. In fact, a solution may not exist: if x is a negative number, there is no number y whose square equals x. Moreover, even if a solution exists, computing the value (or an approximation) requires much more work; calculating square roots is much more computationally demanding than multiplication. The original squaring problem would be called the *forward problem*, and finding the square root would be the corresponding *inverse problem*.

Inverse problems are usually difficult to solve and analyze. Solutions exist under restricted conditions, and constructing such solutions is usually orders of magnitude more computationally expensive (they require much more computer memory and time) than the corresponding forward problem. Of course, looking for a solution makes no sense if such a solution does not exist, and it may be more difficult to find if the solution is not unique.

Let us consider a more sophisticated inverse problem. Suppose that one could measure the darkness level at each point of a photograph. This information could be used to find the average darkness along any line that intersects the picture. Now "invert" the problem. Suppose the average darkness along any line is known. Can we use this knowledge to find the darkness at each point? The mathematical equivalent of this question was affirmatively answered by Johann Radon in 1917. In fact, the Radon transform is a formula that produces the *unique* darkness levels.

This seemingly theoretical result has enjoyed widespread usage. The first real-world application was seen about 50 years later and concerns medical imaging. Take a cross section of the human body (figuratively, of course). Instead of darkness levels, we are concerned with tissue density at each point. By firing an X-ray beam of known intensity through the body and measuring the reduced intensity when the beam exits, one can calculate the average density of the tissue along that line. Repeat this measurement of average density along all possible lines in the cross section. The Radon transform uses these measurements to reconstruct the density of tissue at each point. By performing this process on stacked cross sections of the body, an image of the body's density at each point is formed. Armed with this information, analysts can detect tumors. Early use of this technique yielded a breakthrough in the diagnosis of neural diseases. In this way, the area of nondestructive medical imaging was born. Alan Cormack and Godfrey Hounsfield won the Nobel Prize in Physiology or Medicine in 1979 for their seminal contributions to this area.

Solving inverse problems has become wildly successful. A key to making the theory work is when the problems have *unique* solutions. Inverse problems also have many applications beyond medicine. Seismic imaging is an inverse problem on a large scale. If a sound wave is blasted into the ground and the variable density of the rock layers is known, one can predict the intensity of the reflected sound waves coming up from the surface. The inverse problem measures the reflected waves and mathematically reconstructs the density of the layers beneath, without digging! This is the modern way of looking for oil. Another inverse problem concerns crack detection. It has become necessary to check engine blocks, cylinder heads, and crankshafts for structural flaws. Similar in spirit to seismic imaging, nondestructive electrostatic measurements reveal the structure of the materials. Essentially, these techniques have proven to be useful in physical situations that value noninvasion. But the story is not over. The theorems behind these applications require *exact measurements*, something impossible to achieve in practice. Limited and noisy data produce images with blurring, streaking, phantoms, and other artifacts.

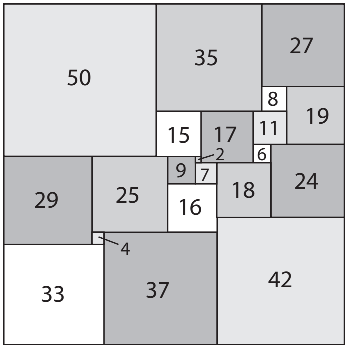

FIGURE 1.11: Duijvestijn's dissection of a square into 21 squares.

More first-class research is needed to develop numerically stable methods to approximate the exact, *unique* solutions.

Perfect Squares

A square that can be dissected into a finite collection of distinct, smaller squares is called a *perfect square*. If no subset of the squares forms a rectangle, then the perfect square is called *simple*.

The Russian mathematician Nikolai Luzin claimed that perfect squares were impossible to construct, but this assertion collapsed when a 55-square perfect square was published by R. Sprague in 1939. In 1978, A.J.W. Duijvestijn delivered a one–two punch by constructing a 21-square simple perfect square (figure 1.11). The number in each square represents the side length of that square. This dissection is *unique* among simple perfect squares of order 21, and there is no simple perfect square of smaller order.

The Bohr–Mollerup Theorem

Students of mathematics usually first encounter the factorial function in conjunction with counting permutations. How many ways can one order the ten letters $\{a, b, c, d, e, f, g, h, i, j\}$? In the first position, there are ten possibilities, leaving nine for the second slot, eight for

Table 1.3
Factorial Function Values

n	0	1	2	3	4	5	6	7	8	9
$n!$	1	1	2	6	24	120	720	5,040	40,320	362880

the third slot, etc. This means that the total number of orderings is $10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdots 3 \cdot 2 \cdot 1 = 3,628,800$. The factorial function is used in many areas of mathematics. Table 1.3 displays how the factorial function grows very quickly. The factorial function can be calculated recursively by using $(n + 1)! = (n + 1) \times n!$. For reasons connected to combinatorial problems, we define $0! = 1$.

The legendary 18th century Swiss mathematician Leonhard Euler thought about how one could extend the factorial function to the positive real numbers. Using table 1.3, we want to "connect the dots" of the points (0,1), (1,1), (2,2), (3,6), (4,24), etc. Of course, there are infinitely many ways to do this, but we would like a way that forces "nice properties" on the resulting function. Euler defined the gamma function—see figure 1.12—as

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

This function enjoys two factorial-like properties: $\Gamma(1) = 1$ and $\Gamma(x + 1) = x\Gamma(x)$ for all $x > 0$. These two properties may be combined to show that $\Gamma(n) = (n - 1)!$ for all positive integers n:

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1)\\ &= (n-1)(n-2)\Gamma(n-2)\\ &= (n-1)(n-2)(n-3)\cdots 2 \times 1 \cdot \Gamma(1)\\ &= (n-1)!\end{aligned}$$

Although these two properties restrict the number of possible extensions to the factorial function, there are still many that work. Bohr and Mollerup noted another property satisfied by the gamma function:

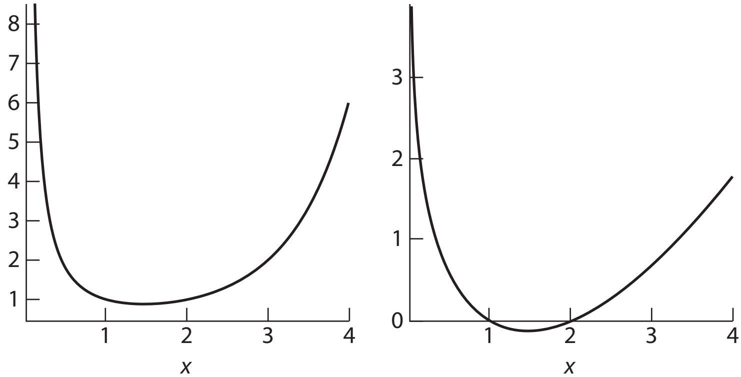

FIGURE 1.12: The functions $\Gamma(x)$ and $\log \Gamma(x)$.

the graph of $\Gamma(x)$ is *log convex*, that is, the function $\log \Gamma(x)$ is convex. To say that a function $f(x)$ is convex means that if $a < b$, then the line segment joining the points $(a, f(a))$ and $(b, f(b))$ lies above the graph of $y = f(x)$. So what makes the gamma function special among all extensions of the factorial function? The Bohr–Mollerup Theorem states that the gamma function is the *unique* function $f(x)$ that is log convex on the positive real axis and that satisfies $f(1) = 1$ and $f(x + 1) = x f(x)$ for all $x > 0$.

The gamma function is widely used in number theory and analysis. Besides the familiar trigonometric functions, for example, sine and cosine, the gamma function is probably the most commonly used special function. Even in statistics, the nonobvious fact that $\Gamma(1/2) = \sqrt{\pi}$ is used in the formula for normal distributions.

The Picard Theorems

When a math student first encounters functions, the concepts of domain (the set of allowable input values) and range (the set of corresponding outputs) are learned. If the domain of a real-valued function is the whole real line, the range can be as large or as small as possible. Simple examples include $f(x) = 5$, whose range is the set $\{5\}$ with just one element, and $f(x) = x$, whose range is the whole real line. What about something in between? The range of $f(x) = 1/(1 + x^2)$ is

$(0, 1]$ (including 1 but excluding 0), while the range of both $\sin(x)$ and $\cos(x)$ equals $[-1, 1]$, and the range of $f(x) = e^x$ is $(0, \infty)$.

Things get more complicated, however, if we extend these functions to having a *complex* variable. Not surprisingly, since the domain has grown, the range generally also grows. Nonconstant polynomials have all complex numbers in their range, a simple consequence of the Fundamental Theorem of Algebra. Both $\sin(z)$ and $\cos(z)$—functions that have a bounded range if the input is real—now also have all complex numbers in their ranges. Some functions that are well-defined on the real line cannot be extended to the whole complex plane, such as $f(z) = 1/(1 + z^2)$, which is not defined at $z = \pm i$.

The Little Picard Theorem makes a broad claim: if a nonconstant function's domain is the whole complex plane and it is differentiable at every point, then the range of the function is the whole complex plane, possibly minus **one** point. For example, the function $f(z) = e^z$ satisfies the conditions of the theorem, and its range contains all complex numbers except for the value 0. So why is this result called the "Little" Picard Theorem? What's so little about it? It's not little, per se; it's simply not as sweeping as a similar result called the Big (or sometimes Great) Picard Theorem. We need more terminology to state this result.

Like functions of a real variable, functions of a complex variable may not be defined at some points. Such points are called *singularities* of a function. Sometimes a singularity is simply a hole in the graph of the function that can be papered over. An example is the function $f(z) = \sin(z)/z$. This function is not defined at $z = 0$, but as z approaches zero, the function's value approaches 1. We call this a *removable singularity*. Another kind of singularity—what mathematicians typically have in mind—is called a *pole*. As an example, $z = 0$ is a pole for the function $1/z$. This is a pole of order 1, while in general a function has a pole of order m at $z = z_0$ if m is the smallest positive integer so that $f(z)(z - z_0)^m$ is either defined or has a removable singularity at $z = z_0$. Finally, we have the scary situation of singularities that are so strong that they are not poles of any order. These are *essential singularities*. An example is $z = 0$ for the function $\exp(1/z)$.

Now what does the Big Picard Theorem claim? Suppose that the function f has an essential singularity at $z = z_0$. Consider a punctured disk centered at z_0, that is, the disk with the center removed. Then the theorem asserts that if we restrict the domain to the punctured disk, the range of the function f is all possible complex values, with *at most one* exception. No matter how small the disk is made, every possible value, with at most one exception, is in the range. The essential singularity $z = 0$ for the function $\exp(1/z)$ misses the value 0 in its range, but it hits every value infinitely many times.