THE GREAT COURSES®

Topic
Science & Mathematics

Subtopic
Mathematics

The Joy of Mathematics

Course Guidebook

Professor Arthur T. Benjamin
Harvey Mudd College

PUBLISHED BY:

THE GREAT COURSES
Corporate Headquarters
4840 Westfields Boulevard, Suite 500
Chantilly, Virginia 20151-2299
Phone: 1-800-832-2412
Fax: 703-378-3819
www.thegreatcourses.com

Printed in the United States of America

Arthur T. Benjamin, Ph.D.

Professor of Mathematics
Harvey Mudd College

Arthur T. Benjamin is a Professor of Mathematics at Harvey Mudd College. He graduated from Carnegie Mellon University in 1983, where he earned a B.S. in Applied Mathematics with university honors. He received his Ph.D. in Mathematical Sciences in 1989 from Johns Hopkins University, where he was supported by a National Science Foundation graduate fellowship and a Rufus P. Isaacs fellowship. Since 1989, Dr. Benjamin has been a faculty member of the Mathematics Department at Harvey Mudd College, where he has served as department chair. He has spent sabbatical visits at Caltech, Brandeis University, and University of New South Wales in Sydney, Australia.

In 1999, Professor Benjamin received the Southern California Section of the Mathematical Association of America (MAA) Award for Distinguished College or University Teaching of Mathematics, and in 2000, he received the MAA Deborah and Franklin Tepper Haimo National Award for Distinguished College or University Teaching of Mathematics. He was named the 2006–2008 George Pólya Lecturer by the MAA.

Dr. Benjamin's research interests include combinatorics, game theory, and number theory, with a special fondness for Fibonacci numbers. Many of these ideas appear in his book (co-authored with Jennifer Quinn), *Proofs That Really Count: The Art of Combinatorial Proof* published by the MAA. In 2006, that book received the Beckenbach Book Prize by the MAA. Professors Benjamin and Quinn are the co-editors of *Math Horizons* magazine, published by MAA and enjoyed by more than 20,000 readers, mostly undergraduate math students and their teachers.

Professor Benjamin is also a professional magician. He has given more than 1,000 "mathemagics" shows to audiences all over the world (from primary schools to scientific conferences), where he demonstrates and explains his calculating talents. His techniques are explained in his book *Secrets of Mental Math: The Mathemagician's Guide to Lightning Calculation and Amazing Math Tricks*. Prolific math and science writer Martin Gardner calls it "the clearest, simplest, most entertaining, and best book yet on the art of calculating in your head." An avid games player, Dr. Benjamin was winner of the American Backgammon Tour in 1997.

Professor Benjamin has appeared on dozens of television and radio programs, including the *Today Show*, CNN, and National Public Radio. He has been featured in *Scientific American*, *Omni*, *Discover*, *People*, *Esquire*, *The New York Times*, the *Los Angeles Times*, and *Reader's Digest*. In 2005, *Reader's Digest* called him "America's Best Math Whiz." ■

Table of Contents

Table of Contents

Table of Contents

SUPPLEMENTAL MATERIAL

The Joy of Mathematics

Scope:

For most people, mathematics is little more than counting: basic arithmetic and bookkeeping. People might recognize that numbers are important, but most cannot fathom how anyone could find mathematics to be a subject that can be described by such adjectives as *joyful*, *beautiful*, *creative*, *inspiring*, or *fun*. This course aims to show how mathematics—from the simplest notions of numbers and counting to the more complex ideas of calculus, imaginary numbers, and infinity—is indeed a great source of joy.

Throughout most of our education, mathematics is used as an exercise in disciplined thinking. If you follow certain procedures carefully, you will arrive at the right answer. Although this approach has its value, I think that not enough attention is given to teaching math as an opportunity to explore creative thinking. Indeed, it's marvelous to see how often we can take a problem, even a simple arithmetic problem, solve it lots of different ways, and always arrive at the same answer. This internal consistency of mathematics is beautiful. When numbers are organized in other ways, such as in Pascal's triangle or the Fibonacci sequence, then even more beautiful patterns emerge, most of which can be appreciated from many different perspectives. Learning that there is more than one way to solve a problem or understand a pattern is a valuable life lesson in itself.

Another special quality of mathematics, one that separates it from other academic disciplines, is its ability to achieve absolute certainty. Once the definitions and rules of the game (the rules of logic) are established, you can reach indisputable conclusions. For example, mathematics can prove, beyond a shadow of a doubt, that there are infinitely many prime numbers and that the Pythagorean theorem (concerning the lengths of the sides of a right triangle) is absolutely true, now and forever. It can also "prove the impossible," from easy statements, such as "The sum of two even numbers is never an odd number," to harder ones, such as "The digits of pi (π) will never repeat." Scientific theories are constantly being refined and improved and, occasionally, tossed aside in light of better evidence. But a mathematical theorem is true forever. We still marvel over the brilliant logical arguments put forward by the ancient Greek mathematicians more than 2,000 years ago.

From backgammon and bridge to chess and poker, many popular games utilize math in some way. By understanding math, especially probability and combinatorics (the mathematics of counting), you can become a better game player and win more.

Of course, there is more to love about math besides using it to win games, or solve problems, or prove something to be true. Within the universe of numbers, there are intriguing patterns and mysteries waiting to be explored. This course will reveal some of these patterns to you.

In choosing material for this course, I wanted to make sure to cover the highlights of the traditional high school mathematics curriculum of algebra, geometry, trigonometry, and calculus, but in a nontraditional way. I will introduce you to some of the great numbers of mathematics, including π, e, i, 9, the numbers in Pascal's triangle, and (my personal favorites) the Fibonacci numbers. Toward the end of the course, as we explore notions of infinity, infinite series, and calculus, the material becomes a little more challenging, but the rewards and surprises are even greater.

Although we will get our hands dirty playing with numbers, manipulating algebraic expressions, and exploring many of the fundamental theorems in mathematics (including the fundamental theorems of arithmetic, algebra, and calculus), we will also have fun along the way, not only with the occasional song, dance, poem, and lots of bad jokes, but also with three lectures exploring applications to games and gambling. Aside from being a professor of mathematics, I have more than 30 years experience as a professional magician, and I try to infuse a little bit of magic in everything I teach. In fact, the last lesson of the course (which you could watch first, if you want) is on the joy of mathematical magic.

Mathematics is food for the brain. It helps you think precisely, decisively, and creatively and helps you look at the world from multiple perspectives. Naturally, it comes in handy when dealing with numbers directly, such as when you're shopping around for the best bargain or trying to understand the statistics you read in the newspaper. But I hope that you also come away from this course with a new way to experience beauty, in the form of a surprising pattern or an elegant logical argument. Many people find joy in fine music, poetry, and other works of art, and mathematics offers joys that I hope you, too, will learn to experience. If Elizabeth Barrett Browning had been a mathematician, she might have said, "How do I count thee? Let me love the ways!" ■

The Joy of Math—The Big Picture

Lecture 1

For many people, "math" is a four-letter word—something to be afraid of, not something to be in love with. Yet, in these lectures, I hope to show you why mathematics is indeed something to love.

To many people, the phrase "joy of mathematics" sounds like a contradiction in terms. For me, however, there are many reasons to love mathematics, which I sum up as the ABCs: You can love mathematics for its applications, for its beauty (and structure), and for its certainty.

What are some of the applications of mathematics? It is the language of science: The laws of nature, in particular, are written in calculus and differential equations. Calculus tells us how things change and grow over time, modeling everything from the motion of pendulums to galaxies. On a more down-to-Earth level, mathematics can be used to model how your money grows. This course discusses the mathematics of compound interest and how it connects to the mysterious number e.

Mathematics can bring order to your life. As an example, consider the number of ways you could arrange eight books on a bookshelf. Believe it or not, if you arranged the books in a different order every day, you would need 40,320 days to arrange them in all possible orders!

Mathematics is often taught as an exercise in disciplined thinking; if you don't make any mistakes, you'll always end up with the same answer. In this way, mathematics can train people to follow directions carefully, but mathematics should also be used as an opportunity for creative thinking. One of the life lessons that people can learn from mathematics is that problems can be solved in several ways.

As a child, I remember thinking about the numbers that add up to 20; specifically, I wondered what two numbers that add up to 20 would have the greatest product. The result of multiplying 10×10 is 100, but could two other numbers that add up to 20 have a greater product? I tried various combinations, such as 9 × 11, 8 × 12, 7 × 13, 6 × 14, and so on. For 9 × 11, the answer is 99, just 1 shy of 100. For 8 × 12, the answer is 96, 4 shy of 100.

As I continued, I noticed two things. First, the products of those numbers get progressively smaller. Second, and more interesting, the result of each multiplication is a perfect square away from 100. In other words, 9 × 11 is 99, or 1 (1^2), away from 100; 8 × 12 is 96, or 4 (2^2), away from 100; and so on. I then tried the same experiment with numbers that add up to 26. Starting with 13 × 13 = 169, I found that 12 × 14 = 168, just shy of 169 by 1. The next combination was 11 × 15 = 165, shy of 169 by 4, and the pattern continued. I also found that I could put this pattern to use. If I had the multiplication problem 13 × 13, I could substitute an easier problem, 10 × 16, and adjust my answer by adding 9. Because 10 and 16 are each 3 away from 13, all I had to do was add 3^2, which is 9, to arrive at the correct answer for 13 × 13, which is 169.

Perhaps nothing is more intriguing in mathematics than the notion of infinity.

In this course, we'll go into more detail about how to square numbers and multiply numbers in your head faster than you ever thought possible. Let's look at one more example here. Let's multiply two numbers that are close to 100, such as 104 and 109. The first number, 104, is 4 away from 100, and the second, 109, is 9 away from 100. The first step is to add 104 + 9 (or 109 + 4) to arrive at 113 and keep that answer in mind. Next, multiply the two single-digit numbers: 4 × 9 = 36. Believe it or not, you now have the answer to 104 × 109, which is 11,336. We'll see why that works later in this course.

Another creative use of mathematics is in games. By understanding such areas of math as probability and *combinatorics* (clever ways of counting things), you can become a better game player. In this course, we'll use math to analyze poker, roulette, and craps.

Throughout the course, you'll be exposed to ideas from high school– and college-level mathematics all the way to unsolved problems in mathematics.

You'll learn the fundamental theorem of arithmetic, the fundamental theorem of algebra, and even the fundamental theorem of calculus. Along the way, we'll encounter some of the great historical figures in mathematics, such as Euclid, Gauss, and Euler. You'll learn why 0.999999… going on forever is actually equal to the number 1; it's not just close to 1, but equal to it. You will also come to understand why $\sin^2 + \cos^2 = 1$, and you'll be able to follow the proof of the Pythagorean theorem and know why it's true.

As I said above, the B in the ABCs of loving mathematics is its beauty. We'll study some of the beautiful numbers in mathematics, such as *e*, pi (π), and *i*. We'll see that *e* is the most important number in calculus. Pi, of course, is the most important number in geometry and trigonometry. And *i* is the imaginary number, whose square is equal to −1. We'll also look at some beautiful and useful mathematical formulas, such as $e^{\pi i} + 1 = 0$. That single equation uses *e*, pi, *i*, 1, and 0—arguably the five most important numbers in mathematics—along with addition, multiplication, exponentiation, and equality.

Another beautiful aspect of mathematics is patterns. In fact, mathematics is the science of patterns. We'll have an entire lecture devoted to Pascal's triangle, which contains many beautiful patterns. Pascal's triangle has 1s along the borders and other numbers in the middle. The numbers in the middle are created by adding two adjacent numbers and writing their total underneath. We can find one pattern in this triangle if we add the numbers across each row. The results are all powers of 2: 1, 2, 4, 8, 16, … . The diagonal sums in this triangle are all *Fibonacci numbers*: 1, 1, 2, 3, 5, 8, 13, … . We'll discuss these mysterious numbers in detail.

Perhaps nothing is more intriguing in mathematics than the notion of infinity. We'll study infinity, both as a number-like object and as the size of an object. We'll see that in some cases, one set with an infinite number of objects may be substantially more infinite than another set with an infinite number of objects. There are actually different levels of infinity that have many beautiful and practical applications. We'll also have some fun adding up infinitely many numbers. We'll see two ways of showing that the sum of a series of fractions whose denominators are powers of 2, such as 1 + 1/2 +

1/4 + 1/8 + 1/16 + …, is equal to 2. We'll see this result both from a visual perspective and from an algebraic perspective.

Paradoxically, we'll look at a simpler set of numbers, 1 + 1/2 + 1/3 + 1/4 + 1/5 + … (called the *harmonic series*), and we'll see that even though the terms are getting smaller and closer to 0, in this case, the sum of those numbers is actually infinite. In fact, we'll encounter many paradoxes once we enter the land of infinity. We'll find an infinite collection of numbers such that when we rearrange the numbers, we get a different sum. In other words, when we add an infinite number of numbers, we'll see that the commutative law of addition can actually fail.

Another problem we'll explore in this course has to do with birthdays. How many people would you need to invite to a party to have a 50% chance that two people will share the same birth month and day? Would you believe that the answer is just 23 people?

The C in the ABCs of loving mathematics is certainty. In no other discipline can we show things to be absolutely, unmistakably true. For example, the Pythagorean theorem is just as true today as it was thousands of years ago. Not only can you prove things with absolute certainty in mathematics, but you can also prove that certain things are impossible. We'll prove, for example, that $\sqrt{2}$ is irrational, meaning that it cannot be written as a fraction with an integer (a whole number) in both the numerator and the denominator.

Keep in mind that you can skip around in these lectures or view certain lectures again. In fact, some of these lectures may actually make more sense to you after you've gone beyond them, then come back to revisit them.

What are the broad areas that we'll cover in this course? We'll start with the joy of numbers, the joy of primes, the joy of counting, and the joy of the Fibonacci numbers. Then we'll have a few lectures about the joy of algebra because that's one of the most useful mathematics courses beyond arithmetic. We'll talk a little bit about the joy of 9 before we turn to the joy of proofs, geometry, and the most important number in geometry, pi.

In the second half of the course, we'll learn about trigonometry, including sines, cosines, tangents, triangles, and circles, and we'll learn about the joy of the imaginary number *i*, whose square is −1. After *i*, we'll learn about the number *e*. We'll talk about the joy of infinity and the joy of infinite series, setting the stage for three lectures on the joy of calculus. After that, we'll study the glory of Pascal's triangle and apply some of the ideas we've learned to the joy of probability and the joy of games. We'll end the course with a mathematical magic show. And that's one more reason to love math: Mathematics is truly magical.

We close our first lecture with a mathematics analogy. Later in the course, we'll see the trigonometric function, the sine function, in three different ways. We'll see, for instance, in terms of a right triangle, that the sine of an angle *a* is equal to the length of the opposite side divided by the length of the hypotenuse. We'll think of the sine function in terms of the unit circle that is given in angle *a*; the sine will be the *y* coordinate of the point (*x*, *y*) on the unit circle corresponding to that angle. We'll even look at the sine function from an algebraic standpoint. That is, we'll be able to calculate the sine of *a*. For any angle *a* written in radians, we can write $\sin a = a - \frac{a^3}{3!} + \frac{a^5}{5!} - \frac{a^7}{7!} + \ldots$. We can write the sine function as an infinite sum.

You may think that you will never benefit from this discussion of the sine function, but mathematics is food for the brain. It can teach you how to think precisely, decisively, and creatively, even if you never use a trigonometric function. Math helps you to look at the world in a different way, whether you use it to quantify decisions in daily life or you come to appreciate a fine proof in the same way that other people appreciate great poetry, painting, music, and fine wine. I invite you to join me as we explore the joy of math together. ■

Suggested Reading

Dunham, *The Mathematical Universe: An Alphabetical Journey through the Great Proofs, Problems, and Personalities*.

Gardner, *Aha!: Aha! Insight and Aha! Gotcha*.

———, *Martin Gardner's Mathematical Games*.

Math Horizons magazine.

Paulos, *A Mathematician Reads the Newspaper*.

Weisstein, Wolfram Mathworld, mathworld.wolfram.com.

Questions to Consider

1. When you think of your experience with mathematics, what adjectives come to mind? What experiences were unpleasant? Were there any experiences that you would describe as joyful?

2. Think of all the places where math intersects your life on a daily basis. For instance, where do you encounter math when reading the daily paper?

The Joy of Numbers

Lecture 2

A concept that we take for granted, but one that took thousands of years for people to figure out, is the idea of negative numbers. Imagine trying to convince someone that there are numbers that are less than 0. How can something be less than nothing?

In everyday mathematics, we use the base-10 number system, also called the *Hindu-Arabic system*. Before this system came into use, quantities had to be represented in a one-to-one correspondence. For instance, if you wanted to represent 23 animals, you'd have to line up 23 stones.

In the base-10 system, we think of a number such as 23 as two rows or groups of 10, followed by a group of 3: 10 + 10 + 3 = 23. A larger number, such as 347, is represented as three groups of 100, four groups of 10, and 7 individual units. The number 0 plays an important role as a placeholder in this system. For instance, 106 would be represented as one group of 100 plus 6 individual units, with zero 10s.

In some areas of science and mathematics, other systems are used, such as base 8. For a number such as 12, instead of counting in groups of 10, we count in groups of 8. Thus, the number 12 would be written as 14 (base 8), and 23 would be written as 27 (base 8). If we were counting 12 stones in base 10, we would group the stones in one row of 10, followed by one row of 2. In base 8, we group the stones in one row of 8 and one row of 4.

If we were counting 63 in base 8, we would group the stones in seven rows of 8 and one row of 7; thus, the number 63 would be written as 77 (base 8). What if we were counting 64 stones? Would we group the stones in eight rows of 8? The answer is no, because when we're working in base 8, we have only the digits 0 through 7. Instead, we have to group the stones in one large block of 64, and the number would be written as 100 (base 8). A number such as 347 would have five blocks of 8, plus three rows of 8, plus 3 individual units. This would be written as 533 (base 8).

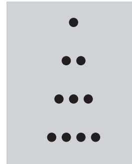
The binary number system, base 2, is used constantly in computers. In this system, we work in powers of 2, and for any number, we write a 1 every time we have a power of 2. The number 12, for example, is $1\times 2^3+1\times 2^2+0\times 2+0\times 1$. Substituting a 1 for each power of 2, we get 1100 (base 2). The number 64 is a power of 2 by itself, with nothing left over. It would be written as a 1 for the 64, but a 0 for the 32, 16, 8, 4, 2, and 1, or 1000000 (base 2).

The hexadecimal system is also used frequently in computers. In this system, instead of having ten digits, 0 through 9, or two digits, 0 and 1, we have sixteen digits. These are the digits 0 through 9, along with A, B, C, D, E, and F, which represent the numbers 10, 11, 12, 13, 14, and 15. In base 10, the number 42 would be represented as four 10s and one 2, but in a hexadecimal system, 42 would be four 16s and one 2, or $4\times 16+2\times 1=66$ (base 16). In the hexadecimal number 2B4, the 2 represents two 16^2s, and the B represents eleven groups of 16, plus 4 units. When you add all those numbers together, you get the number 692 in base 10.

Let's look at the hexadecimal number FADE. This translates to $\mathrm{F}\times 16^3+\mathrm{A}\times 16^2+\mathrm{D}\times 16+\mathrm{E}\times 1$, which when written in terms of base-10 numbers is $15\times 16^3+10\times 16^2+13\times 16+14\times 1=64,222$. Here's a trick question: What would 190 in hexadecimal be? This number would be written as $11\times 16+14$, or $\mathrm{B}\times 16+\mathrm{E}$. The answer is BE, the last word of my question.

Let's turn now to Carl Friedrich Gauss (1777−1855), a great mathematician who seems to have been a genius from a young age. According to one story, when Carl was only 9 or 10 years old, his teacher asked the class to add the numbers 1 through 100. Young Gauss immediately gave the correct answer, 5,050. What Gauss did was to think of the numbers as two groups, 1 through 50 and 51 through 100. He then added those numbers in pairs: 1+ 100 = 101, 2 + 99 = 101, 3 + 98 = 101, … 50 + 51 = 101. The result is fifty 101s, or $50\times 101=5,050$.

We call numbers like 5,050 *triangular numbers*, that is, numbers that can be represented in a triangle. Think of rows of 1 dot, 2 dots, 3 dots, and so on, making a triangle. Using this picture of triangular numbers, we can see another way to come up with a formula for the n^{th} triangular number, that is, another formula for the sum of the first n numbers.



For example, imagine I put together two triangles of the same shape to create a rectangle. I use 1 + 2 + 3 + 4 dots in the form of a triangle and invert another triangle of 1 + 2 + 3 + 4 dots. How many dots are in the rectangle? The rectangle has four rows of 5 dots, which means that if I added 1 + 2 + 3 + 4 twice, I would have 20 dots. If I cut that number in half, I'll have 10 dots for the number that was in a single triangle. If we use one triangle with 1 + 2 + 3 up through n dots and we invert another triangle to create a rectangle that has n rows of $n + 1$ dots, then the n^{th} triangular number plus the n^{th} triangular number is $n(n + 1)$. In other words, the n^{th} triangular number is $\frac{n(n+1)}{2}$.

What is the sum of the first n even numbers, 2 + 4 + 6, all the way up to the number $2n$? Let's reduce this to a problem that we already know how to solve. We can factor out a 2 from each of those terms, which would leave us with 2 times the quantity $(1 + 2 + 3 +\ldots + n)$. We already know that sum is $\frac{n(n+1)}{2}$. Multiplying the sum by 2, the 2s cancel out, and the answer is $n(n + 1)$.

What is the sum of the first n odd numbers? The first odd number is 1, followed by 3, and $1 + 3 = 4$ (or 2^2); the next odd number is 5, and $1+3+5=9$ (or 3^2). Continuing, we start to see a pattern: The sum of the first n odd numbers is n^2. Why is the sum of the first five odd numbers 5^2? If we imagine a square divided into five rows of five squares, and we examine those squares one layer at a time, we can see why the sum of the first n odd numbers is exactly n^2.

Let's look at one more pretty pattern involving odd numbers. We start by adding one odd number, then the next two odd numbers, then the next three odd numbers, and so on: 1, $3 + 5 = 8$, $7 + 9 + 11 = 27$, $13 + 15 + 17 + 19 = 64$, $21 + 23 + 25 + 27 + 29 = 125$. The sums here are all cubes: 1^3, 2^3, 3^3, 4^3, 5^3. We next add these cubes: $1^3 = 1$, $1^3 + 2^3 = 9$, $1^3 + 2^3 + 3^3 = 36$, $1^3 + 2^3 + 3^3 + 4^3 = 100$, $1^3 + 2^3 + 3^3 + 4^3 + 5^3 = 225$. Those sums are perfect squares: 1^2, 3^2, 6^2, 10^2, 15^2, but they're not just any perfect squares—they're the perfect squares of the triangular numbers.

In other words, the sum of the cubes of the first five numbers is equal to the first five numbers summed, then squared. We can also say that the sum of the cubes is the square of the sum. Does this pattern hold for all numbers? As we'll see, the answer is yes.

Let's look now at the patterns in the multiplication table. We'll start by asking a question that you probably haven't thought about since you were in elementary school: Why is 3×5 equal to 5×3? You can draw a picture with dots and count the dots. You might see three rows of five dots in the picture (3×5), or you might see five columns of three dots (5×3). Because both answers are right, they both must represent the same quantity, which is why 3×5 is the same as 5×3.

Why should a negative number multiplied by a negative number equal a positive?

Here's another question you probably haven't thought about in a long time: Why should a negative number multiplied by a negative number equal a positive? The answer is based on the *distributive law* in mathematics, which says: $a(b + c) = (a \times b) + (a \times c)$. Imagine I have a bags of coins, and each bag of coins has b silver coins and c copper coins. Every bag has $(b + c)$ coins, and I have a bags; therefore, the total number of coins is $a(b + c)$. It's also true that I have $(a \times b)$ silver coins (because each bag has b silver coins in it) and $(a \times c)$ copper coins (because each bag has c copper coins). The total number of coins, then, is $(a \times b) + (a \times c)$. Both answers are right, and therefore, they're equal. We can see that the distributive law works when all the numbers are positive.

The distributive law should also work when the numbers are negative. Let's start with an obvious statement: $-3 \times 0 = 0$. We can replace 0 here with $(-5 + 5)$ because that is also equal to 0. If we want the distributive law to work for negative numbers, then it should be true that $(-3 \times -5) + (-3 \times 5) = 0$. We know that $-3 \times 5 = -15$, which leaves us with $(-3 \times -5) - 15 = 0$. We also know that 15 is the only number that results in 0 when we subtract it from 15; that's why $-3 \times -5 = +15$.

Let's look at one last question in this lecture: Can you add up all the numbers in a 10-by-10 multiplication table? You'll need one skill to answer this question: how to square any number that ends in 5. First of all, if you square a number that ends in 5, the answer will always end in 25, such as 35^2, which equals 1,225. To find how the answer begins, multiply the first digit of the original number, in this case 3, by the next higher digit, in this case 4: $3 \times 4 = 12$, so the answer is 1,225.

The original question was: What's the sum of all the numbers in the multiplication table? The numbers in the first row are 1 through 10, the sum of which is 55. That's five pairs of 11, or the triangular number formula: $(10 \times 11)/2 = 55$. The second row of the multiplication table has the numbers 2, 4, 6, 8, 10, and so on, which is twice the sum of the numbers 1, 2, 3, through 10. Thus, that row will add up to 2×55. The third row will add up to 3×55, because it's $3 \times (1 + 2 + 3 + 4 + \ldots + 10)$. The fourth row will add up to 4×55; we can continue to the tenth row, which will add up to 10×55.

If we were to add up all the numbers in the multiplication table, then, we would have $(1 \times 55) + (2 \times 55) + (3 \times 55) + \ldots + (10 \times 55)$. By the distributive law, that's $(1 + 2 + 3 + 4 + 5 + \ldots + 10) \times 55$, and we know the sum of the numbers 1 through 10 equals 55. Thus, if you were to sum all the numbers in the multiplication table, the answer would be 55×55. Returning to the trick we learned earlier, we know that the answer to 55×55 will end in 25 and begin with 5×6, or 30; therefore, the sum of all the numbers in the multiplication table is 3,025. ■

Suggested Reading

Benjamin and Shermer, *Secrets of Mental Math: The Mathemagician's Guide to Lightning Calculation and Amazing Math Tricks*.

Burger and Starbird, *The Heart of Mathematics: An Invitation to Effective Thinking*, chapter 2.

Conway and Guy, *The Book of Numbers*.

Gross and Harris, *The Magic of Numbers*.

Questions to Consider

1. What is the sum of the numbers from 100 to 1,000? Using the formula for the sum of the first n numbers, find a formula for the sum of all the numbers between a and b, where a and b can be any two positive integers.

2. The alternating sum of the first five numbers is $1 - 2 + 3 - 4 + 5 = 3$. Find a formula for the alternating sum of the first n numbers. How about the alternating sum of the squares of the first n numbers?

The Joy of Primes
Lecture 3

The largest prime number that has been found so far, discovered in 2006 by a mathematician named Curtis Cooper, is $2^{32,582,657} - 1$. The resulting prime number is more than 9 million digits long.

Prime numbers, as we'll see, are the building blocks for all the integers around us. We'll restrict our attention in this lecture to positive integers. Let's start by asking a simple question: Which numbers divide evenly into the number 12? The divisors, or *factors*, of 12 are 1, 2, 3, 4, 6, and 12. Similarly, the divisors of the number 30 are 1, 2, 3, 5, 6, 10, 15, and 30. Both of those lists have some divisors in common, namely 1, 2, 3, and 6, and the greatest common divisor (GCD) of those numbers is 6.

Here's a clever way of calculating the GCD of two numbers: We want to find GCD (1,323, 896). Any number that divides evenly into 1,323 and 896 must divide evenly into 896 and into 1,323 − 896. In other words, if a number divides both x and y, then that number must also divide $x - y$. Thus, anything that divides 1,323 and 896 must also divide their difference, 1,323 − 896. How would you do that subtraction in your head? Subtract 900 from 1,323 (= 423), then add 4 back in to get 427. We might also say that any number that divides 896 and 427 must divide their sum, 896 + 427, which is 1,323.

In summary, we've shown that any number that divides the first two numbers, 1,323 and 896, will also divide the next two numbers, 896 and 427. In particular, the greatest of those numbers, the GCD, must be the same. That is, GCD (1,323, 896) = GCD (896, 427). We have now replaced a large number, 1,323, with a smaller number, 427.

This idea goes back to Euclid, an ancient Greek geometer. According to Euclid's algorithm, to find GCD (n, m), divide n by m; the result will be a quotient and a remainder. If $n = qm + r$, then GCD (n, m) will be the same as GCD (m, r).

Let's return to the problem of finding GCD (896, 427). When you divide 427 into 896, you get a quotient of 2 and a remainder of 42, which also means that $896 = 2(427) + 42$. Euclid's algorithm tells us that GCD (896, 427) is the same as GCD (427, 42). Next, 427 divided by 42 is 10 with a remainder of 7, which means that we can simplify GCD (427, 42) to GCD (42, 7). These numbers are now small enough to work with, and we can see that the greatest number that divides evenly into 42 and 7 is 7 itself. Therefore, the GCD of the original two numbers, 1,323 and 896, was shown to be the GCD of the next two numbers, 896 and 427, which was shown to be the GCD of the next two numbers, all the way down to 7, which is the GCD of the first two numbers.

Let's turn now to prime numbers. A number is *prime* if it has exactly two divisors, 1 and itself. A number is *composite* if it has three or more divisors. Both prime and composite numbers must be positive; the number 1 is neither prime nor composite.

The prime numbers are the building blocks of all the integers.

As I said, the prime numbers are the building blocks of all the integers, and this idea is expressed in what's called the *fundamental theorem of arithmetic*, or the *unique factorization theorem*. According to this theorem, every number greater than 1 can be written as the product of primes in exactly one way. As an example, let's look at 5,600. To factor that number, we might say that 56 is 8×7; thus, 5,600 is $8 \times 7 \times 10 \times 10$. That's a factorization, but it's not a prime factorization. The number 8 can be broken into prime factors $2 \times 2 \times 2$; the number 7 is already prime. Both 10s can be factored as 2×5. If we put these together, the prime factorization of 5,600 is $2^5 \times 5^2 \times 7^1$.

How many divisors does 5,600 have? For a number to be a divisor of 5,600, its only prime factors could be 2, 5, and 7. What's the largest power of 2 that could be a divisor of 5,600? The answer is 5, because if we used 2 to a higher power, such as 2^6 or 2^7, then the result wouldn't divide into 5,600. Any divisor of 5,600 will have to be of the form $2^a \times 5^b \times 7^c$, where a could be as small as 0 or as high as 5, b could be as small as 0 or as high as 2, and c could be as small as 0 or as high as 1. If we let $a = 3$, $b = 1$, and $c = 0$, then the divisor would be $2^3 \times 5^1 \times 7^0$, or $8 \times 5 \times 1$, which is 40. What if we chose all 0s? Then, the divisor would be $2^0 \times 5^0 \times 7^0$, or $1 \times 1 \times 1$, which is 1.

The answer to the question of how many divisors 5,600 has will be in the form of $2^a \times 5^b \times 7^c$. We have six possibilities for a, any number between 0 and 5; three possibilities for b; and two possibilities for c. Therefore, 5,600 has $6 \times 3 \times 2$, or 36 divisors.

Now let's look at the complementary idea of least common multiples (LCMs). We look at 12 and 30 again. The number 12 has multiples: 12, 24, 36, 48, 60, 72, … . The number 30 has multiples: 30, 60, 90, 120, 150, 180, … . Comparing those lists, 60 is the smallest multiple of 12 and 30. Recall that GCD (12, 30) = 6, and if we multiply 6×60, we get 360. If we multiply 12×30, we also get 360. That's a consequence of the theorem that for any numbers a and b, GCD (a, b) × LCM (a, b) = ab.

A concept we will use frequently in these lectures is *factorials*. The number $n!$ (n factorial) is defined to be $n \times (n-1) \times (n-2) \times \cdots$, down to 1. For example, 3! is $3 \times 2 \times 1$, which is 6; 4! is $4 \times 3 \times 2 \times 1$, which is 24. Notice that $n!$ can be defined recursively as $n \times (n-1)!$. I claim that 0! is 1 partly because if I want the equation $n! = n \times (n-1)!$ to be true, I need 0! to be 1. Let's look at this idea: $1! = 1 \times 0!$, and $1! = 1$. If we want $1 \times 0!$ to be 1, then 0! must be defined as 1.

The number 10! is 3,628,800, but that pales in comparison to 100!. How many 0s will be at the end of the number 100!? The number 100! has a prime factorization that can be written in the form $2^a \times 3^b \times 5^c \times 7^d \ldots$. To find how many 0s are at the end of that number, we have to ask ourselves how 0s are made.

Every time 2 and 5 are multiplied, the result is a 10, which creates a 0 at the end of a number. For 100!, the only numbers that matter in terms of creating 0s are the power of 2 and the power of 5. The smaller of those numbers, a or c, will be the number of 0s in the result of 100!. Looking at 100!, we see that there will be more powers of 2 in its factorization, so the smaller exponent, the one we're interested in, is the power of 5, the exponent of c. The number of 0s at the end of 100! will be this exponent.

How do we find the number of 5s in the factorization of 100!? There are 20 multiples of 5 in the numbers 1 through 100, and each contributes a factor of 5 to the prime factorization of 100!. Keep in mind that some of those multiples of 5 (namely, 25, 50, 75, and 100) will each contribute an extra factor of 5. Thus, the total contribution of 5s to 100! is 20 + 4, or 24. With 200!, there are 40 multiples of 5, each contributing a 5 to the 200!. All the multiples of 25 each contribute an extra factor of 5, and there are 8 of those up to 200. Finally, the number 125 is 5^3, or $5 \times 5 \times 5$, so it contributes one more factor of 5. Thus, the total number of 0s at the end of 200! will be 49. The number 100! is 9.3×10^{157}, which has 24 zeros on the end.

Now let's return to prime numbers. As we look at larger numbers, the primes become a bit scarcer because there are more numbers beneath them that could possibly divide them. Do the primes ever die out completely? Is there a point after which every number is composite? It seems possible, yet we can prove that the number of primes is, in fact, infinite.

Suppose that there were only a finite number of primes. That would mean that there would have to be some prime number that was bigger than all the other prime numbers. Let's call that number P. Every number, then, would be divisible by 2, 3, 5, …, or P. Now, let's look at the number $P!$. This number, $P!$, will be divisible by 2, 3, 4, 5, 6, 7, …, and every number through P, because it is equal to the product of all those things. Next, let's look at the number $P! + 1$. Can 2 divide evenly into $P! + 1$? No, because 2 divides into $P!$, and if that's true, 2 will not divide into $P! + 1$. Can 3 divide into $P! + 1$? Again, the answer is no, because 3 divides into $P!$, and thus, it can't divide into $P! + 1$. In fact, all the numbers between 2 and P will divide $P!$; therefore, none of them will divide $P! + 1$. That contradicts our assertion that all numbers were divisible by something between 2 and P. Suppose we thought 5 was the biggest prime. We know that the number 5! + 1 will not be divisible by 2, 3, 4, or 5; therefore, 5 could not be the biggest prime. Does that mean 5! + 1 is prime? No. In fact, 5! + 1 = 121, which is 11^2, and 11 is a bigger prime than 5. We know that 5 is not the biggest prime because $P! + 1$ will either be prime or it will be divided by a prime that's larger than P.

Given that there are an infinite number of primes, is it true that we have to encounter a prime every so often? Or would it be possible to find, for example, 99 consecutive composite numbers? I claim that we can. I claim that the 99 consecutive numbers from 100! + 2, 100! + 3, 100! + 4, …, 100! + 100 are all composite. We know that 100! is divisible by 2, 3, 4, …, 100. And we know that since 2 divides into 100!, it will also divide into 100! + 2. Further, since 3 divides into 100!, it will divide into 100! + 3, and so on. Thus, since 100 divides into 100!, it will divide into 100! + 100. Therefore, all those numbers are composite, because the first one is divisible by 2, the second one is divisible by 3, and so on, until the last one, which is divisible by 100.

Let's close with a couple more questions about prime numbers. A *perfect number* is a number that's equal to the sum of all its *proper divisors* (all the divisors except the number itself). For example, 6 has proper divisors 1, 2, and 3. The next perfect numbers are 28, 496, and 8,128. Let's look at the prime factorizations of these numbers.

The prime factorization of 6 is 2 × 3; 28 is 4 × 7; 496 is 16 × 31; and 8,128 is 64 × 127. The first number in all these equations is a power of 2; the second number is one less than twice the original number, and it is also prime. In fact, the mathematician Leonhard Euler (1707−1783) showed that if P is a prime number and if $2^P - 1$ is a prime number, then the result of multiplying $2^{P-1}(2^P - 1)$ will always be perfect. That's true for all the even perfect numbers, but what about odd perfect numbers? No one knows if any odd perfect numbers exist.

A *twin prime* is a set of two prime numbers that differ by 2, for example, 3 and 5. We have found twin primes with more than 50,000 digits, yet we don't know if there are an infinite number of twin primes.

According to *Goldbach's conjecture*, every even number greater than 2 is the sum of two primes: for example, 6 = 3 + 3; 18 = 11 + 7; 1,000 = 997 + 3. This problem has been verified through the zillions, but we don't have proof that it is true for all even numbers. It has been proved, though, that every even number is the sum of at most 300,000 primes. We also have proof that with large enough numbers, we reach a point where every number is of the form $P + QR$, where P is prime and QR is almost prime, meaning that QR is a number that has at most two prime factors, Q and R. Prime numbers have many applications, such as testing the performance, accuracy, and security of computers. ■

Suggested Reading

Gross and Harris, *The Magic of Numbers*, chapters 8−13, 23.

The Prime Pages, primes.utm.edu.

Ribenboim, *The New Book of Prime Number Records*, 3rd ed.

Questions to Consider

1. The primes 3, 5, and 7 form a *prime triplet*, three consecutive odd numbers that are all prime. Why do no other prime triplets exist?

2. To test if a number under 100 is prime, you need to test only whether it is divisible by 2, 3, 5, or 7. In general, to test if a number n is prime, we need to test only if it is divisible by prime numbers less than $\sqrt{n}$. Why is that true?

The Joy of Counting

Lecture 4

The joy of counting ... can really bring you joy, because we're going to see how you can use this to figure out problems that might be of interest to you, such as the number of possible outcomes in a horse race, the chance of winning the lottery, and even figuring out your odds in poker.

Two principles apply to counting: the rule of sum and the rule of product. According to the rule of sum, if I own five long-sleeved shirts and three short-sleeved shirts, then the number of shirts I can wear on any given day is $5 + 3$. According to the rule of product, if I own eight shirts and five pairs of pants, then the number of possible outfits I can wear on any given day is 8×5. If I have ten ties, that would multiply the number of possibilities by a factor of 10. I'd have $8 \times 5 \times 10 = 400$ different outfits.

Knowing those two principles, let's start with a simple question, such as: In how many ways can we arrange a group of letters? For instance, the letters A and B can be arranged in two ways—AB and BA. The letters A, B, and C can be arranged in six ways: ABC, BAC, CAB, ACB, BCA, CBA. There are 24 ways to arrange A, B, C, and D. These numbers are all factorials: $2 = 2!$, $6 = 3!$, and $24 = 4!$.

If we know there are six ways to arrange A, B, and C, let's figure out the number of ways to arrange A, B, C, and D. Starting with ABC, we could put D in the first, second, third, or fourth position. That will to lead to six new ways to arrange ABC and D, where A, B, and C are in their original positions. With the next set of letters, ACB, there are still four places where we can insert the letter D among the original letters. For every one of those six arrangements, we can follow up with four new arrangements. Thus, the number of possibilities is $6 \times 4 = 24$, or $4!$.

Another way of thinking about factorials is to imagine placing five cards on a table in different arrangements. You have five choices for which card you'll put down first. After you've chosen that card, you have four choices for which card goes next. Then, you have three choices for the next card, two choices for the next, and one choice for the last card; the total number of possibilities is $5 \times 4 \times 3 \times 2 \times 1$, or 5!. In general, the number of ways of arranging n different objects is $n!$.

How many different five-digit zip codes are possible? The first digit is anything from 0 to 9; the second digit is from 0 to 9; and so on. For each of the digits, there are 10 choices; therefore, the number of possible zip codes is $10 \times 10 \times 10 \times 10 \times 10 = 10^5 = 100{,}000$. How many zip codes are possible in which none of the numbers repeats? For the first digit, there are 10 choices, but for the second digit, there are only 9 choices; for the third digit, there are 8 choices; for the fourth, 7 choices; and for the fifth, 6 choices. The number of five-digit zip codes with no repeating numbers is $10 \times 9 \times 8 \times 7 \times 6$, or 30,240. Let's apply this approach to horseracing. In a race with 8 horses, how many different outcomes are possible when the outcomes are as follows: one horse finishing first, another finishing second, and another finishing third? Again, there are 8 possibilities for the horse that comes in first, 7 for the horse that comes in second, and 6 for the horse that comes in third: $8 \times 7 \times 6 = 336$ possibilities for the outcome.

How many possible license plates are there if a license plate comes in two varieties? A type I license plate has three letters followed by three numbers. A type II license plate has two letters followed by four numbers. Because we have two types of license plates, the rule of sum will apply here. How many type I license plates are possible? For each of the three letters, there are 26 choices. For each of the three numbers, there are 10 choices. Multiplying those choices, we get 17,576,000 different license plates. How many type II license plates are possible? For the two letters, there are 26 choices each. For the four numbers, there are 10 choices each; altogether, $26 \times 26 \times 10^4 = 6{,}760{,}000$. When we combine type I and type II license plates, the total number of possibilities is 24,336,000.

The branch of mathematics known as combinatorics allows us to solve problems in different ways. For example, we can actually do the license plate problem in one step instead of two. The number of choices for the first letter is 26, and for the second letter, 26 also; whether the license plates are of type I or type II, there are 26×26 ways to get started. The third item on the license plate could be a letter or a number. Thus, there are 26 + 10 = 36 possibilities for the third item. The remaining three items are all numbers; therefore, there are 10 possibilities for each. When we multiple those numbers together, 26 × 26 × 36 × 10 × 10 × 10, we get, again, 24,336,000.

The branch of mathematics known as combinatorics allows us to solve problems in different ways.

What if all the letters must be different on the license plate? In this case, there are 26 choices for the first letter and 25 choices for the second letter, but the third item could be any one of 24 letters or 10 numbers. Therefore, there are 34 possibilities for the third item. Then, because the last three items must all be numbers, there are 10 possibilities for each. Multiply those numbers together, and we get 22,100,000. If all the letters and numbers must be different, then we can still solve the problem in one step, but we have to pursue a more creative strategy. There are 26 choices for the first letter and 25 choices for the second letter. There are 10 choices for the fourth item, which must be a number; 9 choices for the next number; and 8 choices for the last number. The third item can be a number or a letter. There are 24 choices for the letter and 7 choices for the number; therefore, there are 31 possibilities for the third item. Multiplying all those possibilities, 26 × 25 × 31 × 10 × 9 × 8 = 14,508,000.

Let's now talk about winning the lottery. California has a game called Super Lotto Plus, which is played as follows: First, you choose five numbers from 1 through 47. Next, you choose a mega number from 1 through 27. That mega number can be one of the five numbers you picked first, or it can be a different number. For the first step, we pick the first five Fibonacci numbers, 2, 3, 5, 8, and 13, and for the mega number, 21. In how many ways can the state pick its numbers, and which of those are the numbers we picked?

The state has 47 choices for its first number, 46 choices for its second, and so on, or 47 × 46 × 45 × 44 × 43 ways of picking the first five numbers. Then, the state has 27 ways to pick the mega number. It seems like that would be the right answer, but we have overcounted. The state might choose the numbers 1, 10, 20, 30, and 45 or the numbers 10, 20, 45, 30 and 1. They are the same group of five numbers, but we've counted them as different. In how many ways could we arrange those five numbers and still have the same set of five numbers? By dealing the cards earlier, we saw that there were 5! ways of arranging those numbers. Thus, to find the correct answer to this problem, we divide the original number that we came up with by 5!.

In other words, in this problem, we overcounted the possibilities for the numerator, then divided by the denominator to get the correct answer. We saw, then, that the state has 41,416,353 ways to pick its numbers, only one of which is our group of five numbers. Therefore, our chance of winning is just 1/41,416,353. Incidentally, another way to express such products as $47 \times 46 \times 45 \times 44 \times 43$ is to multiply the numerator and the denominator by 42!; thus, the numerator would be 47! and the denominator would be 42!. Those quantities are the same thing, but the second form is cleaner. The number of ways to pick five different numbers out of 47 is $\frac{47!}{5! \times 42!}$.

In general, the number of ways to pick k objects from n objects when the order is not important is $\frac{n!}{k! \times (n-k)!}$. The notation for this is $\binom{n}{k}$. How many 5-card poker hands are possible? We have 52 cards, and we choose 5 of them. The order that you get the cards is not important for a game such as five-card draw. The number of ways of picking 5 out of 52 is $\binom{52}{5}$, which has the formula $\frac{52!}{5! \times 47!}$, which is 2,598,960.

What are the chances of being dealt a specific kind of hand in poker? For instance, what are your chances of being dealt five cards of the same suit, a flush? We have four choices for the suit—spades, hearts, diamonds, or clubs. In how many ways can we pick 5 cards of the same suit, such as hearts, out of the 13 hearts in the deck? By definition, that is $\binom{13}{5}$; thus, we have $4 \times \binom{13}{5} = 4 \frac{13!}{5! \times 8!} = 5{,}148$. The chances of being dealt a flush in poker would be 5,148 divided by the 2,598,960 possible different poker hands. That's about 0.2 percent; about 1 out of every 500 poker hands dealt will be a flush.

What are the chances of being dealt a full house in poker? A full house consists of five cards, three of one value and two of another value. There are 13 choices for the value that will be triplicated and 12 choices for the value that will be duplicated. Let's say our two values are queens and sevens. Next, we have to determine the possibilities for suits of those cards. How many possibilities for suits are there for the 3 queens? The answer is $\binom{4}{3}$. That is, from the 4 queens in the deck—spade, heart, diamond, and club—choose 3 of them: $\binom{4}{3} = 4$. Similarly, how many possibilities for suits are there for the 2 sevens? The answer is $\binom{4}{2} = 6$. Thus, the number of possibilities for a full house is $13 \times 12 \times 4 \times 6 = 3{,}744$.

How many 5-card poker hands have at least 1 ace? To answer this question, you might reason that you first have to choose an ace, then choose 4 other cards from the remaining 51. You have 4 choices for the first ace and $\binom{51}{4}$ ways of picking from the remaining 51 cards. The answer, then, would be $4 \times \binom{51}{4}$.

Unfortunately, that logic is incorrect. There is no "first ace" in the poker hand. To approach the problem by choosing an ace as the first card, then picking 4 other cards is to bring order into a problem where order does not belong. The correct way to do the problem is to break it down into four cases. First, we count those poker hands with 1 ace; then, we count those hands with 2 aces; then, 3 aces; then, 4 aces. Then, we apply the rule of sum to add those hands together, as shown below.

Number of poker hands with 1 ace:	$\binom{4}{1} \times \binom{48}{4}$ *	Number of poker hands with 3 aces:	$\binom{4}{3} \times \binom{48}{2}$
Number of poker hands with 2 aces:	$\binom{4}{2} \times \binom{48}{3}$	Number of poker hands with 4 aces:	$\binom{4}{4} \times \binom{48}{1}$

*Number of ways to choose 1 ace out of 4 in the deck multiplied by the number of ways to choose 4 non-aces out of 48 in the deck.

Adding the cases together, we get 886,656 different poker hands.

Another approach to this problem is to find how many hands have no aces and subtract that answer from the total amount. There are 4 aces in the deck and 48 non-aces, and we can choose any 5 of the non-aces. In how many ways can we choose 5 things out of 48? By definition, the answer is $\binom{48}{5}$, which is 1,712,304. Once we have that value, we can subtract it from the number of possible poker hands, 2,598,960, which leaves us the same number we got before, 886,656.

The possibilities for counting questions in horseracing, lotteries, and poker are endless, as endless as the variations of the games themselves. What happens if we allow wild cards in the game? What if you're playing seven-card stud or Texas hold 'em or blackjack? You can apply mathematics to solving problems in all these games, but you don't want to use math to take all the fun out of games! ■

Suggested Reading

Benjamin and Quinn, *Proofs That Really Count: The Art of Combinatorial Proof*.

Gross and Harris, *The Magic of Numbers*, chapters 1−4.

Tucker, *Applied Combinatorics*, 5th ed.

Questions to Consider

1. How many five-digit zip codes are palindromic (that is, read the same way backward as forward)?

2. In how many ways can you be dealt a straight in poker (that is, five cards with consecutive values: A2345 or 23456 or ... or 10JQKA)? In how many ways can you be dealt a flush (that is, five cards of the same suit)? Compare these numbers to the number of full houses. This explains why in poker, full houses beat flushes, which beat straights.

The Joy of Fibonacci Numbers

Lecture 5

The Fibonacci numbers appear in nature. If you study pineapples, sunflowers, they actually show up there—in computer science, in arts, in crafts, and even in poetry.

In this lecture, we'll talk about the *Fibonacci numbers*. This sequence begins with the numbers 1, 1, 2, 3, 5, 8, 13, 21, and so on. We can find the sequence by adding each number to the number that precedes it. In the 12th century, Fibonacci wrote a book called *Liber Abaci* (*The Book of Calculation)* that was the first textbook for arithmetic in the Western world and used the Hindu-Arabic system of numbers.

The Fibonacci numbers arose in one of the problems from this book that involved a scenario with imaginary rabbits that never die. We begin with one pair of rabbits in month 1. After one month, the rabbits are mature, they mate, and they produce a pair of offspring, one male and one female (month 3). After one month, those offspring mature, mate, and produce a pair of offspring, giving us two pairs of adults and one pair of babies (month 4). In month 5, we'll have three pairs of adults and two pairs of babies.

How many pairs will we have in month 6? We will have all five pairs of rabbits from month 5, plus all the rabbits from month 4 will now have babies, or 5 + 3 = 8. How many rabbits will we have after 12 months? By continuing this process, you can see that we will have 144 pairs of rabbits in month 12.

Let's look at the Fibonacci numbers from a more mathematical standpoint. We define F_1 to be the first Fibonacci number and $F_2 = 1$ to be the second Fibonacci number. We then have what's called a *recursive equation* to find the other Fibonacci numbers. According to this equation, the n^{th} Fibonacci number $(F_n) = F_{n-1} + F_{n-2}$.

F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
1	1	2	3	5	8	13	21	34	55	89	144

What would happen if we were to start adding all the Fibonacci numbers together? For instance, we would see the following:

$$1 + 1 = 2$$
$$1 + 1 + 2 = 4$$
$$1 + 1 + 2 + 3 = 7$$
$$1 + 1 + 2 + 3 + 5 = 12$$
$$1 + 1 + 2 + 3 + 5 + 8 = 20 \ldots .$$

Do you see a pattern with those numbers—1, 2, 4, 7, 12, 20? When those numbers are written as differences (2 − 1, 3 − 1, 5 − 1, 8 − 1, 13 − 1, 21 − 1), we see that they are each one number off from the Fibonacci numbers—2, 3, 5, 8. We'll look at two explanations for why this works.

The first explanation is that if the formula works in the beginning, it will keep on working. We know, for example, that 1 + 1 + 2 + 3 + 5 + 8 = 21 − 1. What will happen when we add the next Fibonacci number, 13, to that system? When we add 21 + 13, we get the next Fibonacci number, 34; further, 21 − 1 + 13 = 34 − 1, and that pattern will continue forever. This is our first example of what's called a *proof by induction*.

The second explanation is a bit more direct. Let's replace the first 1 in the sequence 1 + 1 + 2 + 3 + 5 + 8 = 21 − 1 with 2 − 1. Let's then replace the second 1 with 3 − 2; the 2 with 5 − 3; and so on. We're representing each of those numbers as the difference of two Fibonacci numbers: (2 − 1) + (3 − 2) + (5 − 3) + (8 − 5) + (13 − 8) + (21 − 13).

Look at what happens when we add those numbers together. Starting with (2 − 1) + (3 − 2), we get a +2 and a −2, and those 2s cancel. Then, when we add 5 − 3, the 3s cancel; when we add 8 − 5, the 5s cancel; and so on. This is called a *telescoping sum*. When the dust settles, all that's left of this sum is the 21 on the right that hasn't been canceled yet and the −1 at the beginning that never got canceled. Thus, when we add all those numbers together, we get 21 − 1. The formal equation, what mathematicians call an *identity*, for $F_1 + F_2 + \cdots + F_n$ is $F_{n+2} - 1$. That is, the sum of the first n Fibonacci numbers is equal to $F_{n+2} - 1$.

What would happen if we were to sum the first n even-positioned Fibonacci numbers? That is, what's $F_2 + F_4 + F_6 + \cdots + F_{2n}$? Let's begin by looking at the data: F_2 is 1, F_4 is 3, F_6 is 8, and as we add these numbers up, we have 1, 1 + 3 = 4, 1 + 3 + 8 = 12, and 1 + 3 + 8 + 21 = 33. Do you see the pattern? Rewriting those numbers, we have 2 − 1, 5 − 1, 13 − 1, 34 − 1; those differences are 2, 5, 13, 34—every other Fibonacci number. Thus, the pattern is $F_3 - 1$, $F_5 - 1$, $F_7 - 1$, and $F_9 - 1$.

Let's see why that works. Look at the equation: 1 + 3 + 8 + 21. We leave the 1 alone, but we replace 3 with 1 + 2; we replace 8 with 3 + 5; and we replace 21 with 8 + 13: (1) + (1 + 2) + (3 + 5) + (8 + 13) = 34 − 1. We're adding every other Fibonacci number, and what we really have is 1 + 1 + 2 + 3 + 5 + 8 + 13, which is exactly the same pattern that we had before. The result, then, is 34 − 1, just as we saw before.

What would happen if we were to sum the odd-positioned Fibonacci numbers? That is, what's $F_1 + F_3 + F_5 + \cdots + F_{2n-1}$? We start with 1, then 1 + 2, then 1 + 2 + 5, then 1 + 2 + 5 + 13. We see the numbers 1, 3, 8, and 21. Those are the Fibonacci numbers themselves, not disguised at all. Why does that work? As before, we leave the 1 alone, but we replace 2 with 1 + 1, 5 with 2 + 3, and 13 with 5 + 8. When we add all those together, we have the same Fibonacci sum, except we have an extra 1 at the beginning; that extra 1 will cancel the −1, leaving us with an answer of 21. The formula is $F_1 + F_3 + F_5 + \cdots + F_{2n-1} = F_{2n}$.

Let's now look at a different pattern. Which Fibonacci numbers are even? According to the data, every third Fibonacci number appears to be even. Will this pattern continue? Think about the fact that the Fibonacci numbers start off as odd, odd, even. When we add an odd number to an even number, we get an odd number. Then, when we add the even number to the next odd number, we get another odd number. When we add that odd number to the next odd number, we get an even number, and we're back to where we started: odd, odd, even. That proves that every third Fibonacci number will be even. What's more, anything that isn't a third Fibonacci number won't be even; it will be odd.

What if we look at every fourth Fibonacci number? Believe it or not, every fourth Fibonacci number is a multiple of 3: 3, 21, 144. Moreover, the only multiples of 3 among the Fibonacci numbers occur as every fourth Fibonacci number. Every fifth Fibonacci number is a multiple of 5. Every sixth Fibonacci number is a multiple of 8, and the only multiples of 8 are F_6, F_{12}, F_{18}, F_{24}, ... This theorem reads: The number F_m divides F_n if and only if m divides n.

Forgetting about Fibonacci numbers for just one second, what is the largest number that divides 70 and 90? In other words, what is the greatest common divisor (GCD) of 70 and 90? The answer is 10. Now, what is the largest number that divides F_{70} and F_{90}, the 70th Fibonacci number and the 90th Fibonacci number? Believe it or not, the answer is the 10th Fibonacci number. In general, GCD (F_m, F_n) is always a Fibonacci number, and it's not just any Fibonacci number, but it's the most poetic Fibonacci number you could ask for. That is to say, GCD $(F_m, F_n) = F_{\text{GCD}\,(m,n)}$.

Which Fibonacci numbers are prime? Looking at our list of Fibonacci numbers, we have 2, 3, 5, 13, and 89. It turns out that the first few prime Fibonacci numbers are $F_3, F_4, F_5, F_7, F_{11}, F_{13}, F_{17}, \ldots$. There's a pattern there; except for F_4, which we'll ignore, it looks like we're seeing prime indices. In fact, if the index is composite (except for F_4, which is a special case because it's 2×2, and F_2 and F_1 are both 1), if m is composite, then F_m is guaranteed to be composite. That's a consequence of the theorem that states that m divides n if and only if F_m divides F_n.

F_3	F_4	F_5	F_7	F_{11}	F_{13}	F_{17}
2	3	5	13	89	233	1597

Is it true that every prime index produces a prime Fibonacci number? As is often the case with prime numbers, the answer to that question is hard to pin down. If we go just a little farther out in the sequence to F_{19}, we see that 19 is prime, but F_{19} is 4,181, which is not prime; it can be factored into 113 × 37. The only places where we see primes along the Fibonacci trail are at the Fibonacci indices. In fact, an unsolved problem in math is, Are there infinitely many prime Fibonacci numbers? Even though we don't know if there are an infinite number of prime Fibonacci numbers, we do know that every prime divides a Fibonacci number. In fact, if P ends in 1 or 9, then P divides F_{P-1}. If P ends in 3 or 7, then P divides F_{P+1}. For instance, 7 divides F_8, 21; and 11, which ends in 1, divides F_{10}, which is 55. Then, 13, which ends in 3, divides F_{14}, which is 377, which is 13×29.

We know that if we add consecutive Fibonacci numbers together, we get the next Fibonacci number; that's how Fibonacci numbers are made. Let's now look at the squares of Fibonacci numbers. Starting off, $1^2 = 1$, $2^2 = 4$, $3^2 = 9$, $5^2 = 25$, $8^2 = 64$, $13^2 = 169$, $21^2 = 441$, and so on. Look what happens if we add $1^2 + 1^2$; we get 2, a Fibonacci number. If we add $1^2 + 2^2$, we get 5, a Fibonacci number. If we add $2^2 + 3^2$, $4 + 9$, we get 13, another Fibonacci number. In fact, it looks as if the sum of the squares of Fibonacci numbers is always a Fibonacci number. That is to say, $F_n^2 + (F_{n+1})^2 = F_{2n+1}$.

An unsolved problem in math is, Are there infinitely many prime Fibonacci numbers?

What happens if we start adding up the sums of the squares, not of two consecutive Fibonacci numbers, but of all the Fibonacci numbers? We begin with $1^2 + 1^2 = 2$, $1^2 + 1^2 + 2^2 = 6$, $1^2 + 1^2 + 2^2 + 3^2 = 15$; the sum of the squares of the first five Fibonacci numbers is 40. The sum of the squares of the first six Fibonacci numbers is 104. If we look closely at the results of these additions—2, 6, 15, 40, 104—we see that the Fibonacci numbers are buried inside them. For example, 2 is 1×2, 6 is 2×3, 15 is 3×5, 40 is 5×8, and 104 is 8×13. In fact, in general, $F_1^2 + F_2^2 + F_3^2 + \cdots + F_n^2 = F_n \times F_{n+1}$.

Let's focus on one of the Fibonacci numbers, say, F_4, which is 3. If we multiply its neighbors, F_3 and F_5, we see that 2×5 is 10, which is 1 away from 9, or F_3^2. If look at F_5, which is 5, and multiply its neighbors, 3×8, the result is 24, or 1 away from 25. Do you see the pattern? This pattern works even with the lower numbers; $F_1 \times F_3$ is 1 away from 1^2, and $F_2 \times F_4$ is 1 away from F_3^2. In general, the pattern seems to be as follows: $F_{n-1} \times F_{n+1} = F_n^2 \pm 1$. In fact, we can say it more precisely: $F_{n-1} \times F_{n+1} - F_n^2 = (-1)^n$.

What if we look at the neighbors that are two away from a given Fibonacci number? We begin with F_3, which is 2, and multiply its neighbors two to the left and two to the right. We get $1 \times 5 = 5$, which is 1 away from 4, or 2^2. Let's try the same thing with F_4, which is 3. We multiply its two-away neighbors, 1×8, which is 8, or 1 away from 9, or 3^2. The general pattern is $F_{n-2} \times F_{n+2} - F_n^{\,2} = (-1)^n$.

If we look three away from a given Fibonacci number, we see the same sort of pattern. Looking at F_5, 5, and multiplying its three-away neighbors, we get $1 \times 21 = 21$, which is 4 away from 25. Looking at F_6, 8, and multiplying its three-away neighbors, we get $2 \times 34 = 68$, which is 4 away from 64. The differences between these neighboring multiplications are 1, 1, 4, 9, 25, 64—squares of the Fibonacci numbers.

$$\frac{1}{1} = 1 \qquad \frac{8}{5} = 1.6$$

$$\frac{2}{1} = 2 \qquad \frac{13}{8} = 1.625$$

$$\frac{3}{2} = 1.5 \qquad \frac{21}{13} = 1.615\ldots$$

$$\frac{5}{3} = 1.666\ldots$$

golden ratio: $\dfrac{1+\sqrt{5}}{2} = 1.618\ldots.$

Let's now turn to some division properties of Fibonacci numbers. The ratios of consecutive Fibonacci numbers (shown at right) seem to converge on what's known as the *golden ratio*.

Let's look briefly at the properties of the golden ratio. We start with a rectangle of dimensions 1 and 1.618… and cut out a 1-by-1 square, leaving a rectangle with height of 1 and length of .618… . Rotating the second rectangle 90 degrees, we have a rectangle that is proportional to the first, with height of .618… and length of 1. Thus, the ratio of $\dfrac{1.618...}{1}$ is the same as the ratio of $\dfrac{1}{.618...}$.

Here's another connection between the golden ratio and the Fibonacci numbers, known as *Binet's formula*. Amazingly, this formula, shown below, produces the Fibonacci numbers, and it can be used to explain many of the Fibonacci numbers' beautiful properties. ■

Binet's Formula

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

Suggested Reading

Benjamin and Quinn, *Proofs That Really Count: The Art of Combinatorial Proof*.

Fibonacci Association, www.mscs.dal.ca/Fibonacci.

Knott, *Fibonacci Numbers and the Golden Section*, www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fib.html.

Koshy, *Fibonacci and Lucas Numbers with Applications*.

Livio, *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number*.

Questions to Consider

1. Investigate what you get when you sum every third Fibonacci number. How about every fourth Fibonacci number?

2. Close cousins of the Fibonacci numbers are the Lucas numbers: 2, 1, 3, 4, 7, 11, 18, 29, 47, 76, 123, ... What patterns can you find inside this sequence?

L_0	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	L_{11}	L_{12}
2	1	3	4	7	11	18	29	49	76	123	199	322

For instance, what do you get when you add Lucas numbers that are two apart? What is the sum of the first *n* Lucas numbers? How about the sum of the squares of two consecutive Lucas numbers? What happens to the ratio of two consecutive Lucas numbers?

The Joy of Algebra

Lecture 6

Algebra was invented by an Arab mathematician named Al-Khowarizmi around 825. ... He wrote a book ... *Hisâb al-jabr w'al muqâbalah*, which literally meant the science of reunion and the opposition. Later on, it was interpreted as the science of transposition and cancellation. ... *Al-jabr* is where we get the term *algebra.* ... Later on, computer scientists used the word *algorithm* as any formal procedure of calculating in a particular way. That was named in honor of Al-Khowarizmi.

We begin this lecture by exploring the magic trick we started with. Algebra assigns variables to unknown quantities. When I asked you to think of a number between 1 and 10, I called that unknown number n. The next step was to double that number: $n + n$, or $2n$. The next step was to add 10: $2n + 10$. Then, divide by 2: $\frac{2n+10}{2} = n+5$. Finally, subtract the original number: $n + 5 - n = 5$.

Let's do another trick. This time, think of two numbers between 1 and 20. Let's say you chose the numbers 9 and 2. We'll then start adding these two consecutive numbers to get the next number in a sequence, as shown in the table on the left below.

1	9
2	2
3	9 + 2 = 11
4	2 + 11 = 13
5	11 + 13 = 24
6	13 + 24 = 37
7	24 + 37 = 61
8	37 + 61 = 98
9	61 + 98 = 159
10	98 + 159 = 257

1	x
2	y
3	$x + y$
4	$x + 2y$
5	$2x + 3y$
6	$3x + 5y$
7	$5x + 8y$
8	$8x + 13y$
9	$13x + 21y$
10	$21x + 34y$

Before we continue, note that the sum of the numbers in rows 1 through 10 is 671. The next step is to divide the number in row 10 by the number in row 9. With any two starting numbers, you will find that the first three digits of the answer will always be 1.61. In fact, if you were to continue this process to 20 lines or more and divide the 20^{th} number by the 19^{th} number, you would find ratios getting closer and closer to 1.618…, the golden ratio.

In this trick, we're dealing with two unknown quantities, so let's call the first two numbers chosen x and y. Now, our sequence of additions looks like the table on the right above. The coefficients in each equation are Fibonacci numbers. The sum of lines 1 through 10 is $55x + 88y = 11(5x + 8y)$. Interestingly, the equation $5x + 8y$ is the same as the equation in line 7. To find the sum of lines 1−10, then, I simply multiplied the result in line 7 by 11: $11 \times 61 = 671$.

You can easily multiply any two-digit number by 11 as follows: Using 11×61 as an example, add 6 + 1 and insert the answer, 7, between the 6 and the 1: 671 is the answer. What happens if the numbers add up to something greater than 9? Try 11×85: $8 + 5 = 13$; we insert the 3 in the middle, then carry the 1 to the 8 to get the answer 935.

Why this method works is easy to see if we look at how we would normally multiply 61×11 on paper.

$$\begin{array}{r} 61 \\ \times\ 11 \\ \hline 61 \\ +\,610 \\ \hline 671 \end{array}$$

We see that $1 \times 61 = 61$, and $10 \times 61 = 610$; when we add these two results, we get a 6 on the left and a 1 on the right and, in the middle, 6 + 1.

Returning to the magic trick, we saw how to obtain the sum of lines 1 through 10, but how did we get 1.61 when we divided line 10 by line 9? The answer is based on adding fractions badly: If you didn't know how to add fractions correctly, you might add the numerators together and the denominators together; thus, 1/3 + 2/5 = 3/8. Of course, this answer isn't correct, but it is true that when you add fractions in this way, the answer you get will lie somewhere in between the two original fractions. In general, if we add the numerators and the denominators for $a/b < c/d$, then the resulting fraction (called the *mediant* of those two numbers), $(a + c)/(b + d)$, will lie in between.

The number in line 10 of our magic trick is $21x + 34y$. The number in line 9 is $13x + 21y$. We're interested in that fraction: $(21x + 34y)/(13x + 21y)$. This is the mediant, the "bad fraction" sum, of $21x/13x + 34y/21y$. In the fraction $21x/13x$, the x's cancel, leaving us with 21/13 on the left, which is 1.615… . On the right, we have $34y/21y$, which reduces to 34/21, or 1.619… . As long as the numerator and denominator are positive, then the mediant is guaranteed to lie in between 1.615 and 1.619. As you recall, I asked only for the first three digits of the answer, which is how I knew it was 1.61.

Let's turn now to one of those word problems we all dreaded in school: Find a number such that adding 5 to it has the same effect as tripling it. We don't know the number yet, so let's call it x. If we triple x, we get $3x$, and we want $3x$ to be the same as $x + 5$, or $3x = x + 5$. First, we want to clean up this equation, but we have to keep in mind the golden rule of algebra: Do unto one side as you would do unto the other. Thus, we have to subtract x from both sides: $3x - x = x + 5 - x$. The left side is $3x - x = 2x$, and the right side is $x + 5 - x = 5$. Now we have a much simpler equation: $2x = 5$. We now need to divide both sides by 2. This leaves us with x on the left side and 5/2, or 2.5, on the right. Let's verify the answer: If we triple 2.5, we get 7.5, and if we add 5 to 2.5, we also get 7.5.

Here's another word problem: Find a number such that doubling it, adding 10, then tripling it will yield 90. Again, let's call the original number x. The first step is to double this number and add 10: $2x + 10$. Then, we have to triple that quantity, and it should equal 90: $3(2x + 10) = 90$. We simplify that equation by dividing both sides by 3. On the left, we then have $2x + 10$. On the right, we have 30. Next, we subtract 10 from both sides; the result is $2x = 20$. Of course, now we divide by 2, and we're left with $x = 10$.

Here's another word problem: Today, my daughter Laurel is twice as old as my daughter Ariel. Two years ago, Laurel was three times as old as Ariel. The question is: How old are they today? Here, we have two unknowns, Laurel's age today and Ariel's age today. We'll call those unknowns L and A, respectively.

We know that today Laurel is twice as old as Ariel: $L = 2A$. We also know that two years ago, $L - 2$, Laurel was three times as old as Ariel, $A - 2$. This sentence translates into the equation $L - 2 = 3(A - 2)$. The right side of the equation, $3(A - 2)$, becomes $3A - 6$; thus, $L - 2 = 3A - 6$. Let's now substitute what we learned from the first equation: $L = 2A$. Wherever we see L, we can replace that term with $2A$. The left side of the second equation, then, reads $2A - 2$; the right side still reads $3A - 6$: $2A - 2 = 3A - 6$. Now, we can simplify by adding 6 to both sides to eliminate the 6 on the right: $2A - 2 + 6 = 3A - 6 + 6$, or $2A + 4 = 3A$. Subtracting $2A$ from both sides leaves us with $4 = A$. Therefore, Ariel is 4, and Laurel, who is twice Ariel's age today, is 8.

The way we would do multiplication on paper is really nothing more than an application of FOIL, the distributive law.

The last technique we'll learn in this lecture is FOIL, which we use when we're multiplying several variables together. Suppose we want to multiply the quantity $(a + b)$ by the quantity $(c + d)$. We can write the equation with the answer as follows: $(a + b)(c + d) = ac + ad + bc + bd$. When we multiply the first numbers in the two sets of parentheses, we get ac. When we multiply the outer numbers in the two sets of parentheses, we get ad. We get bc when we multiply the inner numbers and bd when we multiply the last numbers. The name FOIL comes from this technique of multiplying first, outer, inner, last.

FOIL is nothing more than the *distributive law*. According to this law, we can look at $(a + b)(c + d)$ as $a(c + d) + b(c + d)$. By the distributive law again, we can look at $a(c + d)$ as $ac + ad$ and $b(c + d)$ as $bc + bd$. If we put that all together, we get $ac + ad + bc + bd$, which is FOIL.

Let's do an example to solidify that concept: 13×22. The way we would do multiplication on paper is really nothing more than an application of FOIL, the distributive law. That is, 13 is (10 + 3), and 22 is (20 + 2). Multiplying these two quantities together, we get 10×20 for the first term, 10×2 for the outer, 3×20 for the inner, and 3×2 for the last. The result is $200 + 20 + 60 + 6 = 286$.

Let's do a few other examples. First, let's look at $(x + 3)(x + 4)$. The FOIL results for this example would be: x^2, $4x$, $3x$, and 12. Adding those together, we get $x^2 + 4x + 3x + 12$, or $x^2 + 7x + 12$. Now, let's look at $(x + 6)(x - 1)$. Think of $x - 1$ as $x + (-1)$. The FOIL results for this example would be: x^2, $-x$, $6x$, and -6. Adding those together (*combining like terms*) gives us $x^2 + 5x - 6$. Finally, let's look at $(x + 3)(x - 3)$. The FOIL results would be x^2, $-3x$, $+3x$, and -9. The $-3x$ and $+3x$ cancel, leaving us with $x^2 - 9$. You can see, going through the same kinds of calculation, that if we multiply $(x + y)(x - y)$, we get the expression $x^2 - y^2$. We'll see applications of that equation in our next lecture. ■

Suggested Reading

Barnett and Schmidt, *Schaum's Outline of Elementary Algebra*, 3rd ed.

Gelfand and Shen, *Algebra*.

Selby and Slavin, *Practical Algebra: A Self-Teaching Guide*, 2nd ed.

Questions to Consider

1. Pick any two different one-digit numbers and a decimal point. These numbers can be arranged in six different ways. (For instance, if you choose the numbers 2 and 5, you can obtain 52, 25, 5.2, 2.5, .52, and .25.) Next, perform the following steps: Add the six numbers together, multiply that sum by 100, divide by 11, divide by 3, and finally, divide by the sum of the original two numbers you selected. Your answer should be 37. Why?

2. Choose any three-digit number in which the numbers are in decreasing order (such as 852 or 931). Reverse the numbers, and subtract the smaller number from the larger. (Example: 852 − 258 = 594.) Now, reverse the new number you just got and add the two numbers together. (Example: 594 + 495 = 1,089.) Use algebra to show that your final answer will always be 1,089.

The Joy of Higher Algebra

Lecture 7

Then the search went on for hundreds of years to try and find a solution to the quintic equation. That's an equation of fifth degree. Many prominent mathematicians attempted to solve this problem. Lagrange ... Euler ... Descartes, Newton—all of them tried to find a formula, even a messy one, for the quintic equation. They all failed. It wasn't until a Norweigan mathematician named Abel, Niels Abel, actually showed that, in fact, the solution to the quintic equation was futile. ... Which leads me to the following riddle: Why did Isaac Newton not prove that solving the quintic was impossible? The answer is: He wasn't Abel.

Let's begin by looking at the equation we saw at the end of Lecture 6, namely, $(x + y)(x - y) = x^2 - y^2$. This equation can help you learn to square numbers in your head faster than you ever thought possible.

In Lecture 2, we learned how to square numbers that end in 5. For example, to square 65, we know that the answer will end in 25 and begin with the product of 6 and 7, which is 42. The answer is 4,225. Let's start off with an easy number, 13. The number 10 is close to 13, and it's easier to multiply. We'll substitute 10, then, for 13, but we have to keep in mind that if we go down 3 to 10, we must go the same distance up, which gives us the number 16. Instead of multiplying 13×13, we'll multiply 10×16. The result is, of course, 160, and to that, we add the square of the number 3, the distance we went up and down; 3^2 is 9, and $160 + 9 = 169$, or 13^2.

If we do a problem that ends in 5, such as 35^2, we can see why the answer turns out to be the same as it did with our earlier trick. The nearest easy number could be 30 or 40; we go down 5 to 30 and up 5 to 40: $30 \times 40 = 1{,}200$. Now we add 5^2, which is 25, to get 1,225.

Let's try one final example, 99^2. We'll go up 1 to 100 and down 1 to 98: $98 \times 100 = 9{,}800$. To that, we add 1^2, or 1, which means that the answer is 9,801. The reason this trick works is all based on algebra. Let's start with the equation $x^2 = x^2 - y^2 + y^2$. The $-y^2$ and $+y^2$ cancel, leaving us with x^2. As we saw at the end of the last lecture, $x^2 - y^2$ is equal to $(x + y)(x - y)$; that means, then, that x^2 is equal to $(x + y)(x - y) + y^2$. For clarity, let's substitute in the number we just used in the last example, 99^2. If we let $y = 1$, then we have $99^2 = (99 + 1)(99 - 1) + 1^2$; that simplifies to $(100 \times 98) + 1$, which gives us 9,801.

Let's turn to a trick that is even more magical, and it's based on similar algebra. This works for multiplying two numbers that are close together. We'll start with 106×109. The first number, 106, is 6 away from 100; the second number, 109, is 9 away from 100. Now, we add 106 + 9 or 109 + 6, which is 115. Next, we multiply 115 by our easy number, 100: $115 \times 100 = 11{,}500$. Then, multiply 6×9 and add that result to 11,500 for a total of 11,554.

This equation can help you learn to square numbers in your head faster than you ever thought possible.

Again, this trick works through algebra. Let's suppose we're multiplying two numbers, $(z + a)$ and $(z + b)$. Think of z as a number that has lots of zeros in it. If we multiply the numbers using FOIL, we get $(z + a)(z + b) = z^2 + za + zb + ab$. Notice that the first three terms have z's in them, so we can factor out a z to get $z(z + a + b) + ab$. Let's substitute numbers, say, 107×111: $(100 + 7)(100 + 111)$. When z is 100, the answer will be $100(100 + 7 + 11) + (7 \times 11)$. To solve that, we first get 11,800; then we add 7×11 for a total of 11,877.

Let's do another example: 94×91. These numbers are both close to 100, but they're less than 100. The number 94 is −6 away from 100, and the number 91 is −9 away from 100. We now subtract 94 − 9 or 91− 6 to get 85. We multiply 85×100 to get 8,500. To that answer we add $(-6)(-9)$, which is +54. Our answer, then, is 8,500 + 54 = 8,554.

What if one of the numbers is above 100 and one of them is below 100? Let's try 97×106. The number 97 is 3 below 100, and 106 is 6 above 100. We start by adding 97 + 6 = 103. We then multiply $103 \times 100 = 10{,}300$. To that number, we add $(-3)(6)$, which is to say that we subtract 18 from 10,300: 10,300 − 18 = 10,282.

Let's try a simpler problem: 14×17. The nearest easy number to 14 and 17 is 10. Here we multiply 10×21, which is 210. Now we add 4×7, because we were 4 away from 10 and 7 away from 10: $210 + 28 = 238$.

Let's do one more of these problems: 23×28. Those numbers add up to 51, so we multiply $20 \times 31 = 620$. To that, we add 3×8, which gives us an answer of 644. This trick is especially magical when the two numbers at the end add up to 10 because then the multiplication becomes so easy you almost don't have to keep track of the zeros. With 62×68, we multiply 60×70 to get 4,200, to which we add 2×8 to get an answer of 4,216. We're using the same trick that we did for squaring numbers that end in 5. Try 65^2. We're 5 away from 60, and $65 + 5 = 70$. We multiply 60×70, which is 4,200, and add 5^2 to get 4,225.

Now we'll move on to solving quadratic equations. But before we work on quadratic equations, let's have a quick refresher on solving linear equations. For the equation $9x - 7 = 47$, we first add 7 to both sides, which leaves us with $9x = 54$. Dividing both sides by 9 gives us $x = 6$. Now let's try $5x + 11 = 2x + 18$. Subtract 11 from both sides and subtract $2x$ from both sides, leaving $3x = 7$. Solving, we get $x = 7/3$. We should also verify these solutions. In the first equation, we plug in 6 for x, then check to make sure that $9x - 7 = 47$. In the second equation, we plug in 7/3 for x: $5x + 11$ would be $35/3 + 11$, which is 68/3; $2x + 18$ would be $14/3 + 18$, and $14/3 + 54/3$ is also 68/3.

Let's now solve a quadratic equation, $x^2 + 6x + 8 = 0$, using a technique called *completing the square*. Look at $x^2 + 6x + 8$; notice that $(x + 3)(x + 3) = x^2 + 6x + 9$, which would be a perfect square. We can turn our equation into that perfect square by adding 1 to both sides. When we add 1 to both sides, we get $x^2 + 6x + 9 = 1$. We see that $x^2 + 6x + 9$ is the quantity $(x + 3)^2$, which means that quantity is 1. There are only two numbers that yield 1 when squared: 1 and −1. Thus, it must be the case that $x + 3 = 1$ or −1. If $x + 3 = 1$, that means that $x = -2$. If $x + 3 = -1$, that means that $x = -4$. If we plug in $x = -2$, we get $-2^2 + 6(-2) + 8 = 0$, which is true. If we plug in $x = -4$, we get $(-4)^2 + 6(-4) + 8 = 0$, which is also true.

Essentially, using the same logic, we can derive the *quadratic formula*. According to the quadratic formula, any equation of the form $ax^2 + bx + c = 0$ has the following solutions: $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

Let's look at this in terms of the last equation we did: $x^2 + 6x + 8 = 0$. The coefficient behind the x^2, that's a, is equal to 1; the coefficient behind the x, that's b, is equal to 6; and the constant, term c, is equal to 8. Plugging that into the quadratic formula, we get:

$$x = \frac{-6 \pm \sqrt{6^2 - 4(1)(+8)}}{2(1)} = \frac{-6 \pm \sqrt{36 - 32}}{2} = \frac{-6 \pm \sqrt{4}}{2}, \text{ or}$$

$$x = \frac{-6 \pm 2}{2}, \text{ meaning that } x = -4 \text{ or } -2.$$

Here's one more example: $3x^2 + 4x - 5 = 0$. Here, $a = 3$, $b = 4$, and $c = -5$. Plugging into the quadratic formula, we get:

$$x = \frac{-4 \pm \sqrt{4^2 - 4(3)(-5)}}{2(3)} = \frac{-4 \pm \sqrt{76}}{6} = \frac{-4 \pm 2\sqrt{19}}{6} = \frac{-2 \pm \sqrt{19}}{3}.$$

Let's try to plug in the equation $x^2 + 1 = 0$. According to this equation, we're squaring a number and adding 1 to it, and the result is 0; that's impossible. Plugging this into the quadratic formula, a is 1, b is 0, and c is 1, and we see the result shown below. Of course, $\sqrt{-4}$ has no solutions that are real numbers.

$$x = \frac{0 \pm \sqrt{0 - 4(1)(1)}}{2(1)} = \frac{0 \pm \sqrt{-4}}{2}$$

Let's look at another application of quadratics, called *continued fractions*. Look at the fraction $1+\frac{1}{1}$. That fraction is equal to 2, which we can write as 2/1. Now let's add $1+\frac{1}{1+\frac{1}{1}}$; the answer is $1+\frac{1}{2}=\frac{3}{2}$. If we repeat the process, the answer is equal to 1 plus the reciprocal of 3/2, which is 2/3: $1+\frac{1}{3/2}=1+\frac{2}{3}=\frac{5}{3}$. Where are we going?

By looking at these series of 1s, we're getting the fractions 2/1, 3/2, 5/3; let's do the addition one more time: $1+\frac{1}{5/3}$ = 1 + 3/5 = 8/5. Of course, we see the Fibonacci numbers in the resulting fractions, and I claim that this pattern will continue to produce Fibonacci fractions. Imagine that we have 1 + 1 over some other messy term. If that messy term reduces to, say, $1+\frac{1}{F_n / F_{n-1}}$, that simplifies to $\frac{F_n + F_{n-1}}{F_n}$. By definition, $F_n + F_{n-1}$ is equal to F_{n+1}. In other words, once we have the ratio of Fibonacci numbers and repeat this process, we can't help but get a new ratio of Fibonacci numbers.

What does this have to do with quadratics? Suppose we were to continue this process forever. Let's call the result x and solve for x:

$$1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1\ldots}}}=x$$

.

Notice that everything under the topmost fraction bar is itself equal to x; therefore x is equal to $1 + 1/x$. To solve $x = 1 + 1/x$, multiply both sides by x for a result of $x^2 = x + 1$. We subtract x and subtract 1 from both sides, leaving the equation $x^2 - x - 1 = 0$. For the quadratic equation, $a = 1$, $b = -1$, and $c = -1$. Solving, we get two solutions.

$$x = \frac{1+\sqrt{5}}{2} = 1.618\ldots \text{ (golden ratio)}$$

$$x = \frac{1-\sqrt{5}}{2} = -0.618\ldots$$

Only one of these solutions will work, and that is the positive one. We know that the negative solution is incorrect, because we can't add a group of positive numbers and end up with a negative. Incidentally, we have now proved that the ratio of Fibonacci numbers in the long run gets closer and closer to the golden ratio.

This method of solving quadratic equations was known even by the ancient Greeks, but the ancient Greeks did not know how to solve equations in a higher degree, such as a cubic equation: $ax^3 + bx^2 + cx + d = 0$. This problem was first solved by Girolamo Cardano (1501–1576), a mathematician and gambler with a rather shady past. Through various means, he discovered a formula for solving the cubic.

The search went on for a formula to solve any quartic equation, that is, an equation of the form $ax^4 + bx^3 + cx^2 + dx + e = 0$. The formula for this was determined by an Italian mathematician named Lodovico Ferrari (1522–1565). Then, the search went on for hundreds of years to find a solution to the quintic equation, an equation of fifth degree. Finally, in 1824, a Norwegian mathematician, Niels Abel (1802–1829), showed that the attempt to find a solution to the quintic equation was futile. It is impossible to find a single formula that uses nothing more than adding and multiplying and taking roots of coefficients to solve a quintic equation. ■

Suggested Reading

Barnett and Schmidt, *Schaum's Outline of Elementary Algebra*, 3rd ed.

Gelfand and Shen, *Algebra*.

Selby and Slavin, *Practical Algebra: A Self-Teaching Guide*, 2nd ed.

Questions to Consider

1. A little knowledge can sometimes be a dangerous thing. Find the flaw in the following "proof" that 1 = 2:

> Start with the equation $x = y$.
> Multiply both sides by x: $x^2 = xy$.
> Subtract y^2 from both sides: $x^2 - y^2 = xy - y^2$.
> Factor both sides: $(x + y)(x - y) = y(x - y)$.
> Divide both sides by $x - y$: $x + y = y$.
> Substitute $x = y$: $2y = y$.
> Divide both sides by y: $2 = 1$.
> Voila!!?!

2. Using the close-together method, mentally multiply each of the following pairs: 105×103, 98×93, and 998×993. Would you have found these problems easy without knowing this method?

The Joy of Algebra Made Visual

Lecture 8

In this last lecture on algebra, we're going to see what was an earthshattering idea, the idea of connecting algebra with geometry—how you could actually see an equation.

Before we look at the connection between algebra and geometry, let's talk about *polynomials*. Here are some examples: $8x^3 - 5x^2 + 4x + 7$, $x^{10} + 9x^2 - 3.2$. Even x, $x + 7$, and 7 by itself are polynomials. The *degree* of a polynomial is the largest exponent in the polynomial. For instance, the first polynomial above has degree 3 because of the x^3 term. The second polynomial has degree 10 because of the x^{10} term; x by itself has degree 1, as does $x + 7$. The constant polynomial has degree 0. We can think of that as $7(x^0)$. When dealing with polynomials, all the exponents must be whole numbers that are at least 0; no negative or fractional exponents are allowed.

Let's review the law of exponents: $x^a x^b = x^{a+b}$. Why is that true? Initially, we might think $x^a x^b$ should be x^{ab}. Let's do an example. According to the law of exponents, $x^2 x^3 = x^{2+3} = x^5$. If we look at x^2, that's xx, and x^3 is xxx; when we multiply them together, we get $(xx)(xxx)$. That's 5 x's. We also want the law of exponents to be true when the exponent is 0. That is, $x^a x^0$ should equal x^{a+0}, but $a + 0$ is a, which means that $x^a x^0 = x^a$. If we want the law of exponents to work in this case, then x^0 must be 1 so that $x^a x^0$ is still x^a. That's the reason that $x^0 = 1$.

A typical polynomial of degree n looks like this: $ax^n + bx^{n-1} + cx^{n-2} + \ldots$, in which the a, b, c, and so on—all the *coefficients*—can be any real numbers, integers or fractions. The only requirement is that if the equation is of degree n, the coefficient behind the x^n cannot be 0; if it were 0, the equation wouldn't have degree n but a smaller degree.

Now let's explore how we can actually see an equation. We'll start by looking at first-degree equations, that is, linear equations. Let's take one of the simplest linear equations, $y = 2x$, and plug in some values. When x is 0,

y is twice 0, which is 0. When x is 1, y is twice 1, which is 2. When x is 2, y is 4. When x is 3, y is 6. Notice that every time we add 1 to x, we add 2 to y. We'll now plot these points on the *Cartesian plane* (discovered by René Descartes), in which the horizontal axis is the x-axis and the vertical axis is the y-axis. For instance, when we plot (3,6), that means we go three to the right on the x-axis and six up on the y-axis; where those coordinates meet is the point (3,6). If we plot all the points and connect the dots, the result is a line that goes through the points; that's why this is called a *linear equation*.

Let's now change this equation to $y = 2x + 3$. What does that do to the graph? It adds 3 to the same points that we had earlier. The new line is parallel to the old line, but it's now higher than the old line by 3. This graph is called the *graph of the function* $y = 2x + 3$. We also give names to the coefficients behind the x term and the constant term; in this equation, $y = 2x + 3$, 2 is the *slope* of the line, and 3 is the y *intercept*. The slope tells us how much the line is increasing. As we said earlier, if x increases by 1, then y increases by 2. If x decreases by 1, then y decreases by 2. The y intercept tells us where the line crosses the y-axis; in this case, that point is (0,3). The first coordinate will always be 0, and the second coordinate will be the y intercept. Let's generalize this by looking at the equation $y = mx + b$. The slope of this equation is m, and the y intercept is b. For the y intercept, that means that the line will cross the y-axis at the point where $x = 0$ and $y = b$. The fact that the slope is m tells us that if x increases by 2, y will increase by $m \times 2$, or $2m$.

Draw some of these graphs to get a picture of them. Look at the equation $y = .5x - 4$. This equation tells us that the slope is 1/2. This line intercepts the y-axis at −4. We plot the y intercept at (0,4), and every time we increase x by 1, y increases only by 1/2.

Look at another line: $y = -4x + 10$. This line intercepts the y-axis at 10. For every increase of x by 1, the function decreases at a rate of 4. How about a line with 0 slope? Let's plug in a random constant, say $y = 1.618$. We then have a line with 0 slope. By the way, that's still called a linear equation, even though it's an equation of 0 degree. Finally, let's look at a line of infinite slope. Suppose we have the equation $x = 2$. That says $x = 2$ no matter what y is; y could be 0, 1, 100, $-\pi$, and x will always be 2. Plotting the result gives us a vertical line.

Let's now solve a geometry problem using algebra. We start with two equations: $y = 2x + 3$ and $y = -4x + 10$. Where those lines cross, y is equal to both $2x + 3$ and $-4x + 10$. Let's then set those two equations equal to each other (that is, $2x+3=-4x+10$) because where they meet, those two quantities are equal.

Now we add $4x$ to both sides and subtract 3 from both sides, resulting in $6x = 7$. Solving that, we get $x = 7/6$. At the point where the lines cross, remember that y is equal to $2x + 3$ or $-4x + 10$. For $2x + 3$, $y = 2(7/6) + 3$, or $(7/3) + 3$, or 16/3. To verify the solution, when x is equal to 7/6, where is it on the line at $-4x + 10$? Solving, $-4(7/6) = (-28/6) + 10$, or 32/6, or 16/3.

Here's a more practical question: Suppose you were offered two phone plans, and you want to decide which of those plans will save you more money in the long run. One of the plans charges a \$10.00 flat fee, plus \$0.15 for each minute you use. The other plan charges a \$20.00 flat fee, plus \$0.10 for every minute. Which plan should you choose? If you use the phone a lot, then you may want to pay the \$20.00 flat fee and get a lower rate of \$0.10 per minute. If you use your phone only a little, then you may want the \$0.15-per-minute rate with a smaller flat fee.

To find out where the critical point is, we set these two equations equal to each other. The first bill, B, is equal to \$10.00 + \$0.15M, M being the number of minutes you use. The second bill is \$20.00 + \$0.10M. Setting those two equations equal to each other, we get \$10.00 + \$0.15M = \$20.00 + \$0.10M. Putting the M's on one side and the constant terms on the other, we get \$0.05M = \$10.00. We then multiply both sides by \$20.00 to get $M = 200$.

If you use 200 minutes or more, then you want the plan that has the lower per-minute rate. If you use under 200 minutes per month, then you want the plan that has the \$10.00 fee, plus \$0.15 a minute. Again, the solution is worth verifying. In this case, if you used 200 minutes, whether you use the first plan or the second plan, your bill would be \$40.00, which corresponds to the point on the graph where those two lines cross, (200,40).

Let's graduate from first-degree equations to second-degree equations. The equation $y = x^2$ is called a *quadratic equation*, and it's the simplest of second-degree equations. The graph that's drawn from this equation looks like a parabola. If we change the equation to $y = 2x^2$, the graph still has the same basic shape, but y increases much faster than it did in the first equation. If we change the equation to $y = x^2 + 2$, we increase y by 2 everywhere.

The equation $y = x^2$ is called a *quadratic equation*, and it's the simplest of second-degree equations.

The equation $y = (x - 2)^2$ will shift the parabola two units to the right. To see why the graph moves to the right, we look at what happens when $x = 2$. When $x = 2$, then $y = 0^2$, which is 0; thus, we shift to the right at the point on the parabola where $x = 2$. The equation $y = (x + 2)^2$ will shift the parabola to the left. Notice that if we start with the earlier equation, $y = (x - 2)^2$ and subtract 2, that brings the whole parabola down. The equation becomes $x^2 - 4x + 4 - 2$, or $x^2 - 4x + 2$, which looks like a generic quadratic equation: $y = x^2 - 4x + 2$.

Even though the second equation looks different from the equations we've seen before, it's nothing more than a shifted parabola. In fact, the same is true for any quadratic equation. For instance, look at $y = x^2 - 8x + 10$. Using the technique of completing the square, we can rewrite that as $(x^2 - 8x + 16) - 6$, replacing the 10 with $16 - 6$. The quantity in parentheses is equal to $(x - 4)^2$. The equation can be written as $y = (x - 4)^2 - 6$, and the graph is a parabola shifted to the right by 4 and lowered by 6. No matter what the second-degree equation is, it will result in a parabola that intersects the x-axis once, twice, or zero times.

Let's now look at third-degree equations. Looking at the graphs for $y = x^3$, $y = x^3 + 4x^2 + 4x$, and $y = x^3 - 7x + 6$, we see that these cubics cross the x-axis at most three times. The general rule for this is called the *fundamental theorem of algebra*, proved by Gauss. According to this theorem, the graph of a polynomial of degree n will intersect the x-axis at most n times. As we saw, for the quadratic equations, $n = 2$, and for the cubic equations,

$n = 3$. Equivalently, if $P(x)$ is a polynomial of degree n, the number of solutions to the equation $P(x) = 0$ is at most n. For instance, for the equation $2x^{10} - 7x^4 + 5x + 9 = 0$, the fundamental theorem of algebra tells us that there are at most 10 solutions to that problem.

So far, we've been dealing with polynomials, which have exponents that are non-negative integers. Let's now look at some other kinds of exponents. For instance, what does the quantity x^{-1} mean? If we want the law of exponents to be a law for all numbers, we want $x^a x^b = x^{a+b}$. What happens if we plug $a = -1$ and $b = +1$ into the law of exponents? We get $x^{-1}x$ (that is, $x^{-1}x^1) = x^{-1+1}$, which is x^0, or 1. In other words, x^{-1} when multiplied by x gives us 1. This means that x^{-1} is the reciprocal of x—that is, $x^{-1} = 1/x$ for any x not equal to 0. For instance, 3^{-1} is 1/3, and -7^{-1} is $1/-7$, or $-1/7$. But 0^{-1} is forever undefined.

Let's look at 3^{-2}, which is 3^{-1-1}, or (1/3)(1/3), or $(1/3)^2$, or 1/9. In other words, x^{-1} is $1/x^1$; x^{-2} is $1/x^2$. By the same logic, x^{-n} is $1/x^n$. Looking at the graph of $y = 1/x$, for example, we see that as x gets closer to 0 from the right, $1/x$ gets closer to infinity. On the left side, as x gets closer to 0, y gets closer to negative infinity.

Using the law of exponents, what should $9^{1/2}$ mean? We want it to be true that $x^a x^b = x^{a+b}$ even when a and b are fractions. By the law of exponents, $(9^{1/2})(9^{1/2})$ is $9^{1/2+1/2}$; it's also equal to 9^1, which is 9. That tells us that $(9^{1/2})(9^{1/2}) = 9$. In other words, $9^{1/2} = 3$. You might think that if $9^{1/2} = 3$, then also $9^{1/2} = -3$ because $(-3)(-3) = 9$. However, we want $9^{1/2}$ to be well defined; thus, mathematicians define $9^{1/2}$ to be equal to 3, not -3. In general, $x^{1/2}$ is equal to $+\sqrt{x}$. For instance, $\sqrt{9} = 3$, $\sqrt{16} = 4$, $\sqrt{1} = 1$, and $\sqrt{0} = 0$.

Look at a graph of $y = \sqrt{x}$. Notice that we see only the graph to the right of $x = 0$, because to the left of $x = 0$ are the square roots of negative numbers, which we're not quite ready to handle yet. What should $x^{1/3}$ mean? By the law of exponents, $x^a x^b x^c = x^{1/3+1/3+1/3}$, which is x^1, or x. Therefore, $x^{1/3}$ is the cube root of x, which is denoted by $\sqrt[3]{x}$. For instance, $\sqrt[3]{8} = 2$; $\sqrt[3]{27} = 3$; and $\sqrt[3]{2} = \sqrt[3]{2}$, which numerically is about 1.259. We can also look at the graph of the cube root function.

Before we close, let's look at a couple of other important graphs. For instance, we see the equation of the unit circle. The circle is centered around the *origin*—that's the (0,0) point—with a radius of 1; the equation for this is $x^2 + y^2 = 1$, or $y^2 = 1 - x^2$. Technically, we might say that the top half of the circle is $y = \sqrt{1-x^2}$ and the bottom half of the circle is $y = -\sqrt{1-x^2}$, but it's cleaner to put the top half and the bottom half together to get the equation $x^2 + y^2 = 1$.

Let's look at a more general circle. Instead of intercepting the x-axis and y-axis one away from the origin, suppose we intercept them r away from the origin; the equation then is $x^2 + y^2 = r^2$. Here's another example: $x^2 + y^2 = 10^2$, or 100, would be a circle of radius 10. If we shift that circle two units to the right, the equation would be $(x - 2)^2 + y^2 = 10^2$. If we then pushed it up by one unit, the equation would be $(x - 2)^2 + (y - 1)^2 = 10^2$.

In this lecture, we've seen polynomials and how to graph them. We've also talked about the fundamental theorem of algebra and about negative and fractional exponents. In the next lecture, we'll see what joy we can find in the number 9. ■

Suggested Reading

Barnett and Schmidt, *Schaum's Outline of Elementary Algebra*, 3rd ed.

Gelfand and Shen, *Algebra*.

Selby and Slavin, *Practical Algebra: A Self-Teaching Guide*, 2nd ed.

Questions to Consider

1. At the start of a baseball game, your favorite player has a batting average of .200. During the game, he has two hits and strikes out twice. At the beginning of the next game, you notice that his batting average is now .250. How many hits has he had this season?

2. Speaking of batting averages, suppose player A has a better batting average than player B for two consecutive seasons. Must it be the case that player A's combined batting average for both seasons is better than player B's? Surprisingly, the answer is no. Can you find some numbers that support this paradox?

3. Use the fundamental theorem of algebra to prove that if two quadratic polynomials agree for three different values of x, then they must be equal. In general, show that if two n^{th}-degree polynomials agree for $n + 1$ different values of x, then they must be the same polynomial.

The Joy of 9

Lecture 9

Where this becomes I think fun, but also useful, practical, is we can use this idea about 9s as a way of checking our arithmetic. We can use this to check addition, subtraction, and multiplication problems.

Let me begin with a magic trick. Think of a number between 1 and 10. Now, take that number and triple it. Take the number you have now and add 6 to it. Take that number and now triple it again. Now take your answer, probably a two-digit number, and add the digits of your answer. If you still have a two-digit number, add those digits again. Now you're thinking of a one-digit number—I see it; you got the number 9, right?

Let's see why that works. Let's call the first number that you thought of x. Tripling x gives you $3x$. When you add 6, you get $3x + 6$. When you triple that result, you get $3(3x + 6)$, or $9x + 18$. That last equation, $9x + 18$, is the same as $9(x + 2)$; thus, the number you get is guaranteed to be a multiple of 9.

Let's see what the first several multiples of 9 have in common. You may have learned in elementary school that if a number is a multiple of 9, its digits will sum to 9 or a multiple of 9. For example, adding the digits in 18 yields $1 + 8 = 9$, as does adding the digits in 27: $2 + 7 = 9$. The rule is: A number is divisible by 9 if and only if the sum of its digits is a multiple of 9.

Let's do an example: 3,456. Adding the digits together, we get 18, and 18 is a multiple of 9; therefore, 3,456 is a multiple of 9. What about the number 1,234? Its digits add to 10, and if we add the digits in 10, we get 1; therefore, 1,234 is not a multiple of 9. However, that 1 is the remainder when we divide 1,234 by 9. This same rule works for multiples of 3. A number is divisible by 3 if and only if its digits add up to a multiple of 3.

Look again at 3,456; this number is $(3\times1{,}000)+(4\times100)+(5\times10)+6$. We can break 1,000 into 999 + 1, we can break 100 into 99 + 1, and we can break 10 into 9 + 1; and 6 stays 6. If we expand on this, we get $(3\times999) + (4\times99)$

+ (5 × 9), and we're left with a dangling 3 × 1, which is 3; 4 × 1, which is 4; 5 × 1, which is 5; and 6. We know that 3 × 999 is a multiple of 9, 4 × 99 is a multiple of 9, and 5 × 9 is a multiple of 9; thus, all those combine to be some multiple of 9. Plus we have 3 + 4 + 5 + 6, which is 18, also a multiple of 9; and adding 18 to a multiple of 9 still gives us a multiple of 9. With 1,234, the same idea applies. This number is (1 × 1,000) + (2 × 100) + (3 × 10) + 4. Expanding the 1,000s, 100s, and 10s as we did before, we get (1 × 999) + (2 × 99) + (3 × 9); plus we have 1 + 2 + 3 + 4, which equals 10, but 10 can be broken into 9 + 1. That leaves us with (a multiple of 9) + 9 + 1; therefore, as promised, 1,234 is 1 greater than a multiple of 9.

This may sound a bit abstract, but in fact, you do modular arithmetic every day.

We can use this idea about 9s to check addition, subtraction, and multiplication problems. Suppose we want to add 3,456 and 1,234. Let's check the answer, 4,690, using a process called *casting out 9s*. We reduce the number 3,456 by adding all the digits together, giving us 18; we then reduce 18 by adding its digits together to get 9. Then, we reduce 1,234 by adding its digits to get 10, and we add the digits of 10 to get 1. We've changed the original problem, 3,456 + 1,234, to the easier problem of 9 + 1 = 10, and we add the digits of that answer to get 1. When we check our original answer, 4,690, we should get a 1 at the end of the process. The digits of 4,690 add up to 19; the digits of 19 up to 10; and the digits of 10 add up to 1.

Because we got a match, we can have confidence in our answer. If the ending numbers did not match, we'd know that we had made a mistake. Note, however, that we could get a match and still have an incorrect answer. Why does this work? We know, from our earlier calculation, that 3,456 is a multiple of 9; it's $9x + 0$. We also know that 1,234 is $9y + 1$; therefore, when we add $9x + (9y + 1)$, we get $9(x + y) + 1$, which means that in the end, the answer will reduce to 1.

Let's do a bigger problem: 91,787 + 42,864. If we add those numbers together correctly, we get 134,651. To check the answer, we add the digits of 134,651 to get 20; we then add the digits of 20 to get 2. Next, we add the digits of 91,787 to get 32, which simplifies to 5, and we add the digits of 42,864 to get

24, which simplifies to 6. Finally, 5 + 6 = 11, and those digits, 1 + 1, add up to 2. Because the two numbers match, we can have confidence in our answer. Again, this method won't reveal all mistakes. If we accidentally mix up two digits—for example, ending with 561 instead of 651—the numbers will still match, but the error won't be caught.

This method also works for subtraction problems, such as 91,787 − 42,864 = 48,923. If we add the digits of 48,923, we get 26; adding those digits, we get 8. Adding the digits of the first number, 91,787, gives us 32, which reduces to 5. The second number, 42,864, simplifies to 24, which reduces to 6. Because this is a subtraction problem, we subtract 5 − 6, which gives us −1. Remember that −1 is simply the remainder we get when we divide our answer to the original subtraction problem by 9. We can always change that number by adding or subtracting multiples of 9; thus, we'll add a 9 to −1 to get 8. The two reduced numbers match again.

Surprisingly, this method also works for multiplication problems. Let's multiply the same two numbers: 91,787 × 42,864 = 3,934,357,968. We first add the digits of that 10-digit number to get 57; we then add the digits of 57 to get 12, and the digits of 12 to get 3. As before, 91,787 reduces to 5, and 42,864 reduces to 6. Because this is a multiplication problem, we multiply 5 × 6, which gives us 30. Those digits add up to 3, and we have a match again. Why does this work? Basically, this algebraic statement explains it: $(9x + 5)(9y + 6) = 9(9xy + 5x + 6y) + 30$. According to this, if we have a number of the form $9x + 5$ and we multiply that by a number of the form $9y + 6$, we get a number of the form: 9(something) + 30, which is 9(something + 3) + 3.

The ideas behind this method actually extend beyond the number 9. If we want to say 42,864 = 6 + some multiple of 9, the notation we use is 42,864 = 6 (mod 9). To clarify, we say that $a = b$ (mod 9) if $a = (b$ + some multiple of 9). In other words, $a = b + 9k$, where k is some integer. In general, we say that for any integer m (not just the number 9), $a = b$ (mod m) if $a = b$ + some multiple of m. Another way to say that is $a = b + mk$, where k is any integer. Yet another way of saying it is that the number m divides the difference of $a - b$.

We can do what is called modular arithmetic for any integer m. For example, using the same logic we used to demonstrate casting out 9s, we can show that if $a = b$ (mod m) and $c = d$ (mod m), then $a + c$ equals $b + d$ (mod m). Translated, that says that if a and b differ by a multiple of m, and c and d differ by a multiple of m, then $a + c$ and $b + d$ will differ by a multiple of m. Moreover, $ac = bd$ (mod m). If we multiply ($a = b$) by ($a = b$), we get $a^2 = b^2$ (mod m). Multiply that by ($a = b$), and we get $a^3 = b^3$, $a^4 = b^4$, and in general, $a^n = b^n$ (mod m).

This may sound a bit abstract, but in fact, you do modular arithmetic every day. For instance, if the clock reads 12:00 right now, then what time will it read in 17 hours? You might reason as follows: 17 hours is 12 hours + 5 hours; ignoring the 12, the clock will read 5:00. What time will it be in 29 hours, or 41 hours? To get the answer, we just add more multiples of 12, and we can ignore 12 when we're looking at a clock. We're working, then, in mod 12. Here's another example: What will the clock read 1,202 hours from 9:00? To find 1,202 (mod 12), we go around the clock 100 times + 2; 1,202 is 2 (mod 12). In 1,202 hours, the clock will read 9 + 2, or 11:00.

By working in mod 7, we can use this same approach to find the day of the week of any date in history. First, we figure out the day of the week of any date in the year 2007. For this, we need to memorize a year code, which for 2007, is 0. Then, we need to memorize a code for every day of the week. Saturday is 7 or 0 because we're doing this in mod 7. Next, we memorize a code for every month of the year. It's easiest to remember this code if you look at the months in groups of three.

Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.
1	2	3	4	5	6	7 or 0

Jan.	1	Apr.	0	July	0	Oct.	1
Feb.	4	May	2	Aug.	3	Nov.	4
Mar.	4	June	5	Sept.	6	Dec.	6
mnemonic	$= 12^2$		$= 5^2$		$= 6^2$		$=12^2 + 2$

Let's figure out the day of the week of December 25, 2007. Start with the month code for December, 6, and add 25 for the date. Then, for 2007, we add the year code, 0: $6 + 25 + 0 = 31$. We could count the days and wrap around the calendar until we get to 31, but we don't have to, because every seven days, the week repeats. Day 31 will be the same as 31 minus any multiple of 7. The biggest multiple of 7 below 31 is 28, and $31 - 28 = 3$; day 3 in the code is Tuesday. Thus, Christmas in 2007 is a Tuesday.

We know that Thanksgiving 2007 is a Thursday in November, but what is the date? The month code for November is 4. We'll call the date x, and the year code for 2007 is 0. Our equation, then, is $4 + x + 0 = 5$, because Thursday is day 5. What do we add to 4 to get 5 (mod 7)? We must add 1 or something that differs from 1 by a multiple of 7; thus, x will be 1, 8, 15, 22, or 29. The holiday occurs on the fourth Thursday of the month, or the 22nd.

Why does this work? Think about what happens to your birthday as you go from one year to the next—it bumps up by exactly one day. That's because there are usually 365 days in between your birthdays, and 364 is a multiple of 7 ($7 \times 52 = 364$). The exception is that in a leap year, there are 366 days between your birthdays, unless you were born in January or February and the year hasn't leaped yet. If we put all that together, we can figure out the year codes. Remember that 2007 has a year code of 0, but 2008 is a leap year; it will have 366 days. Thus, the year code for 2008 should be 2, except in January or February, when we have to subtract 1. The year code for 2009 is 3; for 2010, 4; for 2011, 5; and for 2012, another leap year, 7. Of course, we can reduce 7 (mod 7) to 0, which means that 2012 has a year code of 0.

Incidentally, the year 1900 has a year code of 0, and knowing that fact, we can derive the year codes for every subsequent date. How could we figure out the year code for 1961, for example? The year 1961 is 61 years after 1900; thus, the calendar will shift 61 times, but it will also shift an extra time for each leap year. There were 15 leap years between 1900 and 1960. We take 1/4 of 61, which is 15, then add $61 + 15$, and that's 76. We could make 76 the year code for 1961, but it's much simpler to look at 76 (mod 7). Subtract the biggest multiple of 7 less than 76, 70, and we get $76 - 70 = 6$, the year code for 1961. Thus, for March 19, 1961, we compute $6 + 4 + 19 = 29$, then subtract 28 for an answer of 1. So March 19, 1961 was a Sunday.

Let's look at one more example: July 22, 1987. We start by finding the year code for 1987: We take 1/4 of 87; that's 21 with a remainder of 3. In this trick, we always ignore the remainder. We add 87 + 21 to get 108 and subtract the biggest multiple of 7, which is 105. Next, 108 − 105 = 3; that's the year code for 1987. To that, we add 0 for the month code of July and 22 for the date: 3 + 0 + 22 = 35. Subtract the biggest multiple of 7, 21, for an answer of 4. July 22, 1987, was a Wednesday.

Here's one last challenge: Pick any four-digit number in which the digits aren't all the same. Let's use 1,618. Now, scramble those numbers to get a different number, such as 8,611. Subtract the smaller number from the larger number. In this case, you'll get 6,993. Next, add the digits: 6 + 9 + 9 + 3 = 27. If you have a two-digit number, add the digits again to get a one-digit number. The resulting number is 9. ■

Suggested Reading

Benjamin and Shermer, *Secrets of Mental Math*, chapters 6, 9.

Gardner, *Mathematics, Magic, and Mystery*.

———, *The Second Scientific American Book of Mathematical Puzzles and Diversions*, pp. 43–50.

Gross and Harris, *The Magic of Numbers*, chapters 15–16.

Reingold and Dershowitz, *Calendrical Calculations*.

Questions to Consider

1. If you take any number and scramble its digits, then subtract the original number from the scrambled one, you always get a multiple of 9. Why?

2. In the Fibonacci sequence 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ..., explain why every fifth number must be divisible by 5.

The Joy of Proofs

Lecture 10

Here's the proof that all numbers are interesting: Suppose that not all numbers were interesting; then, there would have to be a first number that wasn't interesting. But wouldn't that make that number interesting?

Let's start with something that we already know to be true to give an example of an easy proof. We'll use the statement that an even number plus an even number always equals an even number. We say that an integer, a, is an even number if $a = 2b$, where b is any integer. Here's our theorem: If x and y are even numbers, then $x + y$ is also an even number. The following is our proof.

Because x is an even number, then $x = 2a$, where a is an integer. Because y is an even number, then $y = 2b$, where b is an integer. Thus, $x + y = 2a + 2b$, or $2(a + b)$. The quantity $a + b$ is an integer. We know that because a is an integer and b is an integer; therefore, their sum is an integer. Because $x + y$ is twice an integer, $x + y$ is even. Completed proofs end with a filled or empty box sometimes known as a Halmos symbol (■); or with QED (*quod erat demonstrandum*; humorously translated as "quite easily done"); or with ☺.

For the next proof, note that an integer is an odd number if the number is not even. A more mathematical statement of that is: An odd number a is a number that's of the form $2b + 1$. Here's the theorem we'll prove: If x and y are odd numbers, then their product is an odd number.

If x is an odd number, then x is of the form $2a + 1$. If y is an odd number, then $y = 2b + 1$, where a and b are integers; therefore, their product, xy, is $(2a + 1)(2b + 1)$. When we multiply those quantities, we get $4ab + 2a + 2b + 1$; notice that the first three terms are all divisible by 2. Simplifying, we get $2(2ab + a + b) + 1$. Thus, xy is twice an integer plus 1; therefore, xy is an odd number.

Let's graduate from numbers that are even and odd to rational numbers. Those are fractions that are obtained by taking the quotient of two integers. Specifically, we say that a number r is rational, if $r = p/q$, where p and q are integers; of course, q cannot be 0 because that would mean we were dividing by 0. Here's our theorem to prove: The average of two rational numbers is always rational.

We'll use r_1 as a rational number in the form p_1/q_1 and r_2 in the form p_2/q_2; their average is $(r_1 + r_2)/2$. We add the fractions, $p_1/q_1 + p_2/q_2$, and divide the result by 2. The result is $\frac{p_1q_2 + p_2q_1}{2q_1q_2}$. The numerator is the product and sum of integers; the denominator is the product of integers. Because we have an integer in the numerator and the denominator, we know that the average of those two numbers, r_1 and r_2, is rational. The consequence of that statement is that between any two rational numbers, we can find another rational number.

If you were to look at the number line, you might believe, as some of the ancient Greeks once did, that all numbers are rational, that is, that every number out there is equal to some fraction. Pythagoras himself believed that and would have been surprised to learn that one of the numbers derivable from his own theorem was not a rational number. According to the Pythagorean theorem, for any right triangle with side lengths a and b and hypotenuse c, $a^2 + b^2 = c^2$. Using the simplest of right triangles, with side lengths 1 and 1, the hypotenuse would have a length of $\sqrt{2}$ because $1^2 + 1^2 = 2$.

If you were to look at the number line, you might believe, as some of the ancient Greeks once did, that all numbers are rational, that is, that every number out there is equal to some fraction.

We can approximate $\sqrt{2}$ with a decimal expansion ($\sqrt{2} = 1.414213\ldots$), but we cannot write $\sqrt{2}$ exactly as a fraction. We might think that $\sqrt{2}$ is not rational because its decimal expansion doesn't repeat, but how can you be sure that it doesn't repeat? That would also require proof. To prove that $\sqrt{2}$ is irrational, we'll use a *proof by contradiction*. Suppose $\sqrt{2}$ were rational; if that were true, then it would be of the form p/q in lowest terms because we can write all fractions in lowest terms. The equation is $\sqrt{2} = p/q$. Squaring both sides, we get $2 = p^2/q^2$. We can rewrite that equation as $p^2 = 2q^2$, but that says that p^2 is twice a number, which means that p^2 is even. If p^2 is even, then p must be even because if p were odd, its square would be odd. That means that the number p must be of the form $2b$ because it's an even number.

Let's return to our equation, $p^2 = 2q^2$, and replace p with $2b$. When we do that, we have the expression $(2b)^2 = 2q^2$. When we square $2b$, we get $4b^2$, and that's equal to $2q^2$. Dividing both sides by 2, we get $q^2 = 2b^2$. Again, that means that q^2 is even, and if q^2 is even, then q is also even. Now we have a problem. We tried to prove that $\sqrt{2}$ is a rational number, p/q in lowest terms. Then, we showed that both p and q had to be even. The problem is that if p is even and q is even, then that fraction wasn't in lowest terms. If we assume that $\sqrt{2}$ is rational and in lowest terms, we conclude that $\sqrt{2}$ is not in lowest terms; therefore, the only conclusion we can make is that $\sqrt{2}$ is not rational.

Next, we'll prove what some mathematicians call an *existence theorem*: There exist irrational numbers a and b such that a^b is rational. In other words, we can find an irrational number raised to an irrational power that yields a rational number. What makes this an existence proof is that we'll become convinced of its truth without knowing what a and b are. We begin by asking a simple question: Is $\sqrt{2}$ raised to the power of $\sqrt{2}$ rational? If $\sqrt{2}^{\sqrt{2}}$ is rational, then both a and b would be irrational, yet a^b would be rational, and our proof would be complete. What if, however, $\sqrt{2}^{\sqrt{2}}$ is irrational? In that case, we could let $a = \sqrt{2}^{\sqrt{2}}$ (which we're assuming is irrational), and we could let $b = \sqrt{2}$ (which we've just shown is irrational); thus, $a^b = \sqrt{2}^{\sqrt{2}}$.

According to the law of exponents, $(a^b)^c$ is the same as a^{bc}. When we apply that here, we have $(\sqrt{2})^{\sqrt{2}(\sqrt{2})}$. But that's equal to $\sqrt{2}^2$, which is equal to 2. In this case, then, we found an a and a b such that $a^b = 2$, which is rational. The first question we asked was: Is $\sqrt{2}^{\sqrt{2}}$ rational? If the answer was yes, then our proof was complete. If the answer was no, then by choosing a to be $\sqrt{2}^{\sqrt{2}}$ and b to be $\sqrt{2}$, our proof is also complete.

Now let's look at a proof technique called *proof by induction*. We'll look at a problem that we saw in Lecture 2: What is the sum of the first n odd numbers? Recall that 1 = 1, or 1^2; 1 + 3 = 4, or 2^2; 1 + 3 + 5 = 9, or 3^2; 1 + 3 + 5 + 9 = 25, or 5^2; and so on. Will this pattern go on forever? The sixth odd number is 11. When we add that to 25, we get 36, which is 6^2. If we trust the first five results, then the sixth result will follow. Suppose we notice that the sum of the first k odd numbers is k^2. In other words, since the k^{th} odd number is $2k - 1$, we are asserting that 1 + 3 + 5 +… + k^{th} (that's the number $2k - 1$) = k^2. Then, what will be the sum of the first $k + 1$ odd numbers? What is the next odd number? That's $2k + 1$. When we add that to the sum of the first k odd numbers, or k^2, we get $k^2 + 2k + 1$, which is $(k + 1)^2$. In other words, if the sum of the first k odd numbers is k^2, then it's unavoidable that the sum of the first $k + 1$ odd numbers will be $(k + 1)^2$.

Here's another example of a proof by induction: Recall that the sum of the first n odd numbers, which we called triangular numbers, is equal to $n(n + 1)/2$. If we're interested in summing the cubes of the first n numbers, we can find a pattern: $1^3 = 1$, $1^3 + 2^3 = 9$, and $1^3 + 2^3 + 3^3 = 36$. The results are all perfect squares of the triangular numbers; that is, the sum $1^3 + 2^3 + \ldots + 5^3 = 225$, or 15^2, which is equal to $(1 + 2 + 3 + 4 + 5)^2$, or $(5 \times 6)/2$.

We'll assert, then, that the sum of the cubes of the first n numbers equals the n^{th} triangular number squared; that is, $n(n + 1)/4$. We'll start with a base case. We see that the statement works for the number 1; undeniably, $1^3 = 1^2$. Then, we state our *induction hypothesis*: Suppose that the sum of the cubes of the first k numbers is $k(k + 1)/4$. We'll use that fact to show that the statement will continue to be true when we look at the sum of the cubes of the first $k + 1$ numbers. What do we want to see at the end of that? If we replace k with $k + 1$ in the above formula, we want to see $(k + 1)^2(k + 2)^2/4$.

The sum of the first $k + 1$ cubes is equal to the sum of the first k cubes plus $(k + 1)^3$. But what do we know about the sum of the first k^3? By our induction hypothesis, we know that quantity is equal to $k^2(k + 1)^2/4$. We'll add to that the number $(k + 1)^3$. When we add that, we can save ourselves a lot of messy algebra by factoring out the number $(k + 1)^2$ because that divides both the first term and the second term. We're left with the results shown below.

$$(k+1)^2\left(\frac{k^2}{4}+\frac{4(k+1)}{4}\right)=(k+1)^2\left(\frac{k^2+4k+4}{4}\right)=\frac{(k+1)^2(k+2)^2}{4}$$

Let's try a question that doesn't involve any algebra or symbol manipulation: Can we cover an 8-by-8 checkerboard with non-overlapping L-shaped dominoes (called *trominoes*)? An 8-by-8 checkerboard has 64 squares, and if a tromino takes up 3 squares, we won't be able to cover the board evenly, since 64 is not a multiple of 3. If we remove any square at all from the checkerboard, can we cover the rest of the board with trominoes? In this case, 3 divides evenly into 63, so it might be possible. Let's prove that it is, in fact, possible, and that it's also true for 2-by-2 checkerboards, 4-by-4 and 8-by-8 checkerboards, and so on—any 2^n-by-2^n checkerboard.

Let's use a 2-by-2 board, or 2^1 by 2^1, as our base case. We can see that if we remove any square from this board, then the rest of the board can be covered with a single tromino. We assume that this is true for any board of size 2^k by 2^k. We'll now see that it's true for any board of size 2^{k+1} by 2^{k+1}. We have a checkerboard with dimensions 2^{k+1} by 2^{k+1}. We break that checkerboard into four quadrants, so that each quadrant is now of size 2^k by 2^k. Look at the quadrant that we deleted the square from. We know by the induction hypothesis that the rest of that quadrant can be covered with non-overlapping trominoes. But how do we cover those other three quadrants?

Let's look at one of the other three quadrants. We know that if we remove any square from that quadrant, then the rest of it can be covered with non-overlapping trominoes. Let's remove the square closest to the center of the board, then tile the rest of that quadrant. We can do the same for each of the remaining quadrants. We've now covered the entire board except for three squares we removed near the center of the board. But those three squares form a tromino themselves. Thus, when we place a tromino over those squares, we've completely covered our 2^{k+1}-by-2^{k+1} checkerboard. That proof is not only inductive but also constructive. It tells us how we can cover the 8-by-8 board. We could start with the 8-by-8 board, remove one square, and go through the procedure we outlined to systematically cover the rest of the board.

Could we tile an 8-by-8 checkerboard with dominoes? Dominoes have dimensions of 2 by 1; they cover two consecutive squares. We could easily cover an 8-by-8 checkerboard with dominoes, but we couldn't do so if we removed a square from the checkerboard. Could we cover the board with dominoes if we removed any two squares? We could cover the board if we removed two side-by-side squares, but what if we remove two squares in opposite corners? A checkerboard has red squares and white squares; thus, any domino that we place on the board must cover a white square and a red square. When we removed two squares, we were left with 62 squares, which means that we would need 31 dominoes to cover the board. But the two squares that we removed were both the same color. The resulting board has 30 red squares and 32 white squares, so there's no way we could cover it with 31 dominoes.

Let's end this lecture with a proof that all numbers are interesting. For instance, 1 is the first number, which is obviously interesting. Then, 2 is the first even number, which makes it an interesting number, too. Because 3 is the first odd prime number, that's interesting. Then, 4 is the first and only number that spells itself: F-O-U-R. Here's the proof that all numbers are interesting: Suppose that not all numbers were interesting; then, there would have to be a first number that wasn't interesting. But wouldn't that make that number interesting? ■

Suggested Reading

Burger, *Extending the Frontiers of Mathematics: Inquiries into Proof and Argumentation*.

Gross and Harris, *The Magic of Numbers*, chapters 15–16.

Velleman, *How to Prove It*.

Questions to Consider

1. Prove by induction that the sum of the first n Fibonacci numbers is one less than the $(n + 2)^{th}$ Fibonacci number. For example, $1 + 1 + 2 + 3 + 5 + 8 = 21 - 1$. Prove that the sum of Fibonacci numbers that are two apart is always a Lucas number (where the Lucas numbers are 2, 1, 3, 4, 7, 11, 18, 29,...).

2. Place a rook on any point on an 8-by-8 checkerboard. Show that it is possible to move the rook (making only horizontal or vertical moves) in such a way that it visits every square on the checkerboard exactly once and ends at the same point. Use this to prove that if we remove any two squares of opposite color from the checkerboard, then we can cover the remaining squares with 31 dominoes. (Hint: What can you say about the number of steps to walk from one square to another of the same color if only horizontal and vertical steps are allowed?)

3. Prove that the number log 2 is irrational, where the log is base 10. (Hint: Prove by contradiction.) In fact, except when n is a power of 10, log n is irrational.

The Joy of Geometry

Lecture 11

In this lecture, we'll talk about the joy of geometry, the oldest of the mathematical sciences. Geometry literally means *geometria* from the Greek, meaning "to measure the earth"—*geo* for "earth" and *metria* for "measurement."

The term *geometry* is Greek and literally means "to measure the Earth." The oldest textbook on the subject is *The Elements* by Euclid. In this lecture, we'll learn how to measure lengths, angles, and areas.

We begin by defining our terms and looking at basic geometric objects.

- A *point* is an infinitely small dot.
- A *line* is an infinite one-dimensional object and is named by naming two points that lay on the line ($\overleftrightarrow{AB}$).
- A *ray* is like a line except that it has one endpoint and proceeds out infinitely from that point in only one direction. It is named by giving the endpoint and one other point on the ray ($\overrightarrow{OA}$).
- A *line segment* is the portion of a line between two points ($\overline{AB}$).
- An *angle* results when two rays share the same endpoint. It is named by giving one point on one ray, the endpoint, and a point on the other ray ($\angle AOB$).
- Two lines are *parallel lines* ($\parallel$) if they never intersect.
- Two lines are *perpendicular lines* ($\perp$) if their intersection results in four equal, or right, angles.
- All of geometry is built from points and lines and comes from five axioms of Euclidean geometry:

- A straight line segment can be drawn joining any two points.
- Any line segment can be extended indefinitely.
- Given any line segment, a circle can be drawn with that segment as the radius and one endpoint as the center.
- All right angles are congruent and measure 90°.

Given a line and a point not on the line, there is *exactly one* line through the point that is parallel to the original line. This is equivalent to the axiom that the sum of the angles in any triangle is always equal to 180°.

For example, here's a theorem: Every straight line has an angle of 180°. Look at $\angle AOB$. We have a straight line that goes through A and B. We can bisect the line AB at a right angle by drawing a line from O to C; that line is called a *bisector*. We can now obtain $\angle AOB$ by adding $\angle AOC$ to $\angle COB$. Because both of those were right angles, then according to the fourth axiom, they both measure 90°. Adding 90° + 90° gives us 180°.

All of geometry is built from points and lines and comes from five axioms of Euclidean geometry.

Let's prove the *vertical angle theorem*: Suppose AB and CD are lines that intersect at point O. Then $\angle \mathrm{AOC} = \angle \mathrm{BOD}$. (In other words, the measure of angle AOC equals the measure of angle BOD. Some authors write this as $m\angle \mathrm{AOC} = m\angle \mathrm{BOD}$.) We know that the measurement of the line AB is 180°. That means that $\angle AOC + \angle COB$ must sum to 180°. Those two angles are called *supplementary angles* because they add up to 180°. Similarly, if we look at the line COD, we see that $\angle COB + \angle BOD$, because they form a straight line, also sum to 180°. We can now subtract these algebraic equalities to get: $\angle AOC - \angle BOD = 180 - 180 = 0$. In other words, $\angle AOC = \angle BOD$.

Let's look at the *corresponding angle theorem*: If L_1 and L_2 are parallel lines, and a third line crosses the pair (the *transverse line*), then the corresponding angles formed by the third line must be equal. We'll prove that $\angle A$ and

$\angle B$ are equal by drawing a new line from point C to create two new right triangles. By Euclid's fifth postulate, we know that the sum of the angles in any triangle is 180°. Thus, the angles in both of our triangles must sum to 180°. Subtracting, we see that $\angle A - \angle B = 0$. That is, $\angle A = \angle B$.

We know that the sum of the angles of any triangle is 180°, but what about larger objects? Four-sided objects are called *quadrilaterals*, or *4-gons.* Five-sided objects or many-sided objects are called *polygons*. The sum of the angles of any four-sided figure is 360°. By drawing a diagonal from one corner to the opposite corner of a four-sided figure, we create two triangles whose angles each sum to 180°; when we add the angles, we get 360°.

The sum of the angles in a pentagon is 540°. If we cut a small triangle off the top of the pentagon, we're left with a quadrilateral whose angles sum to 360°. When we add the angles of the extra triangle back in, we get 360° + 180° = 540°. We're essentially doing a proof by induction here to show that for any n-sided polygon, the sum of the angles will be always be $(n - 2)180°$.

Perimeter and *area* are two other terms used frequently in geometry. The perimeter of an object is the sum of the lengths of its sides. For a rectangle with a base of length b and a height of length h, this is $2b + 2h$. We define the area of a 1-by-1 square to be 1. We then attempt to define all areas in terms of that unit quantity. For instance, if we have a rectangle that has a height of 3 and a base length of 4, we can show that the area of that rectangle is 12 simply by cutting it into 12 squares of dimensions 1 by 1. Those squares all have area 1; therefore, the area of the rectangle is 12. The area of a rectangle whose sides have positive lengths is $b \times h$.

We can also show that the area of any triangle with a base of length b and a height of length h has an area equal to $\frac{1}{2}(b \times h)$. We adjoin two right triangles, each with a base of b and a height of h, to create a rectangle. The area of that rectangle is bh. Given that the first triangle and the second triangle have the same area, then the area of the first triangle must be $\frac{1}{2}bh$.

It seems odd that any triangle will have an area of $\frac{1}{2}bh$. Imagine we have parallel lines and we place two points on the first line at a distance of b, for base. If those two parallel lines are separated by a distance of h, then no matter where we put the third point on the second line to create a triangle, the area will always be the same: $\frac{1}{2}bh$. Let's look at a triangle with a base of length b and a height of length h. Let's now break that triangle up into two smaller triangles. The triangle on the left will have area a_1 and the triangle on the right will have area a_2.

The triangle on the left is a right triangle, and we know that its area is $\frac{1}{2}bh$. If we split the base into two parts, one part of length b_1 and the other of length b_2, then the area of the triangle on the left is $\frac{1}{2}b_1(h)$ and the area of the triangle on the right is $\frac{1}{2}b_2(h)$. Therefore, the total area is $\frac{1}{2}b_1(h) + \frac{1}{2}b_2(h)$, which is equal to $\frac{1}{2}h(b_1 + b_2)$, but $b_1 + b_2$ was equal to b, the length of the original base. Thus, the total area of that triangle is $\frac{1}{2}bh$.

For a triangle with an obtuse base angle, we have a different proof. Rather than breaking this triangle into two, we'll extend the line that has length $b_1 + b_2$ to create a larger right triangle. We know that the area of a right triangle is $\frac{1}{2}bh$. The length of the base for this triangle is $(b_1 + b_2)$; the area of this triangle, then, is $\frac{1}{2}(b_1 + b_2)h$. The area of the new triangle we created, denoted a_2, is $\frac{1}{2}b_2(h)$ because it's a right triangle. The area of the original triangle, a_1, is what we get when we subtract the triangle with area a_2 from the larger triangle. That's equal to $\frac{1}{2}(b_1 + b_2)h - \frac{1}{2}(b_2)h$. Algebraically, the quantities $\frac{1}{2}(b_2)h$ cancel, and we're left with $\frac{1}{2}(b_1)h$.

The Pythagorean theorem states that given a right triangle with side lengths a and b and hypotenuse length c, $a^2 + b^2 = c^2$. Note that the *hypotenuse* is the length of the side that is opposite the right angle. Imagine that we adjoined four right triangles, each with side lengths a and b and hypotenuse c, to form a square with side lengths $a + b$. In the middle of that square is another square whose sides all have length c. One way we know that we have a square in the middle is by the symmetry of the object we've created: We see that all the angles of this four-sided figure must be equal and, thus, must sum to 360°. All of those angles, then, must be 90°, or right angles.

Let's start again with a picture of a big square with a little square in the middle. The area of the larger square is $(a + b)^2$. We can also compute the area of the big square by finding the areas of the triangles plus the area of the square in the middle. The area of each of the right triangles is 1/2(*ab*); together, those areas add up to $4 \times 1/2(ab)$, plus the area of the square in the middle, or c^2. When we set these two quantities equal to each other, we get: $a^2 + 2ab + b^2 = 2ab + c^2$. The $2ab$'s cancel, leaving $a^2 + b^2 = c^2$.

We can use the same theorem to find the length of any line segment. Suppose we want to determine the length of a line segment between point (0,0) and point (4,3). If we draw a right triangle that starts at (0,0) and ends at (4,3), we know that the length of that line segment will satisfy the Pythagorean theorem. That is, $4^2 + 3^2$ will be the length of that line segment squared. Because that length squared must be 25, then the length of the line must be 5. In general, we can show that starting with any point (0,0), we can find the length to the point (*a*,*b*) by drawing a right triangle; the length of the line segment from (0,0) to (*a*,*b*) will equal $\sqrt{a^2 + b^2}$.

Here's another formula that will come in handy later: To calculate the length of a line that connects any two points, (x_1,y_1) and (x_2,y_2), we can draw a right triangle that has a base of length $x_2 - x_1$ and a height of $y_2 - y_1$. According to the Pythagorean theorem, L^2, L being the length of the line from (x_1,y_1) to (x_2,y_2), is equal to $(x_2 - x_1)^2 + (y_2 - y_1)^2$. The length of the line would be the square root of that quantity.

Let's look at a problem that might seem a little challenging: We want to stretch a rope from the floor in one corner of the room to the ceiling in the opposite corner. What will the length of that rope have to be? Let's put the question in more mathematical terms: We want to calculate the length of a line from the point (0,0,0) in three dimensions to the point (*a*,*b*,*c*).

Instead of looking at the point (*a*,*b*,*c*), let's look at the point (*a*,*b*,0). We know from what we calculated earlier that the length of a line from point (0,0) to point (*a*,*b*) on the plane is $\sqrt{a^2 + b^2}$. Picture a triangle whose base runs across the floor, whose length on one side is $\sqrt{a^2 + b^2}$, and whose length on the other side is *c*. All we're calculating now is the hypotenuse of that triangle. The length that we're interested in must satisfy $L^2 = a^2 + b^2 + c^2$.

Taking the square root of both sides, we find that the length of the line from (0,0,0) to (a,b,c) is $\sqrt{a^2+b^2+c^2}$.

Let's do one more problem: Imagine we start off with two squares, each of length 1, and we put them next to each other. Each has length 1; therefore, they are 1-by-1 squares with an area of 1. Next, we'll build on that rectangle by adding other squares with dimensions that are Fibonacci numbers. What is the area of the resulting rectangle? The right side has length 8, and the top length is 8 + 5, or 13; thus the area is 8×13. We could also calculate the area of the rectangle by adding up the areas of all the individual squares. The sum of those areas is $1^2 + 1^2 + 2^2 + 3^2 + 5^2 + 8^2$, or 8×13, the product of Fibonacci numbers. We've now proved a pattern that we saw earlier; that is, the sum of the squares of the first n Fibonacci numbers is $F_n \times F_{n+1}$.

In our next lecture, we'll turn to the magical number pi. ■

Suggested Reading

Dunham, *Journey through Genius: The Great Theorems of Mathematics*, chapters 1–5.

Kiselev, *Kiselev's Geometry, Book 1: Planimetry*, trans. by Alexander Givental.

Questions to Consider

1. Show that the area of a parallelogram with horizontal base length b and vertical height h has area bh by rearranging a rectangle with the same dimensions.

2. Suppose that you tie a long rope to the bottom of the goalpost at one end of a football field. Then, you run it across the length of the field (120 yards) to a goalpost at the other end, stretch it tight, and tie it to the bottom of that goalpost so that it lies flat on the ground. Now suppose you add just 1 foot of slack to the rope so that you can lift it off the ground at the 50-yard line. How high can the rope be lifted up?

The Joy of Pi

Lecture 12

People have become so enthusiastic about π that people often with tongue-in-cheek—or maybe pi in their cheek—will celebrate π in some fun ways. For instance, I've taken part in many celebrations of π on what's called π day. And π day, because of the digits of π, is celebrated on March 14. That's 3/14 at 1:59, so you have 314159.

Let's begin this lecture on pi (π) by defining some terms. The *radius* (r) of a circle is the distance from the center of the circle to the edge of the circle. The *diameter* of a circle is the distance obtained by drawing a line from one side of a circle to the other side through the center of the circle. The diameter is twice the radius ($d = 2r$). The *circumference* is the distance around the outside of the circle.

Surprisingly, if we divide the circumference of any circle by its diameter, we always get the same number, the constant ratio called pi, (written with the Greek letter π), or about 3.14. Once we know the definition of *pi* (the ratio of the circumference to the diameter of any circle), we can calculate other quantities. For instance, the area of a circle is πr^2.

We can prove this theorem in two ways. Imagine that you have a circle in front of you and you cut through the top of the circle until you hit the radius. You then peel the circle away like an onion. You unwrap the first layer of the circle and lay it down flat; that layer then has a length of $2\pi r$. Then, you peel off the next layer. That next layer has a length that's a little bit less than $2\pi r$. You continue to peel off layers until you can't peel the circle any longer. Once you hit the center point at r, you have a triangle. We know that the area of a triangle is $1/2bh$. The base of the triangle has length $2\pi r$. The height of the triangle is r. Thus, the area of the triangle is $(1/2)(2\pi r)(r)$, which is πr^2.

Let's try another proof of this theorem. Imagine that you slice the circle up like a pizza into lots of triangles. Then, you separate the top half from the bottom half. The two sets of triangles can interlock to form a shape that's almost exactly a rectangle. The length of the bottom of the rectangle is πr

because it came from half of the circumference. The length of the top of the rectangle is πr and the side length is r. Thus, the area of the rectangle is $\pi r(r)$, which is πr^2.

These proofs don't tell us why pi should be the number 3.14…. Here's one way to get a handle on the size of the number pi. Let's look at a circle that has diameter 1. Remember that pi is the ratio of the circumference to the diameter; thus, if we have a circle with diameter 1, then pi will be the circumference of the circle. Next, we draw a square inside the circle. Do you agree that the perimeter of that square is less than the perimeter of the circle? If we can figure out the perimeter of that square, then we'll have a lower bound for the perimeter of the circle.

Surprisingly, if we divide the circumference of any circle by its diameter, we always get the same number, the constant ratio called pi.

Let's break that square, or diamond, into four right triangles and look at just one of those triangles. The triangle will have two sides whose lengths are 1/2 because the radius of the circle was 1. According to the Pythagorean theorem, the length of the hypotenuse, when squared, will be 1/2 + 1/2, or 1/4 + 1/4, or 2/4. We take $(1+\sqrt{5})/2$, or $\sqrt{2}/2$, to get the hypotenuse. Therefore, the perimeter of the diamond we drew in the middle of the circle is $4\left(\frac{\sqrt{2}}{2}\right)$, which is $2\sqrt{2}$, or about 2.828. We know, then, that the perimeter of the circle is larger than 2.828. Finding an upper bound for the size of pi is even easier. We now put the circle inside of a square. The diameter of the circle is 1, which means that it will fit inside of a 1-by-1 square. The perimeter of a square with side length 1 is 4. Thus, the circumference of the circle must be less than 4.

We now have a lower bound and an upper bound for pi. We showed that pi, whatever it is, is somewhere between 2.828 and 4. If we were to expand on this work, we could get better bounds for pi. In fact, the great mathematician Archimedes did so, using the same logic that we used, except instead of using

4-sided squares, he used 96-sided polygons to show that pi was between 3.1408 and 3.1428.

If we were to write pi out, it would be 3.141592653589…, going on forever. In 1761, Johann Heinrich Lambert proved that attempts to find pi exactly are futile. Pi is irrational, which means that it cannot be written as a fraction and its decimal expansion will never repeat.

We know that pi is connected to a circle, but what about other shapes? Let's look at an *ellipse*, which has the equation $x^2/a^2 + y^2/b^2$. In other words, the points (x,y) that lie on the ellipse satisfy this equation. In a drawing of an ellipse, the ellipse touches the x-axis when $x = a$ and when $x = -a$. The ellipse touches the y-axis when $y = b$ and $-b$. If we plug in $x = a$ and $y = 0$, we get $a^2/a^2 + 0^2/b^2$, which is 1. The area of an ellipse is πab. If a and b are equal, then the ellipse becomes a circle. If a is r and b is r, then the equation $x^2/a^2 + y^2/b^2 = 1$ becomes $x^2/r^2 + y^2/r^2 = 1$. When we multiply that by r^2, the equation becomes $x^2 + y^2 = r^2$, which is the formula for the area of a circle of radius r. In that case, the area of the circle would be $\pi(r)(r)$, or πr^2.

We also find pi in the volume of a cylinder that has a circular base of radius r and a height of h. Think of a can of soup. The base of the can has an area of πr^2 and it is then raised up to a height of h; obviously, we have to multiply h by πr^2 to get a volume of $\pi r^2 h$. To calculate the surface area, we have to calculate the area of the top and bottom of the can and the area that goes around the can. The areas of the top and bottom of the can are πr^2. If we were to unwrap the can and flatten it out, it would still have a height of h and its length would be the original circumference of the can, which is $2\pi r$. Thus, the area of the rectangle we get when we flatten out the can is $2\pi rh$. When we put it all together, the surface area is $2\pi r^2 + 2\pi rh$.

The volume of a right circular cone is $(\pi r^2 h)/3$. Think of an upside-down ice-cream cone. We have a circle of radius r on the bottom, and the cone goes straight up to a height of h, then down again to the circle. Exactly three of those cones could fit into a cylinder. The surface area of a right circular cone is $\pi r^2 + \sqrt{r^2 + h^2}$. The volume of a sphere of radius r is $(4\pi r)/3$, and the surface area of a sphere is $4\pi r^2$. Those are best derived using calculus.

Pi also appears in more unusual places. For example, the sum $1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\ldots$, gets closer to $\pi^2/6$ exactly. The sum $\frac{1}{1^4}+\frac{1}{2^4}+\frac{1}{3^4}+\ldots$, gets closer to $\pi^4/90$. Pi also intersects number theory. If we pick two enormous numbers, a and b, at random, the probability that the greatest common divisor of a and b is 1 is exactly $6/\pi^2$, about 60 %.

Another number that we saw earlier in our lectures was $n!$, the number of ways that we can arrange n objects. Believe it or not, $n!$ has an approximation that uses pi. Especially when n is large, this approximation is almost exactly equal to $(n/e)^n \sqrt{2\pi n}$.

The number e is about 2.71828. We'll see more about that number later.

Like phi, the golden ratio, pi has a continued fraction that goes on forever, shown at right. Pi even has a connection to the Fibonacci numbers, especially when we study trigonometry. Look at the formula below. As we add up more of those arc tangents and skip every other Fibonacci number, we get closer to $\pi/4$.

$$\pi = 3+\cfrac{1}{6+\cfrac{9}{6+\cfrac{25}{6+\cfrac{49}{6+\cfrac{81}{\cdots}}}}}$$

$$\frac{\pi}{4} = \tan^{-1}\left(\frac{1}{2}\right) + \tan^{-1}\left(\frac{1}{5}\right) + \tan^{-1}\left(\frac{1}{13}\right) + \tan^{-1}\left(\frac{1}{34}\right) \ldots.$$

When we talk about probability later on, we'll encounter the famous bell curve, which has a height of $1/\sqrt{2\pi}$.

Pi is often celebrated in fun ways. For instance, many people enjoy events on "Pi Day," which is celebrated on March 14 at 1:59, or 314159. The world record right now for memorizing pi is more than 40,000 digits. One way to memorize pi involves a paraphrase of Edgar Allen Poe's poem "The Raven" written by Mike Keith. In Keith's poem, the number of letters in each word equates to the digits of pi. You can also memorize the first 24 digits of pi using this sentence: "My turtle Pancho will, my love, pick up my new mover, Ginger." This sentence uses a phonetic code, in which every digit has an associated consonant sound, as shown in the table below.

1	t or d		6	j, ch, or sh
2	n		7	k or hard g
3	m		8	f or v
4	r		9	p or b
5	L		0	s or z

Note: the consonants for h, w, and y are not represented in this code. (A possible mnemonic is "Danny Marloshkovips.")

Look at the sentence about Pancho and the first five digits of pi (31415). By inserting vowel sounds, we turn 3 into the word *my*; then, for 1415, the t, r, t, and l sounds become *turtle*. Continuing this process, the sentence translates to the first 24 digits of pi. The next 17 digits correspond to "My movie monkey plays in a favorite bucket," and the next 19 digits match with "Ship my puppy Michael to Sullivan's back-rubber." If we want to go up to 100 digits, then the next 40 digits correspond to these two sentences: "A really open music video cheers Jenny F. Jones," followed by, "Have a baby fish knife so Marvin will marinate the goosechick." ■

Suggested Reading

Adrian, *The Pleasures of Pi, e and Other Interesting Numbers*.

Benjamin and Shermer, *Secrets of Mental Math*, chapter 7.

Blatner, *The Joy of Pi*.

Joy of Pi, www.joyofpi.com/.

Questions to Consider

1. Suppose you have a rope around the equator of a basketball. How much longer would you have to make the rope so that it is 1 foot from the surface of the basketball at all points? The answer is 2π feet. Now suppose you have the rope around the equator of the Earth. (Yes, a rope about 25,000 miles long!) How much longer would you have to make that rope so that it is 1 foot off the ground all the way around the equator?

2. Starting with the famous formula for the sum of squares of reciprocals: $1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\frac{1}{25}+\ ...\ = \frac{\pi^2}{6}$, derive a formula for the sum of the squares of the even reciprocals and the sum of the squares of the odd reciprocals.

The Joy of Trigonometry

Lecture 13

[Trigonometry] allows us to calculate areas and measurements often pertaining to triangles that would not be so easily done just using the standard techniques of geometry.

Trigonometry comes from the Greek *trigonometria*—literally, the measurement of triangles. It allows us to calculate measurements pertaining to triangles that we could not easily do using standard geometry techniques. All of trigonometry is based on two important functions known as the *sine function* and the *cosine function*. We will initially define these in terms of a right triangle.

We begin with a right triangle with one angle labeled a. The side that is opposite a is called the *opposite side*. The other side adjacent to a that isn't the hypotenuse is called the *adjacent side*. We define the sine of a (abbreviated as "sin a") to be the length of the opposite side divided by the length of the hypotenuse:

sine = opposite/hypotenuse.

The cosine of a (abbreviated as "cos a") is defined as the length of the adjacent side divided by the length of the hypotenuse:

cosine = adjacent/hypotenuse.

The third most commonly used trigonometric function is the tangent function, which is the sine divided by the cosine. Because sine is opposite/hypotenuse and cosine is adjacent/hypotenuse, the tangent of a (abbreviated as "tan a") is their quotient:

tangent = opposite/adjacent.

We can now calculate some trigonometric values. For instance, let's look at a classic right triangle with side lengths 3, 4, and (hypotenuse length) 5. If the side opposite angle a has length 4, then sin a = 3/5, cos a = 4/5, and tan a = 3/4. Note that the complementary angle to a has a measure of $90 - a$, an angle whose sine is 4/5, cosine is 3/5, and tangent is 4/3. It's no coincidence that the sine of the second angle is the cosine of the first angle, and the cosine of the second angle is the sine of the first angle. Those values come straight from the definition: $\sin(90 - a) = \cos a$, $\cos(90 - a) = \sin a$.

You should also be aware of three other trigonometric functions:

Function	**Reciprocal of function**	**Relationship**
secant	cosine	sec = 1/cos
cosecant	sine	csc = 1/sin
cotangent	tangent	cot = 1/tan

The definitions that we've looked at so far allow us to define the sine, cosine, and tangent only for angles between 0° and 90° because that's all we can fit in a right triangle. A more general view of trigonometric functions allows us to define these for any angle. We begin with the unit circle, which has a radius of 1. The unit circle has the equation $x^2 + y^2 = 1$. We draw an angle of measure a on the unit circle. Let's label the point that corresponds to angle a as (x, y). If we drop a line from (x, y) to the x-axis, we create a right triangle. We know that the length of the base of this triangle is x, the height is y, and the hypotenuse is 1.

What is cos a for that triangle? The adjacent side is length x and the hypotenuse side is length 1; thus, $\cos a = x/1 = x$. Similarly, $\sin a = y/1 = y$. If $\cos a = x$ and $\sin a = y$, then the original point on the circle that we called (x, y) is the point $(\cos a, \sin a)$. Thus, we will *define* the cosine and sine to be the point on the circle that corresponds to angle a. Note that the angle is measured counterclockwise from the x-axis.

Let's find the sine and cosine of 180°. Moving 180° from the x-axis, the x coordinate is -1 and the y coordinate is 0; therefore, $\cos 180 = -1$ and $\sin 180 = 0$. The angle $-a$ is a degrees *clockwise* from the x-axis. Its corresponding point on the unit circle will have the same *x*-coordinate as angle *a*, but the opposite y-coordinate. Thus, $\cos(-a) = \cos a$, and $\sin(-a) = -\sin a$. What happens if we add 360° to angle *a*? That literally takes us full circle; therefore, the cosine and sine will be exactly the same as they were before. That is, $\cos(a + 360) = \cos a$, and $\sin(a + 360) = \sin a$.

Recall that the unit circle is the set of points (x, y) that satisfies $x^2 + y^2 = 1$. Because $(\cos a, \sin a)$ is on the unit circle, that means that $(\cos a)^2 + (\sin a)^2 = 1$. This famous formula is usually written as: $\cos^2 a + \sin^2 a = 1$, or simply as $\cos^2 + \sin^2 = 1$. The box below shows some other angles for our trigonometric vocabulary.

> $\sin 0 = 0$, $\cos 0 = 1$. (That is, at 0°, $y = 0$, and $x = 1$.)
>
> $\sin 30 = 1/2$, $\cos 30 = \sqrt{3}/2$.
>
> $\sin 60 = \sqrt{3}/2$, $\cos 60 = 1/2$.
>
> $\sin 45 = \sqrt{2}/2$, $\cos 45 = \sqrt{2}/2$.
>
> $\sin 90 = 1$, $\cos 90 = 0$.

Notice that we don't have to memorize the tangents because they are simply the sine values divided by the cosine values. Note also that the *arc tangent* of 1 (the angle whose tangent is 1) is 45°. The *arc sine* of 1/2 (the angle whose sine is 1/2) is 30°.

Now we're ready to look at some problems. We see a right triangle with an angle of 30°. What are the lengths of the other two sides of this triangle? Let *b* be the length of the hypotenuse. Since $\sin 30 = 1/2$ and sin 30 (opposite/hypotenuse) $= 10/b$, then $10/b = ½$. Thus, $b = 20$. The length of the hypotenuse is 20. To find the length of the other side *a*, we'll use the Pythagorean theorem. We know that $b = 20$, and because we're dealing with a right triangle, we also know that $10^2 + a^2 = b^2$. We just saw that b^2 is 20^2, or 400, which tells us that a^2 is 300; therefore, $a = \sqrt{300}$, or $10\sqrt{3}$, or approximately 17.3.

We have a base of length 26, a side of length 21, and an angle of 15° between them. Can we find the area of this triangle? First, we'll draw a new line, splitting the triangle into two right triangles. The opposite here has height h and the hypotenuse has length 21; thus, sin 15 = h/21. Hence, h = 21 sin 15, and from our calculator sin 15 = .2588, so h is approximately 5.435. Knowing the height and the length of the base (given as 26), we can find the area of the triangle: ½bh, or ½ (26)(5.435) = 70.66.

We'll now prove one of the most difficult identities in basic trigonometry using a tool from our geometry lecture. (Keep in mind that you may have to go through this proof more than once before it sinks in.) This identity is as follows: $\cos (a - b) = \cos a \cos b + \sin a \sin b$. Here's the tool from geometry: For a line of length L that goes from point (x_1, y_1) to point (x_2, y_2), by the Pythagorean theorem, we showed that L^2 is equal to $(x_1 - x_2)^2 + (y_1 - y_2)^2$.

We start our proof by looking at the unit circle. Focus on the triangle whose vertices are the origin, the point (0, 0); the point (cos a, sin a); and the point (cos b, sin b). We know that two of the side lengths of that triangle are 1 because they are radii of the unit circle. We want to calculate the length of the line L that connects (cos a, sin a) to (cos b, sin b). From the L^2 formula, we see $L^2 = (\cos a - \cos b)^2 + (\sin a - \sin b)^2$. We next expand that equation. The first term expands to $\cos^2 a + \cos^2 b - 2\cos a \cos b$. The second term expands to $\sin^2 a + \sin^2 b - 2\sin a \sin b$. Simplifying, $\cos^2 a + \sin^2 a = 1$, and $\cos^2 b + \sin^2 b = 1$. The expression now reads: $2 - 2\cos a \cos b - 2\sin a \sin b$.

We now rotate the triangle so that the lower side is lying on the x-axis. Note that the lengths of the sides are still 1, and the length L hasn't changed either. The angle that we're looking at is angle a minus angle b, or $a - b$. What is the length of L? Look at the change of the x-coordinates and the change of the y-coordinates. Because the side of the triangle is lying on the x-axis and has a length of 1, that lower point is (1, 0); because the upper point of the triangle corresponds to angle $a - b$, it has coordinates $(\cos(a - b), \sin(a - b))$.

According to the L^2 formula, we add the change in x-coordinates squared and the change in y-coordinates squared: $(\cos(a - b) - 1)^2 + (\sin(a - b) - 0)^2$. When we expand that, we get: $\cos^2(a - b) + 1 - 2\cos(a - b) + \sin^2(a - b)$.

This equation is not as messy as it looks because $\cos^2 + \sin^2 = 1$. Thus, we have: $2 - 2\cos(a - b)$. Now, we have to equate the two expressions that we found for L^2: $2 - 2\cos(a - b) = 2 - 2\cos a \cos b - 2\sin a \sin b$. We divide everything by 2 to get the desired formula: $\cos(a - b) = \cos a \cos b + \sin a \sin b$. Once we have that equation, we can prove many useful identities. (Any truth in trigonometry is typically called a *trigonometric identity*.)

For instance, look what happens when we set $a = 90°$: $\cos(90 - b) = \cos 90 \cos b + \sin 90 \sin b$. But if you memorize $\cos 90 = 0$ and $\sin 90 = 1$, that equation simplifies to: $\cos(90 - b) = \sin b$. We can calculate $\sin(90 - a)$, which is $\cos(90 - (90 - a))$, $= \cos a$. This shows that those formulas are true for *any* angle—not just for angles between 0 and 90 degrees. We have a formula for $\cos(a - b)$, but what about $\cos(a + b)$? We simply replace b with $-b$, so that the formula reads $\cos(a - (-b)) = \cos(a)\cos(-b) + \sin(a)\sin(-b)$. But $\cos(-b)$ is the same as $\cos b$, and $\sin(-b)$ is the negative of $\sin b$. When we plug those in, we get the equation: $\cos(a + b) = \cos a \cos b - \sin a \sin b$. When a and b are the same angle, we have the *double-angle formula*: $\cos(2a) = \cos^2 a - \sin^2 a$. We can do similar calculations with the sine function and show that $\sin(a + b) = \sin a \cos b + \cos a \sin b$. In particular, when a and b are equal, this formula says that $\sin 2a = 2\sin a \cos a$.

Instead of using degrees that go from 0 to 360, mathematicians use a measurement called *radians*, in which $360° = 2\pi$ radians. Hence 1 radian is $360/2\pi$ degrees, approximately 57°. Because the graphs of trigonometric functions come from the unit circle, they have a nice periodic property. The sine and cosine functions can be combined to model almost any function that goes up and down in a periodic way, such as seasons, sound waves, and heartbeats.

We'll close with the *law of sines* and the *law of cosines*. For any triangle, with angles A, B, C, and corresponding side lengths a, b, c:

law of sines	$(\sin A)/a = (\sin B)/b = (\sin C)/c$
law of cosines	$c^2 = a^2 + b^2 - 2ab \cos C$

The law of cosines can be thought of as a generalization of the Pythagorean theorem. With the law of cosines, we can find the length of a missing side, *C*, in a given triangle. In our previous example, the remaining side had length *c*, which satisfies $c^2 = 26^2 + 21^2 - 2(26)(21)\cos 15$. Since $\cos 15 = .9659$, we get $c^2 = 62.2$, or *c* is approximately 7.89. ■

Suggested Reading

Gelfand and Saul, *Trigonometry*.

Maor, *Trigonometric Delights*.

Questions to Consider

1. Although it is useful to memorize the values of sine and cosine for 0°, 30°, 45°, 60°, and 90°, they can be easily derived from basic geometry. Try to do so. Once you know these values, then you can derive exact values for many other angles, as well. Use the double-angle formula to determine the exact value of the sine, cosine, and tangent of 15°.

2. Prove the law of sines, which states that for any triangle with angles *A*, *B*, *C*, and corresponding side lengths *a*, *b*, *c*:

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}.$$

Hint: To prove the first equality, draw a perpendicular line from vertex *A* to the line *BC*. Now compute sin(*A*) and sin(*B*) and compare your answers.

The Joy of the Imaginary Number i

Lecture 14

We saw that real numbers live on the real line, the *x*-axis. Where do complex numbers live? Now we have to start thinking two-dimensionally. Complex numbers live on the plane, what's called the complex plane.

Let's begin by thinking a bit about negative numbers. The ancient Greeks refused to accept the existence of negative numbers, but when we think about numbers on a number line, we can readily understand the concept of negative numbers. We also know how to add, subtract, multiply, and divide negative numbers. In the real world, negative numbers don't have square roots, but let's imagine that they do. Further, let's conduct a thought experiment in which we suppose that the number i satisfies $i^2 = -1$. What else can we learn about this imaginary number i?

If we want the usual laws of arithmetic to work, such as the commutative and associative laws, then we could combine $2i(2i)$ to get $4i^2$. But if we agree that $i^2 = -1$, then $2i(2i)$ would be -4. If we multiply $-2i(-2i)$, we get $+4$. If we multiply $2i(3i)$, then we get $6i^2$, or -6. If we multiply $2i(-3i)$, we get $-6i^2$, and because $i^2 = -1$, we would end up with $+6$. If we multiply an imaginary number, such as $2i$, by a real number, such as 3, we get $6i$.

What can we say about division with imaginary numbers? For a problem such as $6i \div 2i$, we simply solve as we would using algebra; that is, the i's would cancel, and the answer would be 3. As long as we don't divide by 0, we don't get into trouble. The solution to $i \div i$ is 1. How about $1/i$? What would be the reciprocal of this imaginary number? We multiply that number by 1, but we write 1 as the fraction i/i. When we do that, we get i/i^2; i^2 is still -1 and $i \div (-1)$ gives us $-i$.

We can also do the problem $1/2i$ by just multiplying fractions: $1/2(1/i)$, or $1/2(-i)$, which is $-i/2$. Addition and subtraction with imaginary numbers are also easy. For example, $3i + 2i = 5i$; $3i - 2i = 1i$, or i; $2i - 3i = -1i$, or $-i$. How about $2 + i$? The answer is just $2 + i$. A number that's of the form $a + bi$ is called *complex*. A number such as $4i$ is also a complex number, but it's called an *imaginary number*, because the "real part" is 0. Even a number such as 7 is a complex number, but it also happens to be a real number; thus, 7 can be thought of as $7 + 0i$. (Another rule of arithmetic for imaginary numbers is $0i = 0$.)

Let's look at some arithmetic with complex numbers. Sample problems are shown in the table below.

Addition	$(2 + 5i) + (5 + 3i) = 7 + 8i$
Subtraction	$(2 + 5i) - (5 + 3i) = -3 + 2i$
Multiplication	$(2 + 5i)(3 + 4i) = 6 + 15i + 8i - 20 = -14 + 23i$
Division	$\frac{(2+5i)}{(3+4i)} \times \frac{(3-4i)}{(3-4i)} = \frac{(26+7i)}{(3^2+4^2)} = \frac{(26+7i)}{25}$ $(2 + 5i)/(3 + 4i)$

If we add $(2 + 5i) + (5 + 3i)$, the answer is $7 + 8i$. How about $(2 + 5i) - (5 + 3i)$? We subtract the real part, $(2 - 5) = -3$, then we subtract the imaginary part, $(5i - 3i) = 2i$; thus, we get $-3 + 2i$. When we multiply imaginary numbers, we use FOIL, just as we did with polynomials. For the problem $(2 + 5i)(3 + 4i)$, we get: $2(3) = 6$, $5i(3) = 15i$, $2(4i) = 8i$, and $5i(4i) = 20i^2$. The only new idea here is that we can use the fact that $i^2 = -1$. Thus, we have $6 + 15i + 8i - 20$, which simplifies to $-14 + 23i$.

Conjugates help make the division of complex numbers easier.

Here's another problem: $(3 + 4i)(3 - 4i)$. (The second quantity is called the *conjugate* of the first quantity; that is, $a + bi$ has the conjugate $a - bi$. When we use FOIL, we get $9 + 12i - 12i - 16i^2$, which simplifies to $9 - 16(-1)$, $9 + 16 = 25$. Conjugates help make the division of complex numbers easier. For instance, notice that if we multiply any number of the form $a + bi$ by its conjugate, $a - bi$, we get $a^2 - b^2i^2$, but the $-b^2i^2$ becomes $+b^2$; thus, $(a + bi)(a - bi) = a^2 + b^2$.

If we want to find the reciprocal of $a + bi$, $1/a + bi$, we simply multiply both the top and the bottom by the conjugate $a - bi$, as shown below.

$$\left(\frac{1}{a+bi}\right)\left(\frac{a-bi}{a-bi}\right)=\frac{a-bi}{a^2+b^2}$$

Notice that the denominator of this fraction is a real number. In this way, we never have to have a complex number in the denominator; we can always eliminate complex numbers by multiplying the numerator and the denominator by the conjugate of the denominator. For example, to divide $(2 + 5i)/(3 + 4i)$, we simply multiply the top and the bottom by $3 - 4i$. When we do that multiplication, we get $6 + 15i - 8i - 20i^2 = 26 + 7i$ in the numerator and $3^2 + 4^2 = 25$ in the denominator. That is, $(2 + 5i)/(3 + 4i) = (26 + 7i)/25$.

We saw that real numbers exist on the real line, but complex numbers exist on what's called the *complex plane*. Think of the x-axis and the y-axis, as we've been using in geometry. For instance, the number $1 + i$ will have a "real part" of 1, an x-coordinate of 1, and a y-coordinate (called i) of 1. To find $1 + i$, then, we go to the right 1 and up 1. Think of the y-axis as having the number 0 where the two axes meet. As we go up, we see $1i$, $2i$, $3i$,…, and as we go down the imaginary axis, or y-axis, we see $-i$, $-2i$, $-3i$,….

Let's do a few more examples: For $2 + 2i$, we go to the right 2 and up 2. For $-2 + i$, we go to the left 2 and up 1. For $-3 - 2i$, we go to the left 3 and down 2. For $2 + i$, we go to the right 2 and up 1. What happens if we multiply $2 + i$ by 1/2? Then, we'd have the number $1 + 1/2i$, which would be halfway along the line from 0 to the point $2 + i$. Thus, when we multiply by 1/2, the length of the line changes by 1/2. Similarly, if we multiply $2 + i$ by 2, we get $4 + 2i$. The line, or *vector*, that goes from the origin, the point $0 + 0i$, to the point $4 + 2i$ will be twice as long as it was before. When we multiply a complex number by a real number, the line expands by a factor of that real number. If we multiply by a positive number, the line still points in the same direction. If we multiply by a negative number, the line points in the opposite direction.

We can "see" how to add two complex numbers, such as $a + bi$ and $c + di$, by looking at their pictures on the complex plane. We see a line that goes from 0 to $a + bi$ and a line that goes from 0 to $c + di$. Those two lines can form the sides of a parallelogram. The top of the parallelogram is the point at which the sum of $a + bi$ and $c + di$ meet. In other words, we start at $a + bi$, then add the vector that goes to $c + di$ to get the sum.

Look again at the line that goes from 0 to the point $a + bi$. We can define the length of that line to be the length of the complex number. The base of the triangle we see has length a and the height has length b. By the Pythagorean theorem, the hypotenuse of this right triangle will have length $\sqrt{a^2 + b^2}$; we define that to be the length of the complex number $a + bi$. The angle near the origin of this triangle would be the angle associated with the complex number. That angle is measured counterclockwise from the x-axis.

We can "see" how to multiply complex numbers in much the same way. In this case, we use two simple rules: Multiply the lengths from the origin and add the angles. We see, for example, if a complex number $a + bi$ has with an angle of about 30° and length 4 and if the point $c + di$ has an angle of 120° and length 2, to obtain their product, we simply multiply the lengths ($4 \times 2 = 8$) and add the angles ($30° + 120° = 150°$). Hence, the product will be

$$8(\cos 150 + i \sin 150) = \left(\frac{-\sqrt{3}}{2} + i\left(\frac{1}{2}\right) \right) = -4\sqrt{3} + \frac{i}{2}.$$

To summarize, we can add two points, such as $(a + bi)$ and $(c + di)$, by drawing a parallelogram. We can multiply those points by multiplying the lengths and adding the angles.

Here's another example: $(2 + 2i)(-5 + 5i)$. What's the length of the line from the origin to $2 + 2i$? The length of $a + bi$ is $\sqrt{a^2 + b^2}$; thus, the length of $2 + 2i$ will be $\sqrt{2^2 + 2^2}$, or $\sqrt{8}$. The length of $5 + 5i$ will be $\sqrt{5^2 + 5^2}$, or $\sqrt{50}$. When we multiply those lengths together, we get $\sqrt{400} = 20$. The angle that cuts the first quadrant exactly in half at the point 2 + 2*i* is 45°. The angle for −5 + 5*i* is 135°. When we add those angles together, we get 180°.

Incidentally, as we mentioned in the trigonometry lecture, a mathematician would call the measure of the first angle π/4 radians, instead of 45°. The second angle would be 3π/4 radians, instead of 135°. Adding those together, we get 4π/4 radians, or π radians. Returning to the problem, when we multiply those numbers together, we get something that has a length of 20 and an angle of 180°. But 180° means 180° from the origin. Thus, we have a length of 20 pointing in the negative direction, or −20, as the answer.

Why does this rule of multiplying the lengths and adding the angles work? Once again, Euler gives us the equation for this: $e^{i\theta} = \cos\theta + i\sin\theta$ (*e* is a special number that we'll talk about later). Look at the unit circle again. Euler says that we can simplify the point on the unit circle at angle θ can be called $e^{i\theta}$. Note that we would normally call that point cos θ, sin θ if we were in the *x*, *y* plane, but in the complex plane, we call it $\cos\theta + i\sin\theta$, and we can simplify that to $e^{i\theta}$.

Any complex number on the unit circle is of the form $e^{i\theta}$. We can even get beyond the unit circle—that is, a point that has angle θ but has length *R*, represented by $Re^{i\theta}$. For example, the number 2 + 2*i* has a length of $\sqrt{8}$ and an angle of π/4 radians. We can write this in *polar form* by saying

$2 + 2i = \sqrt{8}e^{i\pi/4}$. What if we were to stay on the unit circle and move 90°, or π/2 radians? We then find ourselves at the point *i*, that is, $e^{i\pi/2}$. What happens when we multiply complex numbers? If we write those numbers in polar form—let's say our first number was $R_1e^{i\theta_1}$ and our second number was $R_2e^{i\theta_2}$—then when we multiply those, we get $R_1R_2e^{i(\theta_1+\theta_2)}$. We're just using the laws of arithmetic and the law of exponents. The result, $R_1R_2e^{i(\theta_1+\theta_2)}$, tells us to do exactly what our two simple rules say, namely, multiply the lengths and add the angles.

What would Euler say about (*i*)(*i*)? We said that $i = (e^{i\pi/2})(e^{i\pi/2})$; that would give us $e^{i\pi}$, but we also know that $(i)(i) = -1$; thus, $e^{i\pi} = -1$. That says that e^i multiplied by the angle of π radians (that's 180°) puts us at the real number −1. If we rearrange that equation, it becomes: $e^{i\pi} + 1 = 0$, one simple equation that contains the five most important numbers and some of the most important relations in mathematics. What this "profound" equation says is simply that if you move 180° along the unit circle, you wind up at −1. We can use Euler's equation to derive many complicated trigonometric identities. For example, we know $e^{i(2\theta)} = \cos(2\theta) + i\sin(2\theta)$. But it's also true that $e^{i(2\theta)} = e^{i\theta}\, e^{i\theta} = (\cos\theta + i\sin\theta)^2 = (\cos^2\theta - \sin^2\theta) + i(2\sin\theta\ \cos\theta)$. Comparing the real and imaginary parts gives us $\cos(2\theta) = \cos^2\theta - \sin^2\theta$, and $\sin(2\theta) = 2\sin\theta\cos\theta$.

Complex numbers can also help us with algebra. For instance, without complex numbers, we could not find a solution to the equation $x^2 + 1 = 0$. We know, however, that this equation has at least one solution, namely, i, because $i^2 = -1 + 1 = 0$. We can also find another solution because $(-i)^2$ is also −1. Similarly, the equation $x^2 + 9 = 0$ has two solutions, namely, $3i$ and $-3i$, as does the equation $x^2 + 7 = 0$, which has solutions $\sqrt{7}i$ and $-\sqrt{7}i$. With a more complicated algebraic expression, such as $x^2 + 2x + 5 = 0$, we use the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Plugging into that formula, we get the result shown below:

$$x = \frac{-2 \pm \sqrt{2^2 - 4(1)(5)}}{2} = \frac{-2 \pm \sqrt{-16}}{2} = \frac{-2 \pm 4i}{2} = -1 \pm 2i.$$

Earlier, we discussed the fundamental theorem of algebra, but we can express that in a more polished way using complex numbers. The fundamental theorem of algebra says that if $P(x)$ is a polynomial with real or complex coefficients, then we can always factor it in the form $P(x) = (x - r_1)(x - r_2)\ldots(x - r_n)$, where the roots $r_{1,\ \ldots,}\ r_n$ are complex numbers. We could factor any n^{th}-degree polynomial into these n parts. We said earlier that the polynomial equation $P(x) = 0$ has, at most, n real solutions. In fact, we can say that it has "sort of exactly" n solutions, namely, r_1, r_2, r_3, $\ldots r_n$. (We say "sort of exactly" because it's possible that some of the roots were repeated.)

In this lecture, we've defined imaginary numbers; seen how to add, subtract, multiply, and divide them; and seen how to use them algebraically and geometrically. We've also seen Euler's equation and some of its applications. ■

Suggested Reading

Cuoco, *Mathematical Connections: A Companion for Teachers and Others*, chapter 3.

Nahin, *An Imaginary Tale: The Story of* $\sqrt{-1}$.

Questions to Consider

1. Once you overcome the obstacle of imagining $\sqrt{-1}$, it's easy to imagine the square root of any complex number. For instance, can you find two numbers with a square of i? (Hint: They will both lie on the unit circle.)

2. In general, for every positive integer n, every nonzero complex number has exactly n distinct n^{th} roots. For instance, can you describe all of the n^{th} roots of 1? Express your answer in terms $e^{i\theta}$ and plot the points on the unit circle. Prove that the sum of these roots is always 0.

The Joy of the Number *e*

Lecture 15

Where did the number *e* come from? The number *e* was first used by Isaac Newton, but it was really studied and analyzed and actually named by the great Swiss mathematician, Leonhard Euler. In fact, Euler, I believe, named *e* after himself. He put it as a small *e*. He was being modest, but I think of the so many things he accomplished and discovered, I think he was very proud of this number *e*, and *e* has just so many amazing, amazing uses.

Let's begin by creating *e*. We start with $(1 + 1/10)^{10}$. The result is 2.593…. Next, we look at$(1 + 1/100)^{100}$. We're doing two things here: We're making the base closer to 1, and we're making the exponent much bigger. The result for the second equation is 2.70481…. Let's try again: $(1 + 1/1000)^{1000}$. That result is 2.71692…, still close to 2.7. In fact, as we take this process farther and farther out, as n gets larger and larger, $(1 + 1/n)^n$ gets closer and closer to the magical number e: 2.718281828459…. In mathematical terms, as n goes to infinity, e is the *limit* of $(1 + 1/n)^n$. We can generalize $(1 + 1/n)^n$: For any number x, if we take the limit as n goes to infinity of $(1 + x/n)^n$ we get $e^{x.}$

The number e relates to compound interest. Suppose you put \$1,000 in a bank account that earns 6% each year. After one year, how much money will you have? We can find the answer by multiplying \$1,000 by 1.06, which gives us \$1,060. Assuming that you didn't take the interest out of your account, after two years, you'll have \$1,000 × 1.06 × 1.06, or \$1,123.60. After three years, you'll have $\$1{,}000 \times 1.06^3$ = about \$1,191.02. After t years, you'll have $\$1{,}000 \times 1.06^t$. Let's focus on one year and suppose that instead of being compounded annually, the interest was compounded semiannually. Instead of giving you a lump sum of 6% at the end of the year, the bank gives you 3% after six months and another 3% when the year ends. That's equal to $\$1{,}000(1.03)^2$, or \$1,060.90.

Suppose that your interest was compounded quarterly. That means that every three months, you'll get 1.5% interest. You figure the interest by $1,000(1.015)$^4^. If the interest is compounded monthly, you get 0.5% each month: $1,000(1.0005)$^12^. If the interest is compounded daily, you figure the interest by: $1,000(1 + .06/365)$^365^, which is $1,061.83. If the bank compounds the interest *continuously*, your interest rate will be 6%/n per time period. With $1,000, you'll get

$$\$1,000\left(1+\frac{.06}{n}\right)^n.$$

As we know from the formula we found for e^x, if we raise $(1 + .06/n)$ to the n^{th} power, as n gets larger and larger, we get closer and closer to $e^{.06}$. When we calculate $1{,}000 \times e^{.06}$, we get $1061.84. Thus, with interest compounded continuously instead of daily, you earn an extra penny. But we also have a simpler equation than we had before. The general formula for 6% interest compounded continuously at the end of one year for $1,000 is: $1{,}000(e^{.06})$. For t years, the formula is: $1000e^{.06t}$. Starting with a principal amount p and an interest rate r, after t years with continuous compounding, the general formula for interest is: pe^{rt}.

Let's do another application with e, this one involving homework. My students have turned in a number of homework assignments, but I don't want to grade them. I randomly return the homework to my students for grading, but I don't want any student to be in the position of grading his or her own paper. My question is: How likely is it that nobody gets his or her own paper? Suppose I have three students, A, B, and C. In how many ways can I return their homework papers? We know from our earlier lectures that there are $3! = 6$ ways of returning three homework papers, but only two out of the six ways result in no student getting his or her own homework back. Thus, if I randomly return the homework, the chance that no student gets his or her own homework is 2 out of 6.

If I have four students, then there are 4! = 24 ways of returning the homework papers. Of those 24 ways, only 9 result in no student getting his or her own homework back. The chances that no one gets his or her own homework are 9 out of 24, or 3/8, or .375. If we look at the chances with five students, six students, and so on, we see that the results get closer and closer to the same number. With five students, the chance is .366; with six students, it's about .368; with 100 students, it's .3678….

Those results are strange. Whether I'm returning 100 papers back to 100 students, or 10 papers back to 10 students, or 1,000,000 papers back to 1,000,000 students, the chance that nobody gets his or her own homework is practically .368. This magic number .368 is $1/e$; it's the reciprocal of e, 2.71828. Why should this be? If I have n students in the classroom, the chance that the first student will get his or her own homework is $1/n$, and the chance is the same for the second student and so on. The chance that you won't get your own homework back is $1 - 1/n$; therefore, the chances that no student gets his or her own homework are approximately $(1 - 1/n)^n$. Our earlier formula said that $(1 + x/n)^n$ approaches e^x as n gets large. That's the situation we have here except that x is -1. That is, we have $(1 + -1/n)^n$; as n goes to infinity, that result is e^{-1}, or $1/e$.

Historically, logarithms were useful for converting difficult multiplication problems into more straightforward addition problems.

How does the function e^t grow? Looking at the graph of that function, we see that e^t grows fairly quickly; e^{2t} grows faster, and e^{3t} grows even faster. These are called *exponential functions*. Let's look at the function 5^t. The number 5 is between e, 2.718, and e^2, which is about 7.389. That means that 5^t is between e^t and e^{2t}; therefore, 5 is equal to e raised to some power between 1 and 2. Let's say that 5 is e^r, in which r is some real number between 1 and 2. That means that we can replace 5 in the function 5^t with the number e^r raised to a power of t. Thus, 5^t is the same as $(e^r)^t$. By the law of exponents, that's e^{rt}. To find the number r in this expression, we need to look at logarithms.

Logarithms are based on, initially, the powers of 10: $10^0 = 1$, $10^1 = 10$, $10^2 = 100$, $10^3 = 1{,}000$, and so on. Negatively, $10^{-1} = 1/10$, $10^{-2} = 1/100$, and $10^{-3} = 1/1{,}000$. We say that the logarithm of x, denoted $\log x$, solves the equation $10^{\log x} = x$. The logarithm of x is the exponent to which we have to raise 10 in order to get x. For example, $\log 1{,}000 = 3$ because $10^3 = 1{,}000$. Log $100 = 2$ because $10^2 = 100$. Log $10^y = y$ because we raise 10 to a power of y to get 10^y.

Can we find $\log \sqrt{10}$? The result for $\sqrt{10}$ is $10^{1/2}$; thus, $\log \sqrt{10}$ is 1/2. What is log 512? A calculator tells us that log 512 = about 2.709. Does that seem reasonable? We know that log 100 = 2 and log 1,000 = 3. Because 512 is between 100 and 1,000, it follows that log 512 should also be between log 100 and log 1,000, or between 2 and 3. There are other useful rules for logarithms. For instance, we've said that $\log 10^x = x$ for any x. Another sensible rule is $10^{\log x} = x$. Again, if we think about the definition of log, that makes sense.

Perhaps the most commonly used property of the logarithm is the one that states: The log of the product is the sum of the logs: $\log(xy) = \log x + \log y$. Look at the expression $10^{\log x + \log y}$. According to the law of exponents, $10^{a+b} = 10^a \times 10^b$; thus, the expression would equal $10^{\log x} \times 10^{\log y}$. We know, however, that $10^{\log x}$ is x and $10^{\log y}$ is y, so that gives us $x \times y$. On the other hand, we know from our useful log rule that $10^{\log(xy)}$ is also equal to xy. What have we done here? We've taken 10 to some power and obtained xy. We then took 10 to another power and obtained xy; therefore, the two powers must be equal. Equating these powers tells us that $\log x + \log y$ must equal $\log xy$.

As a corollary to that last rule, we can also show what I call the exponent rule: $\log(x^n) = n \log x$. Let's look at a couple of examples. Historically, logarithms were useful for converting difficult multiplication problems into more straightforward addition problems. Let's illustrate the product rule and exponent rule for logarithms. If log 2 =.301… and log 3 = .477…, then $\log 6 = \log(2 \times 3) = \log 2 + \log 3 = .301\ldots + .477\ldots = .778\ldots$. Can we find log 5 knowing log 2 and log 3? We don't need to use log 3 in this solution, but we do need to use log 10, which is 1; thus, $\log 5 = \log(10 \times 1/2)$, or $\log 10 + \log 1/2$, and we know log 1/2 because 1/2 is 2^{-1}. We now have $\log 10 + \log 2^{-1}$, but by the exponent rule, $\log 2^{-1}$ is $-1 \times \log 2$. This is equal to 1 − log 2, or 1 − .301, or about .699. Earlier in this lecture, we looked at log 512. Note that 512 is 2^9. Log 2^9, by the law of log exponents, is equal to $9 \times \log 2$. Because log 2 is .301, that gives us 2.709, as we saw earlier.

We've been talking about logarithms using base 10, but we can also use logarithms in other bases. We define $\log_b x$ to be the exponent that solves $b^{\log_b x}$. For instance, as we noted above 2^9 is 512; thus, the log(base 2) of 512 is 9 because we have to raise 2 to the 9th power to get 512. The rules for logarithms in other bases are, in fact, virtually unchanged from the rules for base 10: $\log_b b^x = x$, $\log_b(xy) = \log_b x + \log_b y$, and $\log_b(x^n) = n\log_b x$. We can also change from one base to any other base: Log(base *b*) *x* is log $x \div \log b$, where that log could be the log(base 10) or any other base. In chemistry and the physical sciences, the base 10 logarithm is probably the most popular. In computer science, base 2 is the most popular log. But in math, physics, and engineering, by far, the most popular base of the logarithm is the log base *e*, the natural log. ■

Suggested Reading

Adrian, *The Pleasures of Pi, e and Other Interesting Numbers*.

Maor, *e: The Story of a Number*.

Questions to Consider

1. With $10,000 in a savings account earning 3% interest each year, compounded continuously, about how much money will be in the account after 10 years?

2. Starting with the famous formula for *e*: 1 + 1/1! + 1/2! + 1/3! + 1/4! + ... = *e*, determine the following sums:

1/1! + 2/2! + 3/3! + 4/4! + 5/5! + ...
1 + 3/2! + 5/4! + 7/6! + 9/8! + 11/10! + ...
1/1! + 2/3! + 3/5! + 4/7! + 5/9! + 6/11! + ...

The Joy of Infinity

Lecture 16

Let's start with the question, is infinity a number? Technically, it's not. It's treated as if it's a concept that's something that's larger than any number. Technically, of course, there is no largest number because if you thought you found one, then you could add 1 to it, and you'd have an even larger number. But, it's treated as something that's bigger than any other number, but it is itself not a number.

As we go to the right on the number line, we're *approaching* infinity. Sometimes, though, we do treat infinity as a number, represented by the symbol ∞. For instance, we might say that adding all the positive numbers equals infinity, although most mathematicians would say that the sum *goes to* infinity. For the sum to go to infinity means that it will be larger than any number you ask for—larger than a million, a trillion, even a googol.

Though it doesn't get as much attention, the cousin of infinity is negative infinity, denoted by $-\infty$. The sum of all negative numbers gets smaller than any negative number you could ask for. As a mathematical convenience, we make statements such as 1/infinity = 0. That makes sense because if we divide 1 by bigger and bigger numbers, then the quotient gets closer and closer to 0. We can even say $1/-\infty = 0$ because if we divide 1 by negative million, or negative billion, etc., the result gets closer to 0.

On the other hand, we are never allowed to divide by 0; thus, we couldn't say $1/0 = \infty$. The real reason we don't allow that is because 1/0 could be infinity or negative infinity. If we divide 1 by a tiny positive number, the answer will be a big positive number. If we divide 1 by a tiny negative number, the answer will be a big negative number. In other words, as our denominator gets closer to 0 from the right, we're going to infinity; as it gets closer to 0 from the left, we're going to negative infinity. That's why we let 1/0 be undefined. There are some infinite sums that add to something besides infinity. For instance, 1 + 1/2 + 1/4 + 1/8 + 1/16 + … = 2.

Also, 1 + 1/1! + 1/2! + 1/3! +… = *e*. We don't necessarily get infinity as our answer even if we have an infinite sum.

In this lecture, rather than using infinity as a number-like object, we will use it as a size. The size of a set (or *cardinality* of a set) *S* (denoted $|S|$) is the number of elements in the set. For instance, if *S* = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}, then $|S| = 10$. What's the size of the set of all positive integers? Because that set has infinitely many elements, its size is infinity; the size of the set of all even numbers is also infinity. What is the size of the set of all fractions? Because there are more fractions than there are integers, the size of that set is infinite as well. The size of the set of real numbers between 0 and 1 is also infinite. As we will see, however, some infinities are more infinite than others.

Though it doesn't get as much attention, the cousin of infinity is negative infinity.

Let's try a thought experiment. Suppose that every chair in my class is occupied by a student, and no students are chair-less. I could pair up students with chairs and conclude that there are as many students as there are chairs. This is called a *one-to-one correspondence*. We can use this same idea to compare the set of positive odd numbers with the set of positive even numbers. Not only are there an infinite number of both of those objects, but they have the same *order of infinity* because we can pair them up. Those sets, then, are infinite, and they have the same size.

What about the sizes of the sets of all positive integers (1, 2, 3, 4, 5, 6,…) and all positive even integers (2, 4, 6, 8, 10, 12,…)? I claim that that those two sets have the same size—not because they are infinite, but because we can pair them up. Here, 1 is paired with 2, 2 is paired with 4, 3 is paired with 6, and so on.

Mathematicians say that any set that can be paired up with the positive integers is *countable* because we could essentially list all the numbers in the set just by counting. For example, the set of all integers (positive, negative, and 0) is countable. It can be put in one-to-one correspondence with the positive integers because we can list them all with no infinite gaps. If we list the integers as 0, 1, −1, 2, −2, 3, −3, 4, −4, 5, −5…, eventually, we will reach every positive and every negative number. We can't, however, list the positive numbers first, then the negative numbers, because we'd never finish with the first step. These ideas were first put forth by the German mathematician Georg Cantor, but it took mathematicians decades to come to grips with them.

Now, let's look at a larger set of numbers, the set of fractions. Of course, there are an infinite number of positive fractions, but are they countable? We might try listing the fractions out row by row, but that would leave infinite gaps. If we list the fractions out diagonally, however, we see that the set of rational numbers is countable. It has the same size as the set of positive integers.

Can we find a set that is not countable? Surprisingly, the set of real numbers between 0 and 1 is uncountable. We can show this with a proof by contradiction. Suppose you begin your list with the number .31415926…; your second number is .12121212…; your third number is .500000; and your fourth number is .61803399…. I can use your list to create a real number that can't be on the list. I begin with your first number, .31415926…. I add 1 to the first digit of that number to change it to 4. Then, I add 1 to the second digit of your second number to change that digit to 3. I can also change the third digit of your third number, the fourth digit of your fourth number, and so on. In that way, I create a number that is not on your list.

Let's say I created the number .4311…. How do I know that number is not the millionth number on your list? It couldn't be, because it will have a different digit in the millionth digit past the decimal point. The number I've created, then, can't be the first, the second, or the millionth number on your list. Therefore any attempt to list the real numbers is doomed to failure: the list is guaranteed to be incomplete.

We know that the set of positive real numbers and the set of real numbers between 0 and 1 are both infinite, but the first set is countable and the second set is not. We now need to come up with different notations to represent these two different levels of infinity. We use the symbol $\aleph_0$ ("alef nought") to denote the size of the set of positive integrers. (The symbol alef is the first letter of the Hebrew alphabet.) Anything that can be put into correspondence with the positive integers, any countable set, has size, or cardinality $\aleph_0$. The set of real numbers between 0 and 1 has a greater level of infinity; mathematicians usually denote that level of infinity by the letter *c*, where *c* stands for continuum.

Can we find a set that is bigger than *c*? For example, is there twice as much "stuff" in the interval between 0 and 2 as there is in the interval between 0 and 1? Both of these are infinite sets, but there is an elementary way to pair up the numbers between these two sets. Let's look at a triangle. Inside the triangle, we have a segment of length 1 and, at the base, we have a segment of length 2. At the top, we have a laser beam shooting down, connecting every point between 0 and 1 with another point between 0 and 2. We can pair every point in the first interval with a point in the second interval. What we're really looking at here is the function $y = 2x$. Every point on the *x*-axis is associated with a point on the *y*-axis by way of that function. This shows that the size of the set of real numbers between 0 and 2 is the same as the size of the set of real numbers between 0 and 1. In other words, both have size *c*.

What about the size of the set of all real numbers from negative infinity to positive infinity? Is that set bigger than the set between 0 and 1? As long as we draw any function that more or less increases from negative infinity to positive infinity, we can create a one-to-one correspondence. The function we see here is a trigonometric function: $y = \tan(\pi(x - 1/2))$. Between every number from 0 to 1, we can get every real number—positive, negative, and 0. In other words, the size of the set of real numbers is the same as the size of the set of real numbers between 0 and 1. Both still have size *c*.

Can we find a set that has a size bigger than *c*? Let's look at the plane—that's the set of points inside the unit square (side length = 1). If there are an infinite number of points between 0 and 1, there are certainly an infinite number of points in the square that is drawn from 0 to 1 horizontally and from 0 to 1 vertically. Amazingly, however, even this set can be put in one-to-one correspondence with the set of real numbers between 0 and 1. Let's say that x is $0.r_1r_3r_5r_7\ldots$, and y is $0.r_2r_4r_6r_8\ldots$. That's an ordered pair inside the unit square. We will associate that pair with the real number $0.r_1r_2r_3r_4r_5r_6\ldots$. If we start with, say, the point 0.31415926…, that pairs up with the ordered pair

0.3452… and .1196…. Any number between 0 and 1 can be turned into a pair of numbers between 0 and 1, and vice versa. To put it another way, the size of the set called $\mathbb{R}^2$ (the set of all pairs of real numbers; pronounced "R two") is *c*, where *c* stands for "continuum." The sizes of the sets of all triples, quadruples, and so on of real numbers are also *c*. We still haven't found a set that is bigger than the size of the set of real numbers.

There is such a larger set: the set of all curves in the plane. That is, there are more curves than there are real numbers to assign them. Here is another set whose size is bigger than *c*: the set of all subsets of real numbers. That is, there are more subsets of real numbers than there are real numbers to assign them.

In this lecture, we've shown that a set is infinite if the size of that set exceeds any given number. The sets of integers and rational numbers are countable because we can list them; these have a size called $\aleph_0$. The real numbers are uncountable and have size *c*. Finally, there are infinitely many levels of infinity. Here's a question to think about: Are those infinitely many levels of infinity countably infinite or uncountably infinite? ■

Suggested Reading

Burger and Starbird, *The Heart of Mathematics: An Invitation to Effective Thinking*, chapter 3.

Dunham, *Journey through Genius: The Great Theorems of Mathematics*, chapters 11−12.

Maor, *To Infinity and Beyond: A Cultural History of the Infinite*.

Questions to Consider

1. Prove that the number of irrational numbers between 0 and 1 is uncountable.

2. Imagine a red robot that produces 10 billiard balls at a time, numbered 1 though 10, then 11 through 20, then 21 through 30, and so on. Meanwhile, each time the red robot creates 10 balls, an evil green robot destroys a ball. In the first round, it destroys ball 10; in the second round, it destroys ball 20; in the third round, it destroys ball 30; and so on. At the end of the process, which balls remain? (Although this is an infinite process, we can imagine it happening in a finite amount of time. Imagine that round 1 occurs an hour before midnight, round 2 occurs half an hour before midnight, round 3 occurs a third of an hour before midnight, and so on. The challenge is to describe the situation "at midnight.")

3. Bonus question: now suppose instead that after the first round, the green robot destroys ball 1; after the second round, it destroys ball 2; after the third round, it destroys ball 3; and so on. At the end of this process, which balls remain?

The Joy of Infinite Series

Lecture 17

There's so many more things to say about infinite series. I could go on forever about infinite series. In fact, if I gave you a book to read on infinite series, followed by a smaller book to read on infinite series, plus another book, plus another book, plus another book, plus another book, do you know what you'd have? You'd have a book series, wouldn't you?

Let's look at several proofs of the bold statement .999999… = 1. Here's the most elementary proof: We agree that 1/3 = .333333333…. If we multiply $3 \times .333333\ldots$, we get .999999…. We also know that $3 \times .333333 = 3(1/3)$, but 3(1/3) is exactly equal to 1. If we follow that chain of logic, we get: $.999999\ldots = 3 \times .333333\ldots = 3(1/3) = 1$. Here's another proof: Let $S = .999999\ldots$ S; then $10S = 9.999999\ldots$. Subtracting, we get, $9S = 9$, hence $S = 1$. Here's yet another proof: We agree that .999999… must be less than or equal to 1. That means that $1 - .999999\ldots$ is greater than or equal to 0. But $1 - .999999\ldots$. would be 0.000000…. We can say that either that difference is 0 or that it's smaller than any positive number and, thus, must be 0. We have, then, two quantities, 1 and .999999…, whose difference is 0, and if two quantities have a difference of 0, they must be the same.

In summary, we could say that .99 is close to 1 and .999 is even closer to 1, but .999999… is as close to 1 as desired. And for that reason, we say that those quantities are equal. Another way of looking at .999999… is as an infinite sum, the topic for this lecture. Technically, .999999… = .9 + .09 + .009 + .0009…, and we're interested in what happens when we add an infinite number of numbers together. In general, we say that a series, such as $a_1 + a_2 + a_3 + a_4 + \ldots$, has a sum of S if the sum gets arbitrarily close to S. As an example, .9 + .09 + .009 + … gets arbitrarily close to 1.

Let's look at the example: 1 + 1/2 + 1/4 + 1/8 + 1/16 + … = 2. Imagine that the distance between me and a table is 2 feet. If I walk halfway toward the table, I've just walked 1 foot. If I walk half the distance again, I've walked 1/2 foot. If I walk half the distance again, I've walked 1/4 foot. With every step I take, I'm walking half as much as I did with the previous step. Technically, I never reach the table, but I get arbitrarily close to the table. That's why we say that the sum 1 + 1/2 + 1/4 + 1/8 + … = 2. That sum gets as close to 2 as we desire.

As an infinite sum gets closer and closer to a single number, it is said to *converge*.

As an infinite sum gets closer and closer to a single number, it is said to *converge*. If it doesn't converge, it is said to *diverge*. For example, the sum we just looked at converges to 2. The earlier example, .9 + .09 + .009 + …, converges to 1. In contrast, the sum 1 + 2 + 4 + 8 + 16 + … diverges to infinity. A sum can diverge without getting larger. For instance, the sum 1 − 1 + 1 − 1 + 1 − 1 … is first 1, then 0, then 1 again, then 0, and so on. Because that sum is not getting closer to any real number, we say that it diverges. In order for a sum to converge, the terms of the sum must get closer to 0; otherwise, the sum will not get closer to a real number. For example, the series 1 + 1/2 + 1/4 + 1/8 + 1/16 + … is known as a *geometric series*, which has the form $1 + x + x^2 + x^3 + x^4 \ldots$. In order for the terms to be getting closer to 0, the number x must be between −1 and +1.

Here is the formula for the geometric series: For any number x strictly between −1 and +1, the series $1 + x + x^2 + x^3 + x^4 + \ldots = 1/(1 - x)$. Let's look at a proof of that formula. Let $S = 1 + x + x^2 + x^3 + x^4 \ldots$. Multiplying that equation by x, on the left, we have $x(S)$; on the right, we have $x + x^2 + x^3 + x^4 + \ldots$. Taking away the "excess," we have $S - xS$ on the left, or $S(1 - x)$; on the right, we're left with 1. Solving for S, we get $S = 1/(1-x)$. Let's do the example we saw earlier: When $x = 1/2$, then 1+ 1/2 + 1/4 + 1/8 + 1/16 + … = 1/(1 − 1/2), but the denominator, 1 − 1/2, is equal to 1/2; the answer, then is 1/(1/2), which is 2. When $x = -1/2$, the geometric series tells us that

$$1 - 1/2 + 1/4 - 1/8 + 1/16 - \ldots = \frac{1}{1-(-1/2)}$$, or 1/(3/2), which is 2/3.

Let's go back to the number that we started with: .999999…. We can write that number as .9 + .09 + .009 + …. That's not a geometric series yet, but we can factor out a .9 from everything, leaving us with .9(1 + .01 + .001 + .0001 + …). Those terms are the quantity 1/10th raised to higher and higher powers. In other words, we've pulled out a factor of 9/10 and we're multiplying it by $1 + 1/10 + 1/10^2 + 1/10^3 + 1/10^4 + \ldots$. Adding, that infinite series is 1/(1 − 1/10). In other words, we have 9/10(10/9), which is 1. That's our last proof of the fact that .999999… = 1.

When you use the formula for the geometric series, you must be careful that the x you're using is strictly between −1 and 1; if x is greater than or equal to 1 or less than or equal to −1, then the formula doesn't work. For instance, if we let $x = 2$, then the geometric series produces the nonsensical result that $1 + 2 + 4 + 8 + 16 + \ldots = \frac{1}{1-2}$, which is −1.

Let's do an application of the geometric series. Suppose a ball is dropped from a 50-foot building, and the ball always rebounds to 80% of the height from which it was dropped. How far does the ball travel? Obviously, the ball goes 50 feet down originally, but then it travels up 80% of that, or 40 feet. Then, it drops 40 feet and rebounds up 80% of that, or 32 feet. It drops 32 feet, then rebounds up 25.6 feet, then drops 25.6 feet, and so on. What's the total amount that the ball travels? We can write this out as a geometric series, as shown below.

$$50 + 2(50)(.8) + 2(50)(.8)^2 + 2(50)(.8)^3 + \ldots$$

Simplifying: $50 + 80(1 + .8 + .8^2 + .8^3 + \ldots)$

Solving: $50+80\left(\frac{1}{1-.8}\right) = 50+80\left(\frac{1}{1/5}\right) = 450$ ft

If a sum $a_1 + a_2 + a_3 + \ldots$ converges, we know that its terms must go to 0, but does that guarantee that the sum converges? Surprisingly, the answer is no. We can understand this by looking at the *harmonic series*: 1 + 1/2 + 1/3 + 1/4 + 1/5 + …. Before we look at this proof, note that the harmonic series was given its name by the ancient Greeks. They noticed that strings with lengths of 1, 1/2, 1/3, 1/4, and 1/5, and so on, when plucked, tended to produce harmony.

Now let's look at the proof that the harmonic series goes to infinity. If we take 1/2 + 1/3 + … + 1/9, we're adding nine terms, and you would agree that each of those terms is bigger than 1/10. Thus, the sum of those nine terms must be at least 9/10. Now, let's look at the next 90 terms, the numbers 1/10, 1/11, …, 1/99. We've just added 90 more terms, and each of those terms is bigger than 1/100; the sum of those 90 terms is at least 90(1/100), which is 9/10. Thus, the sum of those 90 terms is bigger than 9/10. In the same way, each of the next 900 terms is bigger than 1/1,000, which means that each of those terms is bigger than 9/10. Then, the next 9,000 terms also add to something bigger than 9/10, and the next 90,000 terms add to something bigger than 9/10. In this way, the sum of all these terms is bigger than 9/10 + 9/10 + 9/10 + …. This sum gets arbitrarily large; thus, it diverges to infinity.

Could we scale down the harmonic series somewhat? What if we cut down every term by 100? Does the sum 1/100 + 1/200 + 1/300 + 1/400+ … converge or diverge? We could factor 1/100 out of those terms, and we would be left with 1 + 1/2 + 1/3 + 1/4 + 1/5 + …, but we know that series diverges to infinity and 1/100 of infinity is still infinity. Interestingly, increasing the denominators of the harmonic series slightly brings about enough of a change to get the series to converge. Instead of using the denominators 2, 3, 4…, we use $2^{1.01}$, $3^{1.01}$, and $4^{1.01}$. That makes the denominators a little bit bigger, which makes the fractions a little bit smaller, and the sum, then, will be less than infinity.

Let's now turn to what mathematicians call an *alternating series*. We start with the numbers $a_1 > a_2 > a_3 > a_4 > \ldots > 0$. If these numbers are getting closer and closer to zero, then the sum of $a_1 - a_2 + a_3 - a_4 + a_5 - a_6 + \ldots$ will converge to a single number. For example, 1 − 1/2 + 1/3 − 1/4 + 1/5 − 1/6 + … must converge. To prove this, think of starting at 1, then subtracting 1/2, then adding 1/3, then subtracting 1/4, adding 1/5, subtracting 1/6, adding 1/7, subtracting 1/8, and so on—getting closer and closer to a single point. The sum is honing in on a single number, which incidentally, is .693…, the natural log of 2. The explanation for that, however, requires calculus.

Let's look again at the same series: 1 − 1/2 + 1/3 − 1/4 + 1/5 − 1/6 + …. Notice that the denominators consist of all the positive numbers, and all the odd denominators are counted positively and all the even denominators are counted negatively. Knowing this, we can add that series up in a slightly different way. Consider the series shown below:

$$\left(1-\frac{1}{2}\right)-\frac{1}{4}+\left(\frac{1}{3}-\frac{1}{6}\right)-\frac{1}{8}+\left(\frac{1}{5}-\frac{1}{10}\right)-\frac{1}{12}+\left(\frac{1}{7}-\frac{1}{14}\right)-\frac{1}{16}+$$
$$\left(\frac{1}{9}-\frac{1}{18}\right)-\frac{1}{20}+\ldots.$$

Even though it looks different, this is just a rearrangement of the original series: every odd denominator is added once and every even denominator is subtracted once. Next, we'll group those numbers, which results in

$$\frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\frac{1}{10}-\frac{1}{12}+\frac{1}{14}-\frac{1}{16}+\frac{1}{18}-\frac{1}{20}+\ldots.$$

That is equal to

$$\frac{1}{2}\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\ldots\right)$$, or half the original series.

We started with the series 1 − 1/2 + 1/3 − 1/4…, and when we rearranged it, we paradoxically wound up with half of the original series. In fact, we can rearrange these same sets of numbers to obtain any sum we want. The lesson here is that the commutative law, $a + b = b + a$, can fail when adding infinite numbers of positive and negative terms. ■

Suggested Reading

Adrian, *The Pleasures of Pi, e and Other Interesting Numbers*.

Bonar and Khoury, *Real Infinite Series*.

Questions to Consider

1. Prove 1/2 + 1/6 + 1/12 + 1/20 + 1/30 + ... = 1, where the first denominator is 1×2, the second denominator is 2×3, and so on. (Hint: 1/12 = 1/3 − 1/4).

2. Suppose that in the harmonic series, we throw away all terms with the number 9 in the denominator (i.e., we eliminate such numbers as 1/9, 1/19, 1/29, 1/97, 1/3141592, and so on). Show that this 9-less series converges.

The Joy of Differential Calculus

Lecture 18

The words "calculus," "calcium," and "calculate," they all have the same root, which is "calculus," which literally means *pebble*. Pebbles were the first calculating devices. We learned to count 1, 2, 3 at a time using pebbles. In calculus, we learn to calculate how things grow and change.

In this first lecture on calculus, we'll have fun with functions, seeing how they grow and change over time. In the next lecture, we'll find an approach for approximating any function with a polynomial, the simplest of functions. In our third lecture on calculus, we'll explore the fundamental theorem of calculus, which allows us to calculate areas and volumes that are impossible to find using only the tools of geometry and trigonometry.

We begin with the study of slopes, which we encounter every day. Any time one quantity varies with another quantity, such as in calculating miles per gallon or price per pound, the idea of slope is involved. Mathematically, the simplest slopes are straight lines, where the slope is constant. For instance, we know from our earlier discussion of algebra that the function $y = 2x + 3$ produces a line with a slope of 2. The line for the function $y = 4x - 1$ has a much steeper slope, 4. The line $y = -x$ has a constant slope of −1. Finally, a constant function, which is also a straight line, such as $y = 5$, has a slope of 0.

Lines have the same slope everywhere, but calculus applies our knowledge of lines to curves, which are not nearly as simple. For instance, let's look at the parabola $y = x^2 + 1$. It doesn't make any sense to try to find the slope of a parabola, because it's constantly changing. But we can ask how fast the function is growing at a specific point. When $x = 3$ on this graph, $y = 3^2 + 1$, which is 10. How fast is the function growing at the point (3, 10)? We're interested in the slope of the line that just touches the graph at the point (3, 10). That line is called the *tangent line*, and our mission is to calculate the slope of that tangent line. We need two points to figure out the slope, but we can use a point that's close to (3, 10) that also lies on the parabola. Let's look at $x = 3 + h$; the y value for that would be $(3 + h)^2 + 1$. When we expand that, we get

$h^2 + 6h + 10$. Now we have two points on the parabola. The first point, (x_1, y_1), is equal to (3, 10). The second point, (x_2, y_2), is equal to $(3 + h, h^2 + 6h + 10)$.

To calculate the slope of the line that goes through those two points, we have to calculate the change in y divided by the change in x. The symbol used in calculus to express *change in* is delta, Δ. Thus, to calculate the change in y divided by the change in x, we look at $\Delta y/\Delta x$. Algebraically, that's equal to $(y_2 - y_1)/(x_2 - x_1)$. The change in y is $h^2 + 6h + 10 - 10$, and the change in x is $3 + h - 3$. Simplifying, that's $(h^2 + 6h)/h$; when we divide by h, we're left with $h + 6$. That result tells us that the slope of the line that goes through the point (3, 10) and the point very close to (3, 10) is equal to $h + 6$. As we let h get closer to 0, the slope of that line gets closer to 6. When h is 0, $6 + h$ becomes 6; therefore, the slope of the tangent line is 6 when $x = 3$.

Any time one quantity varies with another quantity, such as in calculating miles per gallon or price per pound, the idea of slope is involved.

We could go through the same argument for other points on the parabola. For instance, we could use the same algebra to find the slope of the point $(x, x^2 + 1)$, which is simply $2x$. When $x = 1$, the slope of that tangent line is 2. When $x = -3$, the slope of that tangent line is -6. When $x = 0$, the slope of that tangent line is 0. In general, for the function $y = x^2 + 1$, the slope at the point x is equal to $2x$, and we represent that with the notation $y' = 2x$. The term for y' is the *slope function* or the *first derivative*. Note that if we raise or lower the function $y = x^2 + 1$, the tangent line still has the same slope as it did before. If we're looking at the function $x^2 + 17$, or x^2, we still have $y' = 2x$.

The official definition of the derivative is as follows: For any function $y = f(x)$, we define y' as $(f(x+h) - f(x))/h$ (that's the change in y divided by the change in x) and we take the limit of that as h goes to 0. Calculating this is called *differentiation*. Other notations for y' include $f'(x)$ and dy/dx. As we just saw, if $y = x^2$, its derivative is $y' = 2x$. By using the same kind of logic we just used, we can come up with some general rules for calculating derivatives. For example, if $y = x^3$, then $y' = 3x^2$. If $y = x^4$, then $y' = 4x^3$. In general, if $y = x^n$, then $y' = nx^{n-1}$. Even when the exponent is 1, $y = x$, the derivative would be $1x^0 = 1$. That makes sense because the slope of the line $y = x$ is constantly 1.

We can also multiply by a constant when we're differentiating. For instance, given that $y = x^2$ has the derivative $2x$, then $y = 10x^2$ would have the derivative $10(2x)$, or $20x$. Here's another simple rule: the derivative of the sum is equal to the sum of the derivatives. For instance, if we know that the derivative of $4x^3 = 12x^2$, the derivative of $8x^2$ is $16x$, the derivative of $-3x$ is -3, and the derivative of 7 (that's a constant function, $y = 7$) is 0, and we want to find the derivative of the sum of all those functions, then we use this rule to get $12x^2 + 16x - 3$.

Now that we know how to calculate some derivatives, let's look at what we can do with this knowledge. We begin with the function $y = x^2 - 8x + 10$. Looking at a graph of that function, we might ask: Where is that function minimized? Remember we said that when a function reaches its low point, the slope of the tangent line is 0. Wherever a function reaches its minimum or its maximum—that is, whenever we go from decreasing to increasing or from increasing to decreasing—the slope of the tangent line is 0. We can find where this function is minimized by finding where the derivative of that function is equal to 0. The derivative of $x^2 - 8x + 10$ is $2x - 8$. When does that equal 0? Solving $2x - 8 = 0$, we get $x = 4$. That function is minimized when $x = 4$.

Let's do another application, this one involving Laurel's Lemonade Stand. For my daughter's lemonade stand, we decided that if she charged x cents per cup, she would sell $(50 - x)$ cups in one day. If Laurel sells $(50 - x)$ cups, then her revenue is $x(50 - x)$, which is $50x - x^2$. That's the revenue function, which we'll call $R(x)$. The graph of that function is an inverted parabola. Where is that function maximized? We set the derivative of $50x - x^2 = 0$; thus, $50 - 2x = 0$. That equals 0 when $x = 25$. If Laurel charges 25 cents, she can expect to earn $25(50 - 25)$, or 625 cents, or \$6.25.

Laurel's sister, Ariel, wants to create a box where Laurel can keep her supplies. She will make the box, without a lid, from a 12-inch piece of cardboard. To create the box, she cuts four x-by-x squares out of the corners of the cardboard and folds up the edges. What will be the volume of the box? The volume of a box is length times width times height. If Ariel cuts out an x-by-x square from each of the corners, then the length of each side will be $12 - 2x$; the width will also be $12 - 2x$, and the height when the tabs are folded up is x. Thus, the volume is $(12 - 2x)(12 - 2x)x$; if we expand that, we get $4x^3 - 48x^2 + 144x$, which we call $v(x)$. How can we maximize that volume? We set the derivative of the volume equal to 0. Using the power rule and sum rule, we get $v'(x) = 12x^2 - 96x + 144 = 12(x^2 - 8x + 12)$. Setting this equal to zero, and dividing by 12 gives us $x^2 - 8x + 12 = 0$. We can then factor that polynomial to get $(x - 6)(x - 2) = 0$.

Of course, the product of those two numbers can be 0 only if one of the numbers is itself 0. That means either $x - 6 = 0$ or $x - 2 = 0$. Thus, to determine where the volume of the box is maximized, we only need to consider when $x = 6$ or when $x = 2$. We can tell, either by looking at the graph of that function or by actually plugging in the numbers, that when we let $x = 6$, the volume of the box is 0. When $x = 2$, however, we get the biggest volume: $(12 - 2x)(12 - 2x)x = (12 - 4)(12 - 4)2 = 128$ cubic inches.

So far, we've solved only problems that involve polynomials, but the power rule is actually even more powerful than it sounds. Again, the rule is that the derivative of x^n is nx^{n-1}, and it works for any exponent n, even negative integers or fractions. For instance, $y = x^{-1}$ is the function $y = 1/x$. The derivative of that, by the power rule, would be $-1(x^{-1-1})$, or $-1(x^{-2})$. In other words, $y' = -1/x^2$. If we were interested in differentiating $y = 1/x^2$, that would be $y = x^{-2}$. If we differentiate that, we get $-2x^{-3}$, or $-2/x^3$. Here's a derivative that we'll see later, $y = \sqrt{x} = x^{1/2}$. If we differentiate that, we get $y' = 1/2(x^{1/2-1}) = 1/2x^{-1/2}$, which equals $1/(2\sqrt{x})$.

We might also be interested in differentiating the trigonometric function and the exponential function. Such functions model how sound waves travel or how money grows, and are well worth memorizing. The derivative of the sine function is the cosine function. That is, if $y = \sin x$, then $y' = \cos x$. The derivative of the cosine function is the negative of the sine function. That is, if $y = \cos x$, then $y' = -\sin x$. The most important function in calculus is the function $y = e^x$ because, as mentioned earlier, the derivative of e^x is $y' = e^x$. This function tells us that when we plug in x, not only do we get a value of y, but we also get the slope of the tangent line—how fast that function is changing. The derivative of the natural log of x, $\ln x$, is equal to $1/x$.

Let's try to clarify why the derivative of $\sin x$ is $\cos x$. We can look at a graph of the sine function just to see how it increases and decreases. Here, we have the graph of $y = \sin x$. Let's estimate the slope at various points along the graph. For instance, when $x = 0$, the slope of the sine function looks close to 1. At the point $x = \pi/2$, at 90°, we have a slope of 0. Down at π, at 180°, we have a slope of −1. At the bottom of the graph, at $3\pi/2$, we again have a slope of 0. At $x = 2\pi$, we're almost back to where we started, and we again have a slope of 1. The pattern of these slopes, 1, 0, −1, 0, 1…, will repeat forever. If we connect the dots of the slope function, we see that it looks very much like the cosine function.

We know that the derivative of the sum is the sum of the derivatives, but is the derivative of the product the product of the derivatives? Unfortunately, the answer is no. Instead, the *derivative of the product* is "the first times the derivative of the second plus the derivative of the first times the second." The product rule is written as follows: If $y = f(x)g(x)$, then $y' = f(x)g'(x) + f'(x)g(x)$. For example, if we're looking at $y = x^2 \sin x$, we know the derivative of each of x^2 and of $\sin x$, and we can use that to find the derivative of their product. The derivative of the product is the first times the derivative of the second, which would be $x^2\cos x$, plus the derivative of the first times the second, which would be $2x(\sin(x))$. When you add those together, you get the derivative: $x^2\cos x + 2x \sin x$.

The Quotient Rule

If $y = f(x)/g(x)$, then

$$y' = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}$$

The quotient rule is shown at right. To remember it, instead of thinking $f(x)/g(x)$, think high over low, or "hi" over "ho." Then, you can remember y' as: ho-di-hi minus hi-di-ho over ho-ho. For instance, suppose we were constructing an elementary model of planetary motion using a yo-yo moving at constant speed. The tangent of x could tell us the slope of the string when the yo-yo is at time x, and the derivative of the tangent of x could tell us how fast that slope is changing at time x. We want to calculate the derivative of the tangent of x; that's $\sin x/\cos x$. By the quotient rule, $\sin x/\cos x$ has derivative

$$\frac{\cos x \sin'(x) - \sin x \cos'(x)}{(\cos(x))^2}, \text{ which is } \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

Thus, $\tan x$ has derivative $1/\cos^2 x$.

As we said at the outset, calculus is the mathematics of how things grow. In general, there are three ways that functions grow. Functions may have a constant growth; those functions are represented by straight lines. Functions may also grow in proportion to their input. For example, a falling body travels faster and faster according to how long it has been traveling. Finally, functions can grow in proportion to their output. Those functions describe how your bank account grows or how the population grows. All of these functions are described by *differential equations*, which sometimes involve taking derivatives of derivatives, called *second derivatives*. Mathematics, being the language of science, is actually expressed through differential equations. For instance, these equations can describe pendulum motion, vibration, pacemakers, even the beating of your heart. In fact, it's safe to say that, on some levels, your life actually depends on calculus. ■

Suggested Reading

Adams, Hass, and Thompson, *How to Ace Calculus: The Streetwise Guide*.

Thompson and Gardner, *Calculus Made Easy*.

Questions to Consider

1. A manufacturer wants to create a can that will contain 1 liter of liquid. Use differential calculus to determine the dimensions of the can that will minimize the surface area of the can. (Hint: A cylinder with base radius r and height h has volume $\pi r^2 h$ and surface area $2\pi rh + 2\pi r^2$. CAN you see why?)

2. For the function $y = x^3$, what is the slope of the tangent line that passes through the point (2,8). What is the equation for that line?

3. Find the dimensions of a rectangle with perimeter P that has the largest area.

The Joy of Approximating with Calculus

Lecture 19

In this lecture, we'll see how this simple idea, slope of tangent line, has many, many beautiful consequences.

We begin with the chain rule, which refers to chains of functions. We know, for example, that the derivative of $\sin x = \cos x$. We also know that the derivative of $x^3 = 3x^2$. Suppose we want to combine those two functions and calculate the derivative of $\sin(x^3)$. You might guess that the derivative of $\sin(x^3)$ is $\cos(x^3)$ or $\cos(3x^2)$. Both answers are wrong, but they're close. The actual answer is $\cos(x^3)(3x^2)$. In general, if we want to take the sine of $g(x)$ and find the derivative of that function—$\sin(g(x))$—the derivative is equal to $\cos(g(x))\, g'(x)$, or $g'(x)\cos(g(x))$.

Let's do another example. Recall that the derivative of the function e^x is still e^x. What about the derivative of e^{x^3}? The chain rule tells us that the answer is e^{x^3} times the derivative of x^3, which is $3x^2$; thus, the derivative we're looking for is $3x^2 e^{x^3}$. In general, $e^{g(x)}$ has derivative $g'(x)\, e^{g(x)}$. We can also improve the first differentiation rule we learned, the power rule. According to this, if $y = x^n$, then the derivative of $x^n = nx^{n-1}$. The derivative of $[g(x)]^n$ would be $n[g(x)]^{n-1}$ times the derivative of $g(x)$. That is, if $y = [g(x)]^n$, then $y' = n[g(x)]^{n-1} g'(x)$. For instance, let's calculate the derivative of $(x^3)^5$. According to the chain rule, that's $5(x^3)^4(3x^2) = 15x^{12}\, x^2 = 15x^{14}$ as the derivative. We can verify this answer because the problem started off as $(x^3)^5$, which is just an unusual way of writing x^{15}, and we know from the power rule that the derivative of x^{15} is, indeed, $15x^{14}$. In general, the chain rule says that if we have a function of a function, $y = f(g(x))$, then $y' = f'(g(x))\, g'(x)$.

Let's now use the chain rule to solve the following cow-culus problem: Claudia the cow is 1 mile north of the *X*-Axis River, which runs east to west. Her barn is 3 miles east and 1 mile north. She wishes to drink from the *X*-Axis River, then walk to her barn in such a way as to minimize her total amount of walking. Where on the river should she stop to drink? If she starts at the point (0,1), her barn is at (3,2). Suppose she decides to drink from the point x along the river, that is, at the point (x,0). As she walks from her starting point to x, she creates a right triangle with one leg of length 1, base length x, and hypotenuse length $\sqrt{x^2+1}$.

Then, Claudia has to walk from point x to her barn. That's another triangle where the base has length $3 - x$, the height is 2, and the hypotenuse is $\sqrt{(3-x)^2+2^2}$. When we expand that, we have $\sqrt{x^2-6x+13}$. The total distance that Claudia walks when she stops at x is $f(x) = (x^2 + 1)^{1/2} + (x^2 - 6x + 13)^{1/2}$. By the chain rule, $f'(x) = \frac{1}{2}(x^2 + 1)^{-1/2}(2x) + \frac{1}{2}(x^2 - 6x + 13)^{-1/2}(2x - 6)$. We want to find the place where the function $f(x)$ is minimized, and to find such a point, we need to find where the derivative is 0. The solution to this equation is $x = 1$, which we can verify.

I gave you the solution of $x = 1$, but how could we have derived it? The fact is that this problem can be solved, if you'll pardon the pun, after just a moment's reflection, without ever using calculus. Imagine that Claudia, as she walks from her original position to the X-Axis River, instead of walking back to her barn at the coordinates (3, 2), walks to the barn's reflection at the point (3, −2). Notice that the distance from her drinking point to (3, 2) is the same as the distance from her drinking point to (3, −2). Since the shortest distance between two points is a straight line, to find the optimal path, we draw a straight line between the original point at (0, 1) to the reflected point at (3, −2). The slope of that line is $-3/3 = -1$. If the line starts at the point (0, 1), then it will hit the x-axis at the point where $x = 1$.

Now let's look at a way to approximate the square root of any number in your head. Our tool for this is the all-purpose approximation formula. The all-purpose formula works for almost any differential function. It says: $f(a + h) \approx f(a) + h\,f'(a)$. Generally, the smaller h is, the better the approximation is.

The reason this formula works is fairly simple. If we go back to the original definition of the derivative, $f'(a) \approx \frac{f(a+h)-f(a)}{h}$. As h goes to 0, that approximation becomes exact. Let's now use this approximation formula to calculate square roots in our heads. We know that the function $f(x) = \sqrt{x}$, has derivative $f'(x) = 1/(2\sqrt{x})$. In particular, if we plug in the value $x = a$, we get $f'(a) = 1/(2\sqrt{x})$. Let's say we want to estimate $\sqrt{106}$. We can break 106 into 100 + 6, and we'll let $a = 100$ and $h = 6$. Our approximation formula tells us that $\sqrt{106} = f(106), \approx f(100) + 6\,f'(100)$. But $f(100)$ is $\sqrt{100} = 10$.

To that, we add $6\,f'(100) = 6/(2\sqrt{100}) = 6/20$. Hence, our approximation is $10 + 6/20 = 10.3$. As it turns out, $\sqrt{106} = 10.295\ldots$.

Let's do another example: $\sqrt{456}$. We know that $\sqrt{400} = 20$, so our first guess is 20 plus an error of 56. We take 20 + 56/2(20), which equals 20 + 1.4 = 21.4. We can get an even better approximation using the process for squaring numbers that we learned in one of our lectures on algebra. Using this process, we know that $21^2 = 441$, which makes our error smaller; h is only 15 instead of 56. In this case, we calculate $\sqrt{456}$ as 21 + 15/2(21) = 21 + 15/42 = 21.357. The exact answer is 21.354.

We've seen a number of parallels between the hyperbolic functions and the trigonometric functions.

Let's return to the approximation formula that says $f(a + h) \approx f(a) + h\,f'(a)$. We plug in $a = 0$ and replace h with x to get a much simpler looking equation. This says $f(x) \approx f(0) + f'(0)x$. Once we have the function f, $f(0)$ is just a number, as is $f'(0)$. If $f(x)$ is some number plus some other number times x, that's the equation of a line with a slope of $f'(0)$. That line goes through the point $(0, f(0))$. In other words, we're approximating the function $f(x)$ with a straight line that goes through the same point, $(0, f(0))$, with the same slope.

Let's look at the graph of $y = e^x$; near the point (0,1), we have a line (actually, it's the line $y = 1 + x$) that looks just like the function e^x, at least when x is close to 0. If we want an even better approximation, then we look for a parabola, a second-degree polynomial, to go through the same point. Because we have one extra degree of freedom, not only will the parabola go through that point with the same slope (the same first derivative), but the parabola will also go through that point with the same second derivative. The magic formula for that is $f(x) \approx f(0) + f'(0)x + (1+\sqrt{5})/2$. Now we have a parabola that matches the function with the same first derivative and second derivative. We call this the second-degree *Taylor polynomial approximation*.

If we want an even better fit, we can get a third-degree approximation by adding a cubic term: $(f'''(0)x^3)/3!$. The reason we use 3! is that now that function will match the original function through the point $(0, f(0))$ with the same first derivative, second derivative, and third derivative. We can also bring this function out to even higher degrees with the same kind of formula. Let's use the function $f(x) = e^x$; we choose this function because $f'(x)$, its first derivative, is e^x. The second derivative is also e^x, as is the third derivative. When we plug those in at 0, $f(0) = 1$, $f'(0) = 1$, $f''(0) = 1$, and $f'''(0) = 1$. That tells us that near the point $x = 0$, e^x is approximately $1 + x + x^2/2! + x^3/3!$.

We're approximating the important function e^x by a cubic polynomial, and when we're close to 0, it's a pretty good fit. The n^{th}-degree Taylor polynomial would be $1 + x + x^2/2! + \ldots + x^n/n!$. If we let n go to infinity, we get perfect accuracy for all values of x. This is called the *Taylor series* of x, and it has amazing consequences. For instance, look what happens when we differentiate the Taylor series for e^x (which is $1 + x + x^2/2! + x^3/3!$...), one term at a time: The derivative of 1 with respect to x is 0. The derivative of x is 1. The derivative of $x^2/2!$ is x. The derivative of $x^3/3!$ is $3x^2/3!$, but the 3's cancel and we're left with $x^2/2!$. The derivative of $x^4/4!$ is $4x^3/4!$, which is $x^3/3!$. When we differentiate the terms of the series for e^x, we get e^x again, which makes sense because the derivative of e^x is e^x.

Let's look at some more important Taylor series, which can be derived in the same way that we derived the e^x series. For instance, sin x has the following

Taylor series: $x - x^3/3! + x^5/5! - x^7/7! + x^9/9!$.... This looks just like the odd terms of the e^x series except that the signs alternate. Let's look at the graph of $y = \sin x$ and its approximation with the function $y = x$. The function $x - x^3/3!$ is an even better approximation, and the fifth-order Taylor approximation, $x - x^3/3! + x^5/5!$, is even better. We can figure out the series for cos x by differentiating the series for sin x. We know that the derivative of sin x is cos x; differentiating the terms of the Taylor series for sin x, we get the series for cos x, namely $1 - x^2/2! + x^4/4! - x^6/6!$.... Those are the even terms of the e^x series, again with the signs alternating.

Now, let's have some more fun with functions. Look at the series for e^{-x}; that's what we get when we take the e^x series and replace all the x's with $-x$. This gives us $e^{-x} = 1 - x + x^2/2! - x^3/3! + x^4/4! -$.... Thus, the e^{-x} series looks like the e^x series except the signs are alternating. If we add the e^x series to the e^{-x} series and divide by 2 (taking the average of those two functions), we get the *hyperbolic cosine function*, or *cosh function*. That is, cosh x = $(e^x + e^{-x})/2$. Look what happens when we add those series together: The odd terms cancel, and the even terms stay the same. Thus, $\cosh x = 1 + x^2/2! + x^4/4! + x^6/6! +$.... One reason that's called the hyperbolic cosine is that its infinite series looks just like the infinite series of the cosine function except that the cosine function has alternating signs.

Similarly, if we subtract those two infinite series, the odd terms survive and the even terms are eliminated. We're left with $\sinh x = (e^x - e^{-x})/2 = x + x^3/3! + x^5/5!$.... It looks just like series for the sine function except that it doesn't have alternating signs. That's called the *hyperbolic sine function*, or *sinh function*. Notice also that $\sinh' x = \cosh x$ and $\cosh' x = \sinh x$. We see hyperbolic functions everywhere in our daily lives. For instance, a hanging cable or piece of rope always fits a cosh curve. In fact, every hanging rope or chain is of the form $y = \frac{1}{a}\cosh\left(\frac{x}{a}\right)$. Note that to differentiate this function, we would use the "chain rule." Where does the word *hyperbolic* come from in these functions? We know that $(\cos\theta, \sin\theta)$ exists on the unit circle since $\cos^2 + \sin^2 = 1$. Similarly, we can show that $\cosh^2 - \sinh^2 = 1$, which means that $(\cosh\theta, \sinh\theta)$ lies on the unit hyperbola, and that's where the word comes from. Another easy property to verify is that $\cosh x + \sinh x = e^x$; that can be verified by the series or by the original definition.

We've seen a number of parallels between the hyperbolic functions and the trigonometric functions, and if cosh + sinh = e^x, then there must be some connection among cosine x, sine x, and e^x. The connection is Euler's equation: $e^{ix} = \cos(x) + i \sin(x)$. We could prove that by the series for e^x , replacing all the x's with ix's. As that i is raised to different powers ($i^0 = 1$, $i^1 = i$, $i^2 = -1$, $i^3 = -1$, $i^4 = 1$) then the sign pattern is: 1, i, −1, −i, 1, i, −1, −i. As we look at that pattern and separate the real part from the imaginary part, we get the series for cos x plus i times the series for sin x. That's the proof of Euler's equation: $e^{ix} = \cos x + i \sin x$. Incidentally, as we observed earlier, when we let $x = \pi$ or 180°, then $e^{i\pi} = -1$, that is, $e^{i\pi} + 1 = 0$. This equation was recently listed as number two on a list in *Physics World* magazine of the 20 greatest equations. ■

Suggested Reading

Adams, Hass, and Thompson, *How to Ace Calculus: The Streetwise Guide*.

Thompson and Gardner, *Calculus Made Easy*.

Questions to Consider

1. What is the value of 1 − 1/1! + 1/2! − 1/3! + 1/4! − 1/5! + ...?

2. Use the approximation formula to derive a method for mentally determining good approximations to $\sqrt[3]{a+h}$, where a is a number with a known cube root. For example, come up with a good mental estimate of $\sqrt[3]{1024}$.

The Joy of Integral Calculus

Lecture 20

We can answer our big problem by chopping it up into little simple problems, then putting all those little simple answers together. That's where the word "integration" comes from. It's a very powerful idea.

Calculus is typically broken into two parts: differential calculus and integral calculus. Differential calculus, as we've studied in our last two lectures, is the mathematics of how things change and grow. Integral calculus is used, among other things, to calculate areas and volumes. The big idea in both differential calculus and integral calculus is to calculate quantities associated with curves using quantities associated with straight lines. For example, in differential calculus, we used our understanding of the slope of a straight line to calculate the slopes of parabolas and trigonometric functions.

In integral calculus, where the goal is to calculate areas, we'll begin by looking at areas we understand, such as the area of a rectangle, and use that knowledge to figure out, for example, the area under a curve. Initially, you wouldn't expect to find much of a connection between the calculation of slopes and the calculation of area, yet those two concepts are intimately connected through the fundamental theorem of calculus. We'll begin this lecture by looking at that theorem.

The original problem of integral calculus is to find the area under some kind of a curve, and we can answer questions about that area using the fundamental theorem of calculus. Suppose we want to carpet a room that is mostly rectangular but has a curved section described by the function $y = f(x)$. According to the fundamental theorem of calculus, to find the area of that region, we first have to find a function, $F(x)$, with $F'(x) = f(x)$. Once we've found that function, we calculate the area with the formula: $F(b) - F(a)$.

Let's look at a specific example, the parabola described by the function $y = x^2$. Suppose we want to find the area under the curve as x goes from 1 to 4. The first step in the fundamental theorem of calculus is to find a function with $F'(x) = x^2$. If we differentiate $x^3/3$, we know from the power rule that we get $3x^2/3$. The 3's cancel and we're left with x^2. Thus, $f(x) = x^3/3$. The next step is to plug in the endpoints to the function we just found. In other words, we calculate $F(4) - F(1) = 4^3/3 - 1^3/3 = 64/3 - 1/3 = 63/3$, which is exactly 21. Therefore, by the fundamental theorem of calculus, the area under the parabola between 1 and 4 is exactly 21. The notation we use for this is shown at right . In this lecture, we'll see how to interpret integrals as infinite sums.

$$\int_1^4 x^2 dx = \frac{x^3}{3}\Big]_1^4$$

The symbol $\int$ is an elongated "s," where "s" stands for "sum."

Let's do another example. We'll calculate the area under the curve for the function $y = \sin x$ as x goes between 0 and π. Before we calculate, we do a bit of guessing. We know that the sine function, at its peak, has a height of 1. We could enclose that entire curve inside a rectangle that has a height of 1 and a length of π; thus the area under the curve can't be bigger than π. To apply the fundamental theorem, we must find a function whose derivative is $\sin x$, and that function is $F(x) = -\cos x$. We then evaluate $F(\pi) - F(0) = -\cos(\pi) + \cos(0) = -(-1) + 1 = 2$; the area under the curve is 2.

If we're looking at a curve that goes above and below the x-axis, then we have to interpret the integral slightly differently. For instance, if we're looking at the function $y = \sin(x)$ as x goes from 0 to 2π, then the area below the x-axis is counted negatively. With that information, what would you expect to find for $\int_0^{2\pi} \sin x dx$? Is there more area above the curve, more area below the curve, or are they equal? Because the function looks symmetrical, we would expect the positive part and the negative part to cancel each other out and give us an answer of 0. Let's apply the fundamental theorem of calculus to see if we get that answer. The anti-derivative of $\sin x$ is still $-\cos x$. We evaluate this at the endpoints 0 and 2π, but cos(2π) is the same as cos 0, so they cancel each other out exactly. Hence, this integral results in 0, as expected.

What is it that makes the fundamental theorem of calculus do its magic? Before we answer that, let's look at a different question: Suppose we have two functions that have the same derivative. Must those two functions be the same? If $f'(x) = g'(x)$, does that mean that $f(x) = g(x)$? The answer is: almost, but not quite. For example, what functions have the derivative $2x$? We know that x^2 has a derivative of $2x$, as do $x^2 + 1$, $x^2 + 17$, and $x^2 - \pi$. Anything that's of the form $x^2 + c$ has a derivative of $2x$, and the only functions that have a derivative of $2x$ are of the form $x^2 + c$. Try to remember this theorem: If two functions have the same derivative, then those two functions differ by a constant. Mathematically, if $f'(x) = g'(x)$, then $f(x) = g(x) + c$.

Knowing this theorem, we're ready to answer the question: What makes the fundamental theorem of calculus do its trick? Our goal is to prove that if we have a function $y = f(x)$ and we want to find the area under the curve between the points $x = a$ and $x = b$, we find a function F whose derivative is f, then evaluate $F(b) - F(a)$ to find the area. We begin with the quantity $R(x)$, which is the area of the region under the curve between a and x. Notice that as we vary x, the region under the curve also varies, and its area will vary. What if we move x on top of a? The area of the region, then, is 0. We're looking at a straight line, which doesn't have any area. Thus, $R(a) = 0$, as will be useful later.

If two functions have the same derivative, then those two functions differ by a constant.

Our goal with the fundamental theorem is to show that the area under the curve from a to b is $F(b) - F(a)$. But, by definition, the area under the curve from a to b, is $R(b)$. Thus, the goal of this theorem is to conclude that $R(b) = F(b) - F(a)$. How are we going to get there? Remember, $R(x)$ is the area under the curve as we go from a to x. What's $R(x + h)$? By definition, that is the area under the curve as we go from a to $(x + h)$. The difference in those quantities, $R(x + h) - R(x)$, is the area as we go from a to $(x + h)$ minus the area as we go from a to x. Almost everything gets canceled there except for the tiny region between x and $(x + h)$.

Looking at a blowup of that region, we see that if h is really small, the region is almost rectangular, and its area, then, is approximately the area of a rectangle with base h and height $f(x)$; thus, its area is approximately h multiplied by $f(x)$. Dividing both sides of this equation by h, we get $(R(x + h) - R(x))/h \cong f(x)$. As we let h go to 0, the expression becomes $R'(x) = f(x)$. And since $F'(x) = f(x)$, we have $R'(x) = F'(x)$. As we said earlier, if two functions have the same derivative, they differ by a constant; therefore, $R(x) = F(x) + c$. That constant must work for every value of x that we plug into it; in particular, it must work when we plug in the value $x = a$. If we plug in $x = a$, then $R(a) = F(a) + c$. Remember, though, that $R(a) = 0$. Solving for c, we find that $c = -F(a)$. Plugging that value into the formula above, $R(x) = F(x) - F(a)$ for all values of x. Because that works for all values of x, in particular, it must work for $x = b$; therefore, $R(b) = F(b) - F(a)$.

Motivated by the fundamental theorem of calculus, here are some techniques for finding anti-derivatives of functions. We use the following notation for anti-derivatives: $\int f(x)dx$, which represents the set of all functions that have derivative $f(x)$. For example, $\int 2xdx$ is simply asking for all functions that have a derivative of $2x$. We know that all functions with a derivative of $2x$ are of the form $x^2 + c$. Thus, $\int 2xdx = x^2 + c$.

Let's look at some other rules for calculating integrals. The power rule for derivatives has a reverse power rule for finding integrals:

$$\int x^n dx = \frac{x^{n+1}}{n+1} + c .$$

For example, the reverse power rules says that $\int x^3 dx = x^4/4 + c$. Multiplying through by constants—real numbers—is as easy as it was for derivatives. Since $\int x^3 = x^4/4 + c$, then $\int 7x^3 = 7x^4/4 + c$.

Recall that the derivative of the sum was the sum of the derivatives. The same sort of rule works for anti-derivatives. That is, the integral of the sum is the sum of the integrals. For instance, we know $\int 7x^3$ and $\int 2x$; therefore, we can calculate $\int 7x^3 + 2x$ just by adding our previous answers. That would be $7x^4/4 + x^2 + c$.

Unfortunately, as we saw with derivatives, the integral of the product is not the product of the integrals. There are some techniques of integration, however, that can help us do these kinds of problems. The equation for a typical bell curve is: $f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$. The bell curve is used to describe numerical quantities, such as exam scores or heights and weights. If we want to find the average value of something that came from a bell-shaped region, then we need to calculate an integral, such as $\int xe^{-x^2} dx$. We'll calculate this integral using the method of integration by guessing. As a guess, we might say it equals e^{-x^2}. By the chain rule, when we differentiate that, we get $(-2x)\, e^{-x^2}$. If it weren't for the −2, we'd have the answer exactly. If we divide through by −2 in our original guess, however, we get: $\int xe^{-x^2} dx = -1/2e^{-x^2} + c$.

What if we wanted to find the area between two different points on a bell curve, such as the area between −1 and 2 under the bell curve e^{-x^2}? The fundamental theorem of calculus tells us to find an anti-derivative. Unfortunately, this function, e^{-x^2}, has no simple anti-derivative. We have to resort to the naïve idea of calculating the area by summing up a number of rectangles, at least theoretically. The notation $\int_a^b f(x)dx$ comes from summing a group of rectangles. Imagine breaking up a region from *a* to *b* into a bunch of little rectangles. We draw a rectangle that starts at the bottom at the point (x,0) and goes to the top of the curve to the point $(x, f(x))$ with a height of $f(x)$; its base is Δx. The area of that rectangle is $f(x)\Delta x$. If we continue to draw rectangles so that we completely cover that spectrum as x goes from a to b, then we're literally summing values of the form

$$\sum_{x=a}^{b} f(x)\Delta x \,.$$

As the widths of those rectangles get smaller and smaller, we get

$$\int_{a}^{b} f(x)dx \,.$$

Thus, when those Δx's go to 0, the Δx becomes a dx.

Let's put this into practice by calculating the area of a circle. We can do this simply by adding up the areas of all the little rings inside. The large circle has a radius of R. We extract one ringlet of that circle, which has a radius of r and a circumference of $2\pi r$. We can flatten that ringlet out and look at the area of the edge, whose length is $2\pi r$ and whose thickness is Δr. The total area will be the sum of $2\pi r(\Delta r)$ as the radius goes from 0 to R. As Δr gets smaller, that sum becomes the integral $\int_0^R 2\pi r dr$. We know the anti-derivative of $2\pi r$ is $F(r) = \pi r^2$. Hence the area of a circle is $F(R) - F(0) = \pi R^2 - \pi 0^2 = \pi R^2$, exactly as expected.

The use of the word *integration* in mathematics comes from the fact that we can answer a big problem by breaking it up into smaller, simpler problems, then putting the simple answers together. For example, we can use integration to figure out the volume of a sphere. One way to create a sphere is by taking a flat circle, such as a lid, and rotating it around the x-axis. Then, we can calculate the volume by chopping the sphere into tiny parts. Chopping off one tiny part, we have a circle with a little bit of thickness, a radius of y, and an area of πy^2. If we call the thickness Δx, then the volume of this small piece is $\pi y^2(\Delta x)$. Because the equation of the original circle was $x^2 + y^2 = R^2$, we can replace y^2 with $R^2 - x^2$. Thus, the sum of $\pi y^2(\Delta x)$ can be written as the sum of $\pi(R^2 - x^2)\Delta x$. We're summing this as x goes from $-R$ to $+R$. In other words, as we let the widths of those slices get smaller, the volume is equal to $\int_{-R}^{R} \pi(R^2 - x^2)dx$. Finding the anti-derivative of that is a fairly simple matter. When we do the algebra, we get $4/3\pi R^3$, which is the volume of a sphere.

Integrals can calculate areas and volume, but also other physical quantities, such as center of mass, energy, and fluid pressure. In fact, along with differential equations, they describe everything from heat to light to sound to electricity. Without a doubt, calculus is an integral part of our daily lives. ■

Suggested Reading

Adams, Hass, and Thompson, *How to Ace Calculus: The Streetwise Guide*.

Thompson and Gardner, *Calculus Made Easy*.

Questions to Consider

1. Using the chain rule, find the derivative of *ln* x, *ln* $3x$, and *ln* $7x$. Explain what you see.

2. Verify the calculation expressed by this limerick:

The integral $z^2\, dz$
From 1 to the cube root of 3,
Times the cosine
Of 3 pi over 9
Is the log of the cube root of *e*.

The Joy of Pascal's Triangle

Lecture 21

You could spend your life looking and studying patterns that live inside of this beautiful triangle. We're going to, in this lecture, define the triangle, we'll explore the triangle, and ultimately understand the triangle.

The next three lectures are devoted to topics in probability. We'll use some calculus in these lectures, as well as some discrete mathematics that depends on one of the most beautiful objects in mathematics, Pascal's triangle. Let's begin by looking at the first six rows of Pascal's triangle, labeled 0 through 5. We create numbers in this triangle by adding two consecutive numbers in a given row to produce the number below. These numbers are denoted $T(n, 0)$, $T(n, 1)$, … $T(n, n)$. For instance, in row 4, we have $T(4, 0) = 1$, $T(4, 1) = 4$, $T(4, 2) = 6$, and so on.

Pascal's Triangle

0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1

The rule for creating the rows of Pascal's triangle is: $T(n, 0) = 1$, $T(n, n) = 1$, which says that the row begins and ends with a 1, and for $T(n, k)$, we take $T(n - 1, k - 1) + T(n - 1, k)$. The 10 that appears in row 5 would be known as $T(5, 2)$ and that's equal to $T(4, 1) + T(4, 2)$. We can use this rule to create rows in the triangle. For instance, row 6 would begin with a 1; then the 6 would be obtained by adding 1 + 5. Then, we add 5 + 10 = 15, 10 + 10 = 20, 10 + 5 = 15, 5 + 1 = 6, and we end with a 1 again. Let's take a look at some patterns inside the triangle. For instance, notice that each row is symmetric. It reads the same way left to right as right to left. Formally, we say $T(n, k) = T(n, n - k)$. If we were to add the numbers in the triangle row by row, we see that row 0 adds to 1, row 1 adds to 2, row 2 adds to 4, row 3 adds to 8, and so on. Those are powers of 2; in general, row n sums to the number 2^n, or $T(n, 0) + T(n, 1) + … + T(n, n) = 2^n$.

I call the next pattern the *hockey stick identity*. It occurs when we add the diagonals in the triangle. For instance, when we add 1 + 3 + 6 + 10 + 15 + 21 + 28, we get 84, which lies below and to the right of 28. This is the hockey stick identity because of its shape: a long stick that juts out in a new direction to give the next entry of the triangle. This rule works whether we're adding diagonally going to the left or the right.

We can understand some of these patterns through combinatorics. Mathematicians typically define $\binom{n}{k}$ as the number of size k subsets of the numbers 1 … n; we defined it as the number of ways to choose k objects from a group of n objects when order is not important. For instance, if I have n students in my class and I need k of them to form a committee, then the number of ways to create that committee is $\binom{n}{k}$.We saw the formula for solving this earlier. But for $k < 0$ or $k > n$, we don't even think of the formula; we just think of the definition, and we get 0. In other words, how many ways could I create a committee with −5 students? Of course, the answer is 0.

How does $\binom{n}{k}$ relate to Pascal's triangle? I claim that $T(n, k) = \binom{n}{k}$. Looking at the first five rows of the triangle, we can see the terms as $\binom{n}{k}$; thus, row 4 (1, 4, 6, 4, 1) is $\binom{4}{0}, \binom{4}{1}, \binom{4}{2}, \binom{4}{3}, \binom{4}{4}$. If we calculate $\binom{4}{2}$ by the formula, we get $\frac{4!}{2!(2!)} = \frac{24}{2(2)} = 6$. At least in the first five rows, it looks as if my claim is true. Let's prove this idea. We know that the boundary numbers for $\binom{n}{k}$ satisfy $\binom{n}{0} = 1$; $\binom{n}{n} = 1$. Thus, the boundary conditions are as expected.

The triangle condition was $T(n, k) = T(n - 1, k - 1) + T(n - 1, k)$. Will that growth condition, or *recurrence relation*, remain true as we look at the numbers $\binom{n}{k}$? Can we show that $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$? One way we can show this is true is by using algebra. That is, we add the terms using the factorial definition; we then put those terms over a common denominator of $k!(n - k)!$, add the fractions, and when the dust settles, we get $\frac{n!}{k!(n-k)!}$, or $\binom{n}{k}$.

We can also use a combinatorial proof. Returning to the original question, from a class of n students, how many ways can I create a committee of size k? On the one hand, we know the answer to that question is $\binom{n}{k}$. On the other hand, we can answer that question through something known as *weirdo analysis*. Imagine that student number n is the weirdo. Among the $\binom{n}{k}$ committees, how many of them do not use the weirdo? We're looking at size k committees from the class of students 1 through $n-1$. By definition, that's $\binom{n-1}{k}$. How many of those committees must use student n? If student n is on the committee, then we must choose $k - 1$ more students to be on the committee from the remaining $n - 1$ students. Again, by definition, we're looking at $\binom{n-1}{k-1}$. There are $\binom{n-1}{k}$ committees without the weirdo and $\binom{n-1}{k-1}$ with the weirdo; their sum is the total number of committees. Hence, the number of size k committees is $\binom{n-1}{k-1} + \binom{n-1}{k}$.

Comparing our two answers to the same question, we get

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

We've shown that the $\binom{n}{k}$ terms (called *binomial coefficients*) have the same boundary conditions as Pascal's triangle. They will continue to grow in the same way as the entries of Pascal's triangle; therefore, they are the elements of Pascal's triangle.

All the patterns of Pascal's triangle can be expressed in terms of binomial coefficients. For example, let's look at the pattern we saw earlier, that the elements of row n sum to 2^n. In terms of binomial coefficients, this says

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n} = 2^n.$$

We express this idea using *sigma notation*:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Sigma is the Greek letter Σ and is read: "the sum as k goes from zero to n of…". Here's our combinatorial proof, beginning with the question: How many committees can we form from a class of size n? We can break up the question by considering the size of the committee and adding the answers, as shown below. Now we ask: Why is the number of committees 2^n? We can answer this using the rule of product. To create a committee, we go through the classroom student by student and decide whether or not each student will be on the committee. For each student, we have two choices, on or off, from student 1 up through student n. That's $2 \times 2 \times 2 \times 2 \times 2 \times \ldots \times 2$ n times, or 2^n ways to create a committee.

$$\text{Committees of size } 0 = \binom{n}{0}$$

$$\text{Committees of size } 1 = \binom{n}{1}$$

$$\text{Committees of size } 2 = \binom{n}{2}$$

$$\ldots$$

$$\text{Total committees} = \sum_{k=0}^{n} \binom{n}{k}$$

Another useful theorem in mathematics is the *binomial theorem*, which we can find inside Pascal's triangle. Remember this equation from basic algebra: $(x + y)^2 = x^2 + 2xy + y^2$, as appears in row 2 of Pascal's triangle (1, 2, 1). We see that $(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$, and the coefficients in that expression are row 3 of the triangle (1, 3, 3, 1). The expression $(x + y)^4$ would be $x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$, and those coefficients are row 4 (1, 4, 6, 4, 1). In general, for $(x + y)^n$, the coefficients are the numbers in the n^{th} row of Pascal's triangle. Specifically, the coefficient of $x^k y^{n-k}$ is $\binom{n}{k}$.

We can think of $(x + y)^n$ as $(x + y)(x + y)(x + y)(x + y)\ldots$ n times. There's only one way to get an x^n term, and that's by taking x from the first expression times x from the second expression times x from the third expression, all the way down to x from the last expression. There are n ways to create an $x^{n-1}y$ term simply by deciding which y's we will use, then letting the rest of the terms be

x's. For $x^{n-2}y^2$, we choose two terms to be y's and all the rest x's. There are $\binom{n}{2}$ ways to pick two y's here; thus, the coefficient of $x^{n-2}y^2$ is $\binom{n}{2}$.

To summarize, the binomial theorem says:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

This simple formula can be applied to produce many beautiful identities. For example, if we let $x = 1$ and $y = 1$, the binomial theorem tells us that

$$\sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n.$$

Here's another identity:

$$\sum_{k=0}^{n} k\binom{n}{k} = n2^{n-1}.$$

One way to prove this is to let $y = 1$ in the binomial theorem; thus:

$$(x+1)^n = \sum_{k=0}^{n} \binom{n}{k} = x^k.$$

Let's differentiate both sides of this equation with respect to x. When we differentiate the left side, we get $n(x+1)^{n-1}$. When we differentiate the right side, each summand has derivative of

$$\binom{n}{k} kx^{k-1}.$$

Hence

$$n(x+1)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} kx^{k-1} .$$

When we set $x = 1$, all the x's disappear, and we're left with

$$\sum_{k=0}^{n} k \binom{n}{k} = n2^{n-1} .$$

We can prove this same theorem combinatorially. For example, from a class of n students, how many ways can we create a committee of any size with a chair? If the committee has size k, there are $\binom{n}{k}$ ways to create the committee. Once we've done that, there are k ways to choose the chair of the committee. Thus, the number of committees of size k with a chair is

$$\binom{n}{k} k ;$$

the total number of committees over all possible values of k is

$$\sum_{k=0}^{n} k \binom{n}{k} .$$

There's also a more direct way of answering this question. To create a committee of any size from a class of n students, first, we have n ways to pick a chair. Once we've done that, we have to choose a subset of the remaining $n - 1$ students to serve on the committee. How many possible committees can we form from the remaining $n - 1$ students? As we saw earlier, that's 2^{n-1}. The number of committees of this type, then, is $n2^{n-1}$.

Let's look at some other patterns in Pascal's triangle. We summed the rows of the triangle earlier; let's now sum the diagonals of the triangle. We write it as a right triangle to make the pattern easier to see. Summing the first diagonals, we get 1, 1, 2, 3, 5, 8, and so on. These sums are the Fibonacci numbers. In any given row of Pascal's triangle, how many of the numbers are odd? The top row has one odd number, the next row has two, the next row also has two, the next row has four, and so on.

The number of odd numbers in each row of Pascal's triangle is always a power of 2. In fact, it's 2 raised to the number of 1's in the binary expansion of n. Let's look at an example of this. Row 81 of Pascal's triangle has the numbers $\binom{81}{0}, \binom{81}{1}, \binom{81}{81}$. How many of those binomial coefficients are odd? The number 81, written in terms of powers of 2, is 64 + 16 + 1, which in binary notation is 1010001. There are three 1's in that binary expansion of 81, so that will be our exponent. The number of odd numbers in row 81 of Pascal's triangle is $2^3 = 8$. The positions of the 8 odd numbers in row 81 of Pascal's triangle are those numbers that can be formed using a subset (possibly empty) of the numbers 1, 16, and 64. They are: 0, 1, 16, 64, 1 + 16 = 17, 1 + 64 = 65, 16 + 64 = 80, and 1 + 16 + 64 = 81.

Let's end on a holiday note with "The Twelve Days of Christmas." What is the total number of gifts received by the end of the 12 days? On the k^{th} day, you received 1 + 2 + 3 +…+ k, but we know that's equal to $k(k + 1)/2$, which is also equal to the binomial coefficient $\binom{k+1}{2}$. For example, on the 12^{th} day of Christmas, you receive 1 + 2 + 3 + … + 12 gifts. That's equal to (12)(13)/2, or 78 gifts; it's also equal to $\binom{13}{2}$. All the numbers of gifts you receive (1, 3, 6, 10) lie on Pascal's triangle. In fact, when we summed those numbers earlier, we got the hockey stick identity. In general, if we sum the numbers at right, the hockey stick identity tells us that we get $\binom{14}{3}$

gifts altogether. Calculating, that's (14)(13)(12)/3!, or 364 gifts. By the end of the song, you've received one gift for every day of the year, except Christmas. ■

Suggested Reading

Benjamin and Quinn, *Proofs That Really Count: The Art of Combinatorial Proof*.

Gross and Harris, *The Magic of Numbers*, chapter 6.

Questions to Consider

1. There are three odd numbers in the first two rows of Pascal's triangle. How many odd numbers are in the first 4 rows, the first 8 rows, and the first 16 rows? Find a pattern. Can you prove it? Also, describe the resulting picture of Pascal's triangle if you remove all the even numbers from it (or simply replace each even number with 0 and replace each number with 1).

2. Choose any number inside Pascal's triangle and note the six numbers that surround it. For example, if you choose the number 15 in row 6, then the six surrounding numbers are 5 and 10 (above it), 6 and 20 (beside it), and 21 and 35 (below it). Now draw two triangles around that number so that each triangle contains three of those numbers. For example, the first triangle would contain 5, 20, and 21, while the second the triangle would contain 10, 6, and 35. Show that the product of both sets of numbers will always be the same. For instance, $5 \times 20 \times 21 = 2{,}100$ and $10 \times 6 \times 35 = 2{,}100$. This is sometimes called the *Star of David theorem*, because the two triangles form a star with the original number in the middle.

The Joy of Probability

Lecture 22

If there are only two things that I want you to remember about mean and variance it's this. Most normal random variables have the property that the probability that you are within 1 standard deviation away from the mean is about 68%; 2/3 is easy to remember. There is a 95% chance that you are within 2 standard deviations of the mean.

The easiest events to understand are those that have equally likely outcomes. For instance, the flipping of a coin has two possible outcomes, heads or tails. The probability of either outcome is 1/2. Probability is expressed as a number between 0 (impossible) and 1 (certain). In rolling a fair six-sided die, there are six possible outcomes, each of which has an equal probability of occurring. The probability of rolling any specific number is 1/6; of rolling an even number is 3/6, or 1/2; and of rolling a number that is 5 or larger is 2/6, or 1/3.

There are eight sequences in which you can flip a coin three times, and each of those sequences is equally likely. Once you've flipped two heads, for example, the chance that the third flip is a head is still 1/2. There are eight possible equally likely outcomes, but there is only one way to flip three heads, so the probability of that outcome is 1/8. The probability of flipping two heads or one head is 3/8. The probability of flipping all tails is 1/8. In general, if you flip a coin n times, you have two equally likely possibilities for the first outcome, the second outcome, the third outcome, and the n^{th} outcome; therefore, there are 2^n different ways of flipping the coin n times. How many of those ways of flipping the coin result in exactly k heads? Among those n coin flips, choose k of them to be heads; the other ones will have to be tails. The number of ways of picking k heads is $\binom{n}{k}$. The probability of flipping k heads is $\binom{n}{k}/2^n$.

What is the probability that at least two people in a group of n people will have the same birth month and day? With just 23 people, there is at least a 50% chance that two people in the room will have the same birthday. To see why that's true, let's answer the negative question: What's the probability that everyone in the room has a different birthday? In other words, what are the equally likely events in this situation? If we write down lists of birthdays for everyone in the room, how many possible lists could we create? We'd have 365 choices for the first list, 365 choices for the second, and 365 choices for the last. The total number of lists that are possible is 365^n.

The notion of *independence* is important in probability problems.

How many ways can we create lists in which all the birthdays are different? There would be 365 choices for the first birthday, 364 choices for the second, 363 choices for the third, and so on, down to $366 - n$ choices for the last one. The probability that all those birthdays are different would be $365 \times 364 \times 363 \times \ldots (366 - n)/365^n$. The probability that there's at least one match among those people is $1 - \dfrac{365!}{365^n(365-n)!}$. If we plug in some numbers, we find that the probability of a birthday match with 10 people is 12%. With 20 people, the probability is greater than 40%, and with just 23 people, the probability is 50.7%. With 100 people, the probability of a birthday match is 99.99996%.

The notion of *independence* is important in probability problems. Two events, A and B, are independent if the occurrence of A does not affect the probability that B will occur. For example, the outcome of a coin flip has no influence on the outcome of a roll of a die. For independent events, the probability of A and B is the probability of A times the probability of B, or $P(A \text{ and } B) = P(A) \times P(B)$. For example, the probability of flipping heads is 1/2; the probability of rolling a 3 is 1/6. The probability that both events will occur is their product, 1/12.

What's the probability of rolling five 3's in a row? The probability of rolling the first 3 is 1 out of 6, and the probability of rolling each of the other 3's is also 1 out of 6. Because each of those rolls is an independent event, the probability of rolling five 3's in a row is $1/6^5$. What's the probability of rolling the first five digits of pi in order? Even though this sequence seems more random than the previous one, the probability of rolling this specific sequence is also $1/6^5$. The probability of rolling the numbers 1, 2, 3, 4, and 5 in that order is $1/6^5$. But if we allow any order, each sequence has a probability of $1/6^5$. We can arrange the numbers 1 through 5 in 5!, or 120, ways. Thus, the probability of rolling 1, 2, 3, 4, 5 in any order would be $5!/6^5$.

If I roll a six-sided die ten times, what's the probability that I will roll a 3 exactly two of those times? I could roll a 3, then another 3, then 8 numbers that are not 3s. The probability of that sequence is 1/6 for the first 3, 1/6 for the second 3, and 5/6 for each succeeding number that is not a 3. Thus, the probability of seeing the specific outcome of a 3, followed by a 3, followed by 8 non-3's would be $1/6^2(5/6)^8$. However, there are $\binom{10}{2}$ ways of rolling two 3's., or $\binom{10}{2}$ sequences that have a probability of $1/6^2(5/6)^8$; therefore, the answer to the original question is $\binom{10}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8$.

This is an example of a *binomial probability problem*, one of the most important kinds of problems that appear in probability. In general, when we perform an experiment, such as flipping a coin, n times, each of those experiments has a success probability of p. The number of successes, such as the number of heads, is x (called a *binomial random variable*, meaning that it has two possibilities). In that situation,

$$P(x=k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

Let's look at a *geometric probability question*: Suppose I roll a six-sided die repeatedly until I see a 3. The probability that the first 3 will appear on the 10^{th} roll is $(5/6)^9(1/6)$.

Let's now switch our attention to problems involving *dependence*. For these problems, we need to know the *conditional probability formula*: The probability of A given B is the probability of A and B divided by the probability of B, or $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$. Let's say I roll a six-sided die and the outcome of that roll is x. Find the probability that x is equal to 6 given that x is greater than or equal to 4. We know that the probability of getting any particular outcome is 1/6, but the probability that the outcome will be greater than or equal to 4 is 1/3. The formula gives us that same conclusion. According to the formula, the probability that $x = 6$ given that x is greater than or equal to 4 is

$$P(x = 6 \mid x \geq 4) = \frac{P(x = 6 \text{ and } x \geq 4)}{P(x \geq 4)}.$$

The numerator has redundant values: $x = 6$ and $x \geq 4$; thus, we can rewrite the numerator as $P(x = 6)$. In the denominator, the probability that $x \geq 4$ is 3/6. Therefore, the probability is

$$\frac{1/6}{3/6} = \frac{1}{3}.$$

What about the probability that x is even given $x \geq 4$? Using the same idea, that's

$$\frac{P(x \text{ is even and } x \geq 4)}{P(x \geq 4)} = \frac{P(x = 4 \text{ or } 6)}{P(x \geq 4)} = \frac{2/6}{3/6} = \frac{2}{3}.$$

If A and B are independent events, the conditional probability formula tells us that the probability of A happening given B happens is the probability of A and B divided by the probability of B. But because A and B are independent, the probability of A and B is the probability of A times the probability of B divided by the probability of B:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A),$$

which agrees with our notion of independence.

Another important concept in probability is *expected value*. The expected value of a random variable x, which we denote $E[x]$, is the weighted average value of all the possible values that x can take on. Specifically,

$$E[x] = \sum kP(x = k).$$

Let's say x could take on three values: 0 with a probability of 1/2, 1 with a probability of 1/3, or 2 with a probability of 1/6. $E[x]$ is a weighted average of the numbers 0, 1, and 2, where those weights are the probabilities. In this case, $E[x] = 0(1/2) + 1(1/3) + 2(1/6) = 2/3$.

Expected values have some properties that we might…expect. For instance, if a is a constant, $E[ax] = aE[x]$. The expected value of $x + y$ is the expected value of x plus the expected value of y: $E[x + y] = E[x] + E[y]$. That's true even if we add n random variables. In other words, the expected value of the sum is the sum of the expected value:

$$E[x_1 + x_2 + \ldots + x_n] = E[x_1] + E[x_2] + \ldots + E[x_n].$$

That's true for any random variables, independent or not. Now we can apply the expected value of the sum to derive the expected value of a binomial random variable. Suppose I flip a coin n times, each with heads probability p, and x is the number of heads that I get. What is the expected number of heads when I perform this experiment n times? Your intuition might tell you that if p is 1/2, and I flip the coin n times, we expect about half the results to be heads. If the probability of heads is 2/3, then we expect the number of heads to be 2n/3. Thus, $E[x] = np$. We can derive this using an easy method that looks at each individual coin flip. Here's the easy method: x_i is equal to 1 if the i^{th} flip is heads and 0 if it's tails. In other words, $x_i = 1$ with probability p and $x_i = 0$ with probability $1 - p$. Then, the total number of heads will be $x_1 + x_2 + x_3 + \ldots + x_n$. In other words, we're just counting the 1's. Thus, $E[x_i] = 1(p) + 0(1 - p) = p$. In this way, $E[x]$, which is $E[x_1] + E[x_2] + \ldots + E[x_n]$ (the expected value of the sum is the sum of the expected value) is equal to $p + p + p + p + \ldots + p$, a total of n times, which is np.

Variance measures the spread of x. If $E[x] = \mu$ (as in *mean*), then the variance of x (Var(x)) is defined as $E[(x - \mu)^2]$. In other words, the measure of the spread is the expected squared distance from the mean. The standard deviation of x is the square root of that quantity. Here are some handy formulas for variance and standard deviation: Though we defined the variance of x in one way, in practice, it's often easier to calculate it as $E[x^2] - E[x]^2$. For example, in the problem we saw earlier, if the probability that $x = 0$ is 1/2, the probability that $x = 1$ is 1/3, and the probability that $x = 2$ is 1/6, then $E[x^2]$ is a weighted average of all the possible values of x^2, which is $1/2(0^2) + 1/3(1^2) + 1/6(2^2) = 1$. As we saw earlier, $E[x] = 2/3$; thus, $\text{Var}(x) = E[x^2] - E[x]^2 = 1 - (2/3)^2 = 5/9$. Another property of variance that is worth knowing is as follows: If x_1 through x_n are independent random variables, then the variance of the sum is the sum of the variances.

So far, we've been dealing with discrete random variables, questions that have nice integer answers. But many random processes have continuous answers. We can address continuously defined quantities using calculus. We describe the probability of continuous quantities by a *probability density function*, a curve that stays above the x-axis and whose area under the curve is 1. With this function, the probability that x is between a and b is the area under the curve between a and b. Let's use a probability density function of $x^2/9$. This is a legal probability density function since $\int_0^3 \frac{x^2}{9} dx$. The probability that x is between 1 and 2 is

$$\int_1^2 \frac{x^2}{9} dx = \frac{7}{27} .$$

Continuous random variables have similar formulas to discrete random variables. For instance, the expected value of x if x is a continuous random variable, instead of being a weighted sum of the possible values of x, is a weighted integral of the possible values of x. Specifically, $E[x]$ is the integral of x times the density function of x with respect to x. Similarly, to find the expected value of x^2, we take the integral of x^2 times the density function of x.

Perhaps the most important continuous random variable of all is the normal distribution—the original bell-shaped curve. The most famous of these is the bell-shaped curve that has a mean of 0 and a variance of 1, but these curves can have different sizes. The most general bell curve has a mean of μ and a variance of σ^2. This has a rather imposing probability density function:

$$\frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2\sigma^2}(x-\mu)^2} .$$

Every normal distribution has the following property: The probability that a continuous random variable is within one standard deviation of its mean is about 68%, within two standard deviations, about 95%. The fact that the normal distribution is so common stems from the *central limit theorem*, which says that if we add up many independent random variables, we always get approximately a normal distribution. In the coin flip, where a single coin flip could be heads or tails (probability 1/2), we can show that the expected number of heads in a single coin flip is 1/2. The variance of a single coin flip is 1/4. If we flip a coin 100 times, the expected number of heads is 50. The variance of the number of heads is $100 \times 1/4 = 25$. Thus, the standard deviation is 5. Though we can't predict the outcome of a single coin flip, we've got a good handle on the outcome of 100 coin flips. That is, the outcome has an expected value of 50 and a standard deviation of 5. Since this has an approximate normal distribution, there is about 95% chance that the number of heads will be between 40 and 60, that is, $50 \pm 2(5)$. We'll see how to exploit more of this kind of information in our next lecture on mathematical games. ■

Suggested Reading

Burger and Starbird, *The Heart of Mathematics: An Invitation to Effective Thinking*, chapter 7.

Gross and Harris, *The Magic of Numbers*, chapters 8−13.

Questions to Consider

1. If 10 people are each asked to think of a card, what are the chances that at least two of them will think of the same card?

2. In the game of Chuck-a-Luck, you bet $1 on a number between 1 and 6; then, three dice are rolled. If your number appears once, you win $1; if your number appears twice, you win $2; and if your number appears three times, you win $3. (If your number does not appear, you lose $1.) On average, how much should you expect to lose on each bet?

The Joy of Mathematical Games

Lecture 23

Now, I want to say I'm not advocating that you all go out there and start gambling. What I am saying is that if you are going to gamble, you may as well be smart about it.

Let's start with horseracing and Harvey, a horse who likes to run in the rain. If it rains tomorrow, Harvey has a 60% chance of winning the race, but if it doesn't rain tomorrow, he has a 20% chance of winning the race. Our notation for this is: $P(\text{win} \mid \text{rain}) = .60$ and $P(\text{win} \mid \text{no rain}) = .20$. The question is: What's the probability that Harvey will win the race? That answer depends on the actual probability that it will rain. The probability that Harvey will win is the weighted average of the probability that he wins when it rains and the probability that he wins when it doesn't rain. If the probability of rain is 50%, the expression is: $P(\text{win}) = (.60)(.50) + (.20)(.50) = .40$. If the probability of rain is 70%, the expression is: $P(\text{win}) = (.60)(.70) + (.20)(.30) = .48$.

Suppose that the probability of rain on race day is 99%. Harvey's chances for winning now should be almost 60%. We take a weighted average of 60% and 20%, giving 60% a weight of .99 and 20% a weight of .01. The weighted average of those numbers is .596; thus, Harvey has a 59.6% chance of winning, as our intuition told us. The probability that Harvey will win is governed by the *law of total probability*, which states, in general: If an event B has two possible outcomes, B_1 or B_2, then $P(A) = P(A \mid B_1)\text{P}(B_1) + P(A \mid B_2)P(B_2)$. Similarly, if B has n possible mutually exclusive outcomes, B_1 or B_2 or … B_n, then $P(A) = P(A \mid B_1)\text{P}(B_1) + \ldots + P(A \mid B_n)P(B_n)$.

Let's use this formula to analyze the game of craps. To play craps, you roll two dice. Let's call the total of those two dice the number B. If B is 7 or 11, you win immediately. If B is 2 or 3 or 12, you lose immediately. If B is 4, 5, 6, 8, 9, or 10, you keep rolling the dice until you get a sum of B—your original total—or a 7. If a sum of B shows up first, you win, and if a 7 shows up first, you lose. According to the law of total probability, the probability of winning (event A) is $P(A) = P(A \mid B_1)\text{P}(B_1) + \ldots + P(A \mid B_n)P(B_n)$. In craps, the B event is the total of the dice. It's easier to determine the probability of winning at craps overall once we know what the number rolled is, and the law of total probability allows us to break this problem up into more manageable pieces according to the numbers rolled.

We'll put all the information we need in a "craps table" (shown below); then, we can figure out some of these probabilities. How do we find the probability of seeing any particular number? Imagine that one of the dice is green and the other one is red. There are 6 possible outcomes for the green die and 6 for the red die, or $6 \times 6 = 36$ possibilities for the green/red combination. Even though we're only interested in the total of some number between 2 and 12 (and those are not equally likely), we're just as likely to see a green 3 and a red 5 as a green 6 and a red 2. Thus, each of the 36 outcomes has the same probability. Note that there is one way to roll a total of 2. There are two ways to roll a total of 3 (a green 2 and a red 1 or a green 1 and a red 2). All the possible outcomes are listed in the matrix below. To find the number of possible outcomes for each number, we just count the number of times a given number appears out of 36.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Knowing these outcomes, we can now start to fill in our craps table, focusing first on the shaded rows.

B	P(win | B)	P(B)	Product
2	0	1/36	0
3	0	2/36	0
4	1/3	3/36	3/108 = .027777…
5	4/10	4/36	16/360 = .044444…
6	5/11	5/36	25/396 = .063131…
7	1	6/36	6/36 = .166666…
8	5/11	5/36	25/396 = .063131…
9	4/10	4/36	16/360 = .044444…
10	1/3	3/36	3/108 = .027777…
11	1	2/36	2/36 = .055555…
12	0	1/36	0

For instance, the probability of winning given that $B = 2$ is 0; if you roll a 2, you've lost immediately. The probability of winning if you roll a 3 is also 0, as is the probability of winning if you roll a 12. On the other hand, the probability of winning if your first roll is a 7 is 100%, or 1, as is the probability of winning if $B = 11$.

Now we turn to some of the trickier probabilities. For instance, what's the probability of winning given that $B = 4$? There are two ways to answer that question. If the initial roll is 4, you keep rolling the dice until you see either another 4 or a 7. If a 4 shows up before a 7, you win. If a 7 shows up before a 4, you lose. From our matrix, we see that there are three ways out of 36 to roll a 4 and six ways out of 36 to roll a 7. The chance of winning on the next roll after you've rolled a 4 would be 3/36. But what are the chances that you win two rolls after that first roll? You didn't roll a 4 or a 7 on the next roll ($P = 27/36$); then you did roll a 4 or a 7 on the following roll ($P = 3/36$). Multiplying those probabilities, we get (27/36)(3/36). You could win on the next roll—that is, no 7 or 4, no 7 or 4, followed by a 4. That has the probability $(27/36)^2(3/36)$, and so on.

This is an infinite series—a geometric series—and we know how to sum those. We factor out 3/36 to get $1 + 27/36 + 27/36^2 + \ldots$, which has the form of a geometric series, $1 + x + x^2 + x^3 + \ldots$, and we know that equals $1/(1 - x)$. When we do the algebra, we get 1/3 as the probability of rolling a 4 before rolling a 7. Another way of answering that question is a bit more intuitive. If we look at our matrix again, we see that there are three ways to roll a 4 and six ways to roll a 7. Thus, there are twice as many ways to roll a 7 as there are to roll a 4; therefore, it would make sense that you would be twice as likely to roll a 7 before you roll a 4. The only numbers that are relevant to winning in the matrix are the three 4's and the six 7's. One of those will be the first number that you roll, and three of those possibilities allow you to win and six cause you to lose. That's why $P(\text{win} \mid B{=}4) = 3/9 = 1/3$, which agrees with our previous calculation.

Let's use this easier method to answer the next question: What's the probability that you win given that $B = 5$? There are four ways to roll a 5 in our matrix, and there are six ways to roll a 7. What's the probability that the next number you roll is a 5 before you roll a 7? Of the ten possibilities, four of them are good and six of them are bad, so the chance will be 4/10. What about the probability that you win given that your initial roll was a 6 or an 8? Now you've got a better chance of winning because there are five ways to win and six ways to lose with each number; your chance of winning is 5/11. Using this information, we now look at our completed craps table. The law of total probability tells us to multiply column 2 and column 3; the product then goes in column 4. To get the total probability of winning, we add up all those products, which gives us 244/495 = .492929…, or a 49.3% chance of winning and a 50.7% chance of losing.

If you know the rules of craps, you know that you can bet against the shooter. Every time the shooter loses, you win, except if the shooter's initial roll is a double 6. In that case, the shooter loses, but you don't win or lose; the result is called a *push*. That event adds to your losing probability by (1/36)(1/2) = 1/72 = .014, which makes up for the difference in the 49.3% chance of losing and 50.7% chance of winning. Putting these numbers together, the expected value when you bet $1.00 on craps is as follows: 1(.493) −1(.507) = −.014. In other words, if you bet $1.00, then your expected value is −1.4 cents. That doesn't seem like much, but if you play the game long enough, you'll go broke.

The expected value is −1.4 cents. The variance of a single bet is almost $1.00. If you make 100 bets and on average you lose 1.4 cents for every bet, then after 100 bets, you will be down about $1.40. The variance of the sum is equal to the sum of the variances, so the variance after 100 bets will be 100, but the standard deviation—the quantity we most care about—is $\sqrt{100}$ = $10.00. Thus, your expected loss is $1.40, but the standard deviation is $10.00. You're probably going to lose, but there's a chance that you'll still be on the positive side after 100 bets. After 10,000 bets, you'll be down $140. Because the standard deviation grows with the square root of the number of bets, it will be about $100. You now have less than a 20% chance of being in the black after 10,000 bets. After 1,000,000 bets, you will be down $14,000 with a standard deviation of 1,000. You will almost certainly be within two standard deviations of your expected loss; thus, you have a 95% chance of being down somewhere between $16,000 and $12,000; there's a 99% chance that you'll be within three standard deviations—somewhere between $11,000 and $17,000 down.

A game that is easier to look at is roulette. In American roulette, we have 18 red numbers, 18 black numbers, and 2 green numbers—the 0 and the double 0. If you bet on red, you win $1.00 with a probability of 18/38; you lose $1.00 with a probability of 20/38. Your expected value here is (18/38) − (20/38) = −2/38 = −.0526. You will be down about 5 1/4 cents for every bet. After 100 bets, you will be down about $5.26 with a standard deviation of $10.00. After 10,000 bets, you will be down $526 with a standard deviation of $100. Thus you will almost certainly be down somewhere between $200 and $800.

In American roulette, we have 18 red numbers, 18 black numbers, and 2 green numbers.

Let's close with something called the Gambler's Ruin problem. In this problem, with each bet, you win $1.00 with probability p and you lose $1.00 with probability $1 - p$, or q. You begin with d dollars and your goal is to reach n dollars. Let's say d = \$60 and n = \$100. The Gambler's Ruin theorem has a beautiful formula for figuring out your chance of reaching n dollars without going broke: $\frac{1-(q/p)^d}{1-(q/p)^n}$, as long as $q/p \neq 1$. When $q/p = 1$, which happens when p is 1/2, then the answer is d/n. Let's look at the implications of this formula. If you walk into a casino and play a fair game (p = 1/2), what are the chances that you will go from \$60 to \$100 before reaching \$0? The answer is 60%. If the game is fair and you start 60% of the way toward your goal, then you will reach your goal with a probability of 60%.

In a game such as craps, however, where your probability of winning is 49.3%, your chance of reaching your goal is about 28%. If your probability of winning is 49% instead of 49.3%, your chance of reaching your goal goes to 19%. If you play a game such as roulette, where your probability of winning is 47.3% on any given play of the game, you have only a 1.3% chance of reaching your goal without going broke. On the other hand, if you know a little bit of gambling theory, you might be able to play blackjack with a 51% probability of winning, which means you can reach your goal with a probability of 93%. ■

Suggested Reading

Gardner, *Martin Gardner's Mathematical Games*.

Packel, *The Mathematics of Games and Gambling*.

Questions to Consider

1. When dealing cards on a table, what is the probability that an ace will appear before a jack, queen, or king appears?

2. If you are dealt two cards at random from a deck of 52 cards, what is the probability that one of the cards is an ace and the other card is a 10, jack, queen, or king?

The Joy of Mathematical Magic

Lecture 24

Something else I love to play with are magic squares. What I've brought here is, in fact, the smallest magic square you can create using the numbers 1 through 16. If you took the time to verify, you would see that every row and every column adds to the same number—in this case, 34. I've done such an extensive study on magic squares that I propose to create one for you right before your very eyes.

We begin this lecture with a trick that involves phone numbers and seems to be intriguing to many people. You may need a calculator to follow along. Let's call the first three digits of your phone number x and the last four digits y. Here are the steps to follow: Multiply the first three digits by 80: $80x$. Add 1: $80x + 1$. Multiply by 250: $(80x + 1)250$. Add the last four digits of your phone number: $(80x + 1)250 + y$. Add the last four digits again: $(80x + 1)250 + y + y$. Subtract 250: $(80x + 1)250 + y + y - 250$. Simplify and divide by 2: $(20{,}000x + 2y)/2$. Answer: $10{,}000x + y$ = your phone number. When we get to the number $10{,}000x + y$, we're just attaching four 0's to x, then adding the number y, which leaves us with the phone number.

Let's now turn to magic squares. We'll create a magic square using my daughter's birthday, December 3, 1998. In the first row, we write: 12, 3, 9, 8. Adding those digits, we get 32. Now, we have to fill out the rest of the square in such a way that every row and every column adds to 32. The result is on the left below.

12	3	9	8
8	9	11	4
9	10	2	11
3	10	10	9

A	B	C	D
C−	D+	A−	B+
D+	C+	B−	A−
B	A− −	D+ +	C

All the rows and columns in this square add to 32, as do the diagonals, the square in the middle, the squares in each of the corners, and the corners themselves. In fact, the four corners are the original numbers. To create a birthday magic square of your own, suppose that the original birth date had numbers *A*, *B*, *C*, and *D*. Begin by writing *A*, *B*, *C*, and *D*, in every row, column, and diagonal in the arrangement shown on the right above. This kind of magic square, where every row and column has the same four numbers is called a *Latin square*. To make the Latin square a bit more magical, we start with in the lower left-hand corner. We leave the *B* alone, but we change the *C* that's in the third row, second column, to *C* + 1 (designated *C*+). Right now, the first diagonal will not add up correctly, so we fix that by changing *A* to *A*−. With *D*, then, that group adds up correctly. To get all the groups to *balance*, we fill out the rest of the square as shown on the right above. Notice that every row, column, diagonal, and group of four is balanced. We can now go back through this process to fill in the square for the birthday we started with.

Let's see how to do instant cube roots in your head. In order to do this, you first have to memorize a table of the cubes of the numbers 1 through 10.

Here's a mathematical game that was inspired by a TV show: Mathematical Survivor. To keep the game simple, we start with six positive, one-digit numbers. In fact, however, this can be done with any number of numbers, and it will always work. Let's use the first six digits of pi: 3, 1, 4, 1, 5, 9. Choose any two of those six numbers to be removed. If we remove 3 and 5, we're left with 1, 4, 1, 9. To replace the numbers we removed, we multiply the two numbers, add them, then add those two results: 3(5) = 15, 3 + 5 = 8, and 15 + 8 = 23; that becomes the fifth number. Now, we have 1, 4, 1, 9, and 23. We then repeat the process. Let's say we eliminate 1 and 4. We multiply them, add them, then add the results: 1(4) = 4, 1 + 4 = 5, 4 + 5 = 9, leaving the list as 1, 9, 23, and 9. Repeating the process, we remove 9 and 23: 9(23) = 207, 9 + 23 = 32, 207 + 32 = 239. The list is now 1, 9, 239. We then remove 1 and 239: 1(239) = 239, 1 + 239 = 240, 239 + 240 = 479. Now we're left with just two numbers, and when we go through the process, the result is 4,799. Surprisingly, when we start with 3, 1, 4, 1, 5, 9, no matter what order we eliminate the numbers in, we will always end up with 4,799.

We started with the numbers 3, 1, 4, 1, 5, 9. To do the trick, I used numbers that are one greater than the original numbers, in this case, 4, 2, 5, 2, 6, 10. I then multiplied these numbers together, which results in 4,800. From that answer, I subtracted 1 to get 4,799. In general, if we start with the numbers a_1, a_2, … a_n, the mathematical survivor will be: $(a_1 + 1)(a_2 + 1) \ldots (a_n + 1) - 1$. How does this work? Suppose you start with the numbers a_1 through a_n; I start with the numbers $a_1 + 1$ through $a_n + 1$. While you're playing your game, I play a much simpler game. That is, whenever you choose two numbers, I also choose the corresponding numbers, but all I do is multiply mine together. At the end of the game, my numbers are simply the product of all the original numbers that I chose. Notice, however, that every time you replace the numbers a and b with $ab + (a + b)$, I replace $(a + 1)$ and $(b + 1)$ with $(a + 1)(b + 1) = ab + a + b + 1$. My new number is one greater than your new number. That means that our lists begin one number apart everywhere, and they remain one number apart everywhere. For example, if you start with 3, 1, 4, 1, 5, 9, I start with 4, 2, 5, 2, 6, 10. When you replace 3 and 5 with 23 by multiplying, adding, and adding, I simply multiply 4 and 6 to get 24. My 24 is one greater than your 23. Term by term, my list of five terms is one greater than your list of five terms, and that will remain true at each step in the problem. Because I know that my list is guaranteed to be the product of my six numbers, 4,800, then you're going to be left with a number that's one less than mine, 4,799.

We've learned to do all kinds of amazing mental calculations in this course; let's see how to do instant cube roots in your head. In order to do this, you first have to memorize a table of the cubes of the numbers 1 through 10. Here's the table .

$1^3 = 1$	$3^3 = 27$	$5^3 = 125$	$7^3 = 343$	$9^3 = 729$
$2^3 = 8$	$4^3 = 64$	$6^3 = 216$	$8^3 = 512$	$10^3 = 1000$

Notice that each of the last digits in the cubes is different. Also note that when you cube a number, it ends in the same number (for example, 1^3 ends in 1, 4^3 ends in 4), or it ends with 10 minus that number (for example, 2^3 ends in 8, 8^3 ends in 2). Suppose someone tells you that a two-digit number cubed is 74,088. First, listen for the thousands. In this case, it's 74,000. We know that 4^3 is 64 and 5^3 is 125. That means that 40^3 is 64,000 and 50^3 is 125,000. This cube must lie between 64,000 and 125,000, or between 40^3 and 50^3. That tells us that our answer must be 40-something. Because we know the answer is a perfect cube, all we have to do is look at the last digit of that cube, in this case, 8. Only one number when cubed ends in 8, namely, 2. Thus, the last digit of the original two-digit number had to be 2, and 42 must be the original cube root.

Let's do one more example. Suppose I cube a two-digit number and I tell you that the answer is 681,472. Once again, listen for the thousands—that's 681,000. The number 681 is between 8^3 and 9^3, or 512 and 729. That means that the original number must begin with 8. The last digit of the cube is a 2. Only one number when cubed ends in 2, namely, 8. Thus, the original number had to be 88.

I'd like to end, finally, with a card trick. I'm not going to give you the secret to this card trick, but I have confidence that with all the math you've learned in this course, you will be able to figure it out if you watch it a few times. This trick works with the 10's, jacks, queens, kings, and aces from the deck—20 cards. I begin by shuffling the cards to my heart's content—or your heart's content. When you tell me to stop, I will keep the cards in that order, but I will ask you to choose whether I should turn some of the cards face up or face down or whether I pair up some of the cards and keep them in the same order or flip them. In this way, we "randomize" the cards. Finally, I deal the cards out into four rows of five, but you choose whether I deal each row out from left to right or from right to left.

The cards now are in a completely random order. I consolidate the cards by folding the rows together. You choose whether I fold the left edge, the right edge, the top, or the bottom. Recall that when we started this trick, I shuffled the cards to your heart's content. I can tell that your heart was content because if we look at the cards that are now face up, we have here the 10, jack, queen, king, and ace of hearts. I hope in this course you've been able to experience the joy of math, indeed, the magic of math, as much as I have. ■

Suggested Reading

Benjamin and Shermer, *Secrets of Mental Math: The Mathemagician's Guide to Lightning Calculation and Amazing Math Tricks*.

Gardner, *Mathematics, Magic, and Mystery*.

Questions to Consider

1. Suppose I cube a two-digit number and the answer is 456,533. What was the original two-digit number, that is, the cube root?

2. How was the final card trick of this lecture done? Here's a hint: At the beginning of the trick, when did the magician offer the choice of "face up or face down" and when did the magician offer the choice of "keep or flip"? As for why the folding procedure works, you might look at the hint given in the second problem of Lecture 10.

Glossary

algebra: Literally, the reunion of broken parts; the manipulation of both sides of an equation to solve for an unknown quantity.

algebraic proof: Establishing the truth of a statement through algebraic manipulation.

anti-derivative: A function whose derivative is a given function.

axiom: A statement that is accepted without proof, such as: "For any two points, there is exactly one line that goes through them."

binomial probability: If an experiment is performed n times, and each experiment independently has a probability p of success, then this is the probability that exactly k successes will occur; numerically equal to

$$\binom{n}{k} p^k (1-p)^{n-k} .$$

binomial theorem: How to expand $(x + y)^n$; the coefficients of the expansion appear on Pascal's triangle. More precisely, it says:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} .$$

calculus: The branch of mathematics that deals with limits and the differentiation and integration of functions of one or more variables. See also **differential calculus** and **integral calculus**.

central limit theorem of probability: The average of a large number of random variables tends to have a normal (bell-shaped) distribution.

circumference: The perimeter of a circle.

combinatorial proof: Establishing the truth of a statement by counting a set in two different ways.

combinatorics: The mathematics of enumeration; the only subject that really counts.

complex number: A number of the form $a + bi$, where i is an imaginary number.

composite number: A positive number with three or more divisors.

conditional probability: The probability that an event occurs, given that another event has occurred.

cosine: For a given angle a, cosine a, is the x-coordinate of the point on the unit circle associated with angle a.

derivative: The rate of change of a function at a given point.

diameter: The length of a line segment obtained by drawing a line from one side of a circle through the center of the circle to the other side of the circle.

differential calculus: The mathematics of how things change and grow.

differential equation: An equation satisfied by a function and its derivatives. For example, the function $y = e^{kx}$ satisfies the differential equation $y' = ky$.

differentiation: The process of calculating derivatives.

e: A number of "exponential" importance, the number e is equal to 2.71828…, which is the limit of $(1 + 1/n)^n$ as n approaches infinity.

equilateral triangle: A triangle that has three equal side lengths.

Euler's equation: A formula that brings algebra, geometry, and trigonometry together: $e^{ix} = \cos x + i \sin x$. When $x = \pi$, it follows that $e^{i\pi} + 1 = 0$.

exponent: The exponent of a^n is the number n. When n is positive, a^n equals a multiplied n times; when n is negative, a^n equals $1/a$ multiplied n times; $a^0 = 1$.

factorial: The number $n!$ is the product of the numbers from 1 through n.

Fibonacci numbers: The numbers obtained in the sequence 1, 1, 2, 3, 5, 8, 13,…, where each number is the sum of the previous two numbers.

fundamental theorem of algebra: Any polynomial of degree n has at most n roots. This is because any polynomial of degree $n \geq 1$, with real or complex coefficients, can be factored as $c(x-r_1)(x-r_2)(x-r_3)...(x-r_n)$, where $c, r_1, r_2, ..., r_n$ are real or complex numbers and x is the variable.

fundamental theorem of arithmetic: Every positive number can be factored into prime numbers in a unique way.

fundamental theorem of calculus: For any positive function $y = f(x)$, the area under the curve $y = f(x)$ that lies above the x-axis and between a and b is equal to $F(b) - F(a)$ where $F(x)$ is a function with derivative $f(x)$.

geometric probability: If an experiment is performed until a success occurs, and each experiment has probability p of success, then this is the probability that the first success will occur on the n^{th} trial; numerically equal to $(1-p)^{n-1}p$.

geometric series: A useful infinite series that says for all numbers x with absolute value less than 1, $1 + x + x^2 + x^3 + ... = 1/(1 - x)$.

geometry: The mathematics of measurement.

golden ratio (phi): the value $(1+\sqrt{5})/2$ = 1.618…, a number with many beautiful properties; in the limit, the ratio of ever larger consecutive Fibonacci numbers.

harmonic series: The infinite sum 1 + 1/2 + 1/3 + 1/4 + 1/5 + …, which diverges to infinity.

hyperbolic functions: cosh $x = (e^x + e^{-x})/2$ and sinh $x = (e^x - e^{-x})/2$ are called hyperbolic functions because they satisfy $\cosh^2 x - \sinh^2 x = 1$, and therefore, (cosh x, sinh x) is a point on the unit hyperbola. Also, tanh x = sinh x/cosh x. Many relationships satisfied by hyperbolic functions are analogous to ones satisfied by the usual trigonometric functions of sine, cosine, and tangent.

i: The square root of negative one, located 1 unit above zero on the imaginary axis. It is one of two solutions to the equation $x^2 + 1 = 0$, the other solution being negative i.

imaginary number: The square root of a negative number.

induction, proof by: To prove that a statement is true for all positive integers, prove it for the number 1, and show that if it is true for the number k, then it will continue to be true for $k + 1$.

infinite series: The sum of infinitely many numbers. We say that an infinite sum of numbers converges to S means that as you add more and more terms you get closer to S, eventually getting as close as you want.

infinity: The number of numbers, larger than any number. (The more you contemplate it, the more your mind gets <u>numb</u>er!)

integer: A whole number, which can be positive, negative, or zero.

integral calculus: The mathematics of determining a quantity, such as volume or area, by breaking the quantity into very small parts.

integration: The process used to calculate areas and volumes by making use of the fundamental theorem of calculus.

isosceles triangle: A triangle with two sides of equal length.

law of cosines: For any triangle with side lengths a, b, c: $c^2 = a^2 + b^2 - 2ab \cos C$, where C is the angle opposite side c.

law of sines: For any triangle with side lengths a, b, c with corresponding angles A,B,C, $(\sin A)/a = (\sin B)/b = (\sin C)/c$.

law of total probability: The probability that an event A occurs can be determined by first considering whether or not another event B occurs: specifically, P(A) = P(A|B)P(B) + P(A|not B) P(not B), where P(A|B) denotes the probability that A occurs, given that B occurs.

logarithm: The exponent needed to obtain one number from another. More precisely, the base b logarithm of a is the number x that satisfies $b^x = a$. The power of 10 needed to obtain a given number is called the base-10 logarithm; for example, the base 10 logarithm of 1,000 is 3.

modular arithmetic: The mathematics of remainders.

normal distribution: Popularly known as the bell-shaped curve, a random variable with a normal distribution has about a 68% chance of being within one standard deviation away from its mean and about a 95% chance of being within two standard deviations away from its mean.

perfect number: A number that is equal to the sum of all its proper divisors. For example, 6 is perfect because 6 = 1 + 2 + 3.

pi: The ratio of the circumference of any circle to its diameter, denoted by the Greek letter π.

polynomial: A sum of terms of the form ax^n where the number a is called the coefficient and the exponent n must be an integer greater than or equal to zero.

prime number: A positive number that has exactly two divisors, 1 and itself.

probability: The likelihood of an event. An event with probability near 1 is nearly certain; an event with probability near 0 is nearly impossible.

Pythagorean theorem: In any right triangle with side lengths a, b, c: $a^2 + b^2 = c^2$, where c is the length of the hypotenuse.

quadratic formula: The equation $ax^2 + bx + c = 0$ has the solution $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. The word "quadratic" comes from the word for "square."

radian: The angle equal to $180/\pi$ degrees.

radius: The distance from the center of a circle to the edge of the circle; equal to 1/2 the diameter of the circle.

rational number: A number that can be expressed as the ratio of two integers.

reciprocal function: A function times its reciprocal function is 1. For example, the reciprocal of cos x is 1/cos x (also known as sec x).

second-degree equation: A function of the form $y = ax^2 + bx + c$.

sine: For a given angle a, sine a, is the y-coordinate of the point on the unit circle associated with angle a.

tangent: Sine divided by cosine.

theorem: A mathematical truth derivable from axioms and the rules of logic.

trigonometry: The branch of mathematics that deals with the relationships between the sides and angles of triangles.

variable: A non-constant numerical quantity.

variance: Measures how much the values of a variable spread around the mean of that variable. Square root of the variance is known as the standard deviation.

Bibliography

Reading:

Adams, Colin, Joel Hass, and Abigail Thompson. *How to Ace Calculus: The Streetwise Guide*. New York: W. H. Freeman, 1998. A lighthearted but very clear guide to the concepts and techniques of calculus. Highly readable without too many technical details.

Adrian, Y. E. O. *The Pleasures of Pi, e and Other Interesting Numbers*. Hackensack, NJ: World Scientific Publishing, 2006. A collection of beautiful infinite series and products that often simplify to some function of pi or *e*. Designed in a unique format that lets the reader first marvel over the number patterns before presenting the proofs later in the book.

Barnett, Rich, and Philip Schmidt. *Schaum's Outline of Elementary Algebra*, 3rd ed. New York: McGraw Hill, 2004. The Schaum's outlines emphasize learning through problem solving. This book has 2,000 solved problems and 3,000 practice problems.

Benjamin, Arthur T., and Jennifer J. Quinn. *Proofs That Really Count: The Art of Combinatorial Proof*. Washington, DC: Mathematical Association of America, 2003. Most numerical patterns in mathematics, from Fibonacci numbers to numbers in Pascal's triangle, can be explained through elementary counting arguments.

Benjamin, Arthur T., and Michael Shermer. *Secrets of Mental Math: The Mathemagician's Guide to Lightning Calculation and Amazing Math Tricks*. New York: Three Rivers Press, 2006. Learn the secrets of how to mentally manipulate numbers, often faster than you could do with a calculator, and other magical feats of mind.

Blatner, David. *The Joy of Pi*. New York: Walker Publishing, 1997. An entertaining history of the number 3.14159..., filled with numerical facts, trivia, and folklore.

Bonar, Daniel D, and Michael J. Khoury. *Real Infinite Series*. Washington, DC: Mathematical Association of America, 2006. A widely accessible introductory treatment of infinite series of real numbers, bringing the reader from basic definitions to advanced results.

Burger, Edward B. *Extending the Frontiers of Mathematics: Inquiries into Proof and Argumentation*. New York: Key College Publishing, New York, 2007. Artfully crafted sequences of mathematical statements gently guide readers through important and beautiful areas of mathematics.

Burger, Edward B., and Michael Starbird. *The Heart of Mathematics: An Invitation to Effective Thinking*. Emeryville, CA: Key College Publishing, 2000. This award-winning book presents deep and fascinating mathematical ideas in a lively, accessible, readable way.

Conway, John H., and Richard K. Guy. *The Book of Numbers*. New York: Copernicus, 1996. Ranging from a fascinating survey of number names, words, and symbols to an explanation of the new phenomenon of surreal numbers, this is a fun and fascinating tour of numerical topics and concepts.

Cuoco, Al. *Mathematical Connections: A Companion for Teachers and Others*. Washington, DC: Mathematical Association of America, 2005. This book delves deeply into the topics that form the foundation for high school mathematics.

Dunham, William. *Journey through Genius: The Great Theorems of Mathematics*. New York: Wiley, 1990. Each of this book's 12 chapters covers a great idea or theorem and includes a brief history of the mathematicians who worked on that idea.

———. *The Mathematical Universe: An Alphabetical Journey through the Great Proofs, Problems, and Personalities*. New York: Wiley, 1994. Similar to this course, this book contains 25 chapters, each devoted to some beautiful aspect of mathematics. Everything from numbers to geometry to logic to calculus appears in this extremely well-written book.

Gardner, Martin. *Aha!: Aha! Insight and Aha! Gotcha*. Washington, DC: Mathematical Association of America, 2006. This is my first recommendation for a young reader. This two-volume collection (which is not part of Gardner's *Mathematical Games* series [below]) contains simply stated problems and puzzles, with diabolically clever solutions. Adults will enjoy it, too.

———. *Martin Gardner's Mathematical Games*. Washington, DC: Mathematical Association of America, 2005. This single CD contains 15 books by Gardner, comprising 25 years of his "Mathematical Games" column from *Scientific American*. The most recent book from this series is *The Last Recreations: Hydras, Eggs, and Other Mathematical Mystifications* (Springer).

———. *Mathematics, Magic, and Mystery*. New York: Dover Publications, 1956. The original classic book on magic tricks based on mathematics. Includes tricks based on simple algebra and geometry to curious properties of numbers.

———. *The Second Scientific American Book of Mathematical Puzzles and Diversions*. Chicago: University of Chicago Press, 1987. A collection of many of Gardner's early writings, including a chapter on some of the magical properties of the number 9.

Gelfand, I. M., and M. Saul. *Trigonometry*. New York: Birkhauser, 2001. A basic, accurate, and easy-to-read introduction to trigonometry.

Gelfand, I. M., and A. Shen. *Algebra*. New York: Birkhauser, 2002. This algebra book focuses on why things are true and does not simply present a collection of disjointed techniques for the reader to master.

Gross, Benedict, and Joe Harris. *The Magic of Numbers*. Upper Saddle River, NJ: Pearson Prentice Hall, 2004. This book introduces the beauty of numbers, the patterns in their behavior, and some surprising applications of those patterns, while teaching the reader to think like a mathematician.

Kiselev, Andrei Petrovich. *Kiselev's Geometry, Book 1: Planimetry*. Alexander Givental, trans. El Cerrito, CA: Sumizdat, 2006. Assuming only very basic knowledge of mathematics, Kiselev builds the edifice of geometry from the bottom up, supplying both bricks and mortar in the process. The book is very much self-contained.

Koshy, Thomas. *Fibonacci and Lucas Numbers with Applications*. New York: Wiley-Interscience Series of Texts, Monographs, and Tracts, 2001. A comprehensive collection of amazing facts about Fibonacci numbers and related sequences, including applications and historical references.

Livio, Mario. *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number*. New York: Broadway Books, 2002. This book is written for the layperson and delves into the number phi, also known as the golden ratio.

Maor, Eli. *e: The Story of a Number*. Princeton, NJ: Princeton University Press, 1994. The history of the number *e*, its many beautiful mathematical properties, and surprising applications.

———. *To Infinity and Beyond: A Cultural History of the Infinite*. Princeton, NJ: Princeton University Press, 1991. This book examines the role of infinity in mathematics and geometry and its cultural impact on the arts and sciences.

———. *Trigonometric Delights*. Princeton, NJ: Princeton University Press, 1998. A very readable treatment of the history and applications of trigonometry.

Math Horizons. Magazine published quarterly by the Mathematical Association of America (www.maa.org); aimed at undergraduates with an interest in mathematics, with high quality exposition on a wide variety of mathematical topics, including stories about mathematical people, history, fiction, humor, puzzles, and contests.

Meng, Koh Khee, and Tay Eng Guan. *Counting*. River Edge, NJ: World Scientific Publishing, 2002. A user-friendly introduction to counting techniques, accessible at the high school level.

Nahin, Paul J. *An Imaginary Tale: The Story of* $\sqrt{-1}$. Princeton, NJ: Princeton University Press, 1998. The history of the imaginary number *i*, its many beautiful mathematical properties, and surprising applications.

Packel, Edward. *The Mathematics of Games and Gambling*. Washington, DC: Mathematical Association of America, 2006. This book introduces and develops some of the important and beautiful elementary mathematics needed for analyzing various games, such as roulette, craps, blackjack, backgammon, sports betting, and poker.

Paulos, John Allen. *A Mathematician Reads the Newspaper*. New York: Anchor, 1997. What sort of questions does a mathematician think about when reading about current events?

Reingold, Edward M. and Nachum Dershowitz, *Calendrical Calculations: The Millenium Edition*. Cambridge: Cambridge University Press, 2001. A history and mathematical description of every known (and ancient) calendar system.

Ribenboim, Paulo. *The New Book of Prime Number Records*, 3rd ed. New York: Springer-Verlag, 2004. Everything you wanted to know about prime numbers, from their history to open problems.

Selby, Peter H. and Steve Slavin. *Practical Algebra: A Self-Teaching Guide*, 2nd ed. New York: Wiley, 1991. This book is written for people who need a refresher course in algebra, teaching the basic algebraic skills alongside problems with real-world applications.

Thompson, Silvanus P., and Martin Gardner. *Calculus Made Easy*. New York: St. Martin’s Press, 1998. Martin Gardner revised this old classic to be accessible to all readers.

Tucker, Alan. *Applied Combinatorics*, 5th ed. New York: Wiley, 2006. An enjoyable and comprehensive introduction to the art of counting, appropriate at the collegiate level.

Velleman, Daniel J. *How to Prove It*. New York: Cambridge University Press, New York, 2006. The book prepares students to make the transition from solving problems to proving theorems. The book assumes no background beyond high school mathematics.

Wapner, Leonard M. *The Pea and the Sun: A Mathematical Paradox*. Wellesley, MA: A. K. Peters, Ltd, 2005. This book provides a very accessible introduction to the notion of infinite sets and their paradoxical consequences, for example, how an object can be rearranged so that its volume increases.

Internet Resources:

Blatner, David. *The Joy of* π. Everything you ever wanted to know about the mysterious number pi. www.joyofpi.com.

Fibonacci Association. Fibonacci fanatics (like the author) may wish to join the Fibonacci Association. www.mscs.dal.ca/Fibonacci.

Knott, Ron. *Fibonacci Numbers and the Golden Section*. The first Web site for fascinating Fibonacci facts and folklore. www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fib.html.

Mudd Math Fun Facts. A collection of beautiful mathematical facts that can be appreciated by math students of all ages. This site is created by my colleague Professor Francis Su of Harvey Mudd College. www.math.hmc.edu/funfacts/.

"Online Mathematics Textbooks." Lists 65 college-level math textbooks that are available online for free! www.math.gatech.edu/%7Ecain/textbooks/onlinebooks.html.

The Prime Pages. This site contains prime number research, records, and resources. http://primes.utm.edu/.

Weisstein, Eric. *Wolfram Mathworld*. This site proclaims itself to be the Web's most extensive mathematics resource, and that is probably not an exaggeration. This site was created, developed, and nurtured by Dr. Eric Weisstein, who began compiling an encyclopedia of mathematics when he was a high school student. http://mathworld.wolfram.com/.