THE GREAT COURSES®

Topic
Science & Mathematics

Subtopic
Mathematics

What Are the Chances? Probability Made Clear

Course Guidebook

Professor Michael Starbird
The University of Texas at Austin

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

Copyright © The Teaching Company, 2006

Printed in the United States of America

This book is in copyright. All rights reserved.

Without limiting the rights under copyright reserved above,
no part of this publication may be reproduced, stored in
or introduced into a retrieval system, or transmitted,
in any form, or by any means
(electronic, mechanical, photocopying, recording, or otherwise),
without the prior written permission of
The Teaching Company.

Michael Starbird, Ph.D.

University Distinguished Teaching Professor of Mathematics
The University of Texas at Austin

Professor Michael Starbird is Professor of Mathematics and a University Distinguished Teaching Professor at The University of Texas at Austin. He received his B.A. degree from Pomona College in 1970 and his Ph.D. in mathematics from the University of Wisconsin, Madison, in 1974. That same year, he joined the faculty of the Department of Mathematics of The University of Texas at Austin, where he has stayed, except for leaves as a Visiting Member of the Institute for Advanced Study in Princeton, New Jersey; a Visiting Associate Professor at the University of California, San Diego; and a member of the technical staff at the Jet Propulsion Laboratory in Pasadena, California.

Professor Starbird served as Associate Dean in the College of Natural Sciences at The University of Texas at Austin from 1989 to 1997. He is a member of the Academy of Distinguished Teachers at UT. He has won many teaching awards, including the Mathematical Association of America's Deborah and Franklin Tepper Haimo Award for Distinguished College or University Teaching of Mathematics, which is awarded to three professors annually from among the 27,000 members of the MAA; a Minnie Stevens Piper Professorship, which is awarded each year to 10 professors from any subject at any college or university in the state of Texas; the inaugural award of the Dad's Association Centennial Teaching Fellowship; the Excellence Award from the Eyes of Texas, twice; the President's Associates Teaching Excellence Award; the Jean Holloway Award for Teaching Excellence, which is the oldest teaching award at UT and is presented to one professor each year; the Chad Oliver Plan II Teaching Award, which is student-selected and awarded each year to one professor in the Plan II liberal arts honors program; and the Friar Society Centennial Teaching Fellowship, which is awarded to one professor at UT annually and includes the largest monetary teaching prize given at UT. Also, in 1989, Professor Starbird was the Recreational Sports Super Racquets Champion.

The professor's mathematical research is in the field of topology. He recently served as a member-at-large of the Council of the American Mathematical Society and on the national education committees of both the American Mathematical Society and the Mathematical Association of America.

Professor Starbird is interested in bringing authentic understanding of significant ideas in mathematics to people who are not necessarily mathematically oriented. He has developed and taught an acclaimed class that presents higher-level mathematics to liberal arts students. He wrote, with coauthor Edward B. Burger, *The Heart of Mathematics: An invitation to effective thinking*, which won a 2001 Robert W. Hamilton Book Award. Professors Burger and Starbird have also written a book that brings intriguing mathematical ideas to the public, entitled *Coincidences, Chaos, and All That Math Jazz: Making Light of Weighty Ideas*, published by W.W. Norton, 2005. Professor Starbird has produced three previous courses for The Teaching Company, *Change and Motion: Calculus Made Clear*; *Meaning from Data: Statistics Made Clear*; and, with collaborator Edward Burger, *The Joy of Thinking: The Beauty and Power of Classical Mathematical Ideas*. Professor Starbird loves to see real people find the intrigue and fascination that mathematics can bring. ■

Table of Contents

INTRODUCTION

LECTURE GUIDES

Table of Contents

SUPPLEMENTAL MATERIAL

Acknowledgments

These lectures were prepared in collaboration with Thomas Starbird, Ph.D., a principal member of the technical staff at the Jet Propulsion Laboratory, Pasadena, California. Michael and Thomas Starbird were assisted by Nathanael Ringer, a Ph.D. student in Financial Mathematics at The University of Texas at Austin. Thanks to Lucinda Robb, Noreen Nelson, Pamela Greer, Alisha Reay, and others from The Teaching Company, not only for providing excellent professional work during the production of this series of lectures, but also for creating a supportive and enjoyable atmosphere in which to work. Finally, thanks to my wife, Roberta, and children, Talley and Bryn. ■

What Are the Chances? Probability Made Clear

Scope:

Many of the most significant events of our lives involve random chance—the people we meet, the accidents that befall us, the weather, the stock market, the games we play, the professions we fall into. Whether we are assessing the chance of being struck by lightning, the chance of winning the lottery, or the chance that it will rain tomorrow, we are confronted with trying to describe in as precise a manner as possible the likelihood of an outcome that is uncertain. Probability is the study that accomplishes the seemingly impossible feat of giving a meaningful numerical value to the likelihood that an event will occur when we admit that we do not and cannot know what will happen.

The basic strategy of probability is clear and simple. When we flip a fair coin, one of two equally likely outcomes will occur; namely, it will land on heads or tails. Thus, we define the probability of landing on heads as 1 out of 2, that is, 1/2. Or, if we roll a fair die, because there are six equally likely possible outcomes, the probability of rolling any one of them, say a four, is simply 1 out of 6, or 1/6. In other words, when probability involves equally likely outcomes, the concept of probability is simply a matter of counting. However, we soon find that the "simple matter of counting" is often not simple at all and frequently leads to surprises. A famous example is that among any random group of 50 people, there is a 97% chance that two or more of them have the same birthday. Our intuition about the likelihood of events, particularly rare events, often diverges sharply from the truth. As we explore probabilistic surprises, we will refine our intuition about the probability of random events and will learn more specifically what is surprising and what is not. We will learn why coincidences are so common and why we must learn to expect the unexpected.

In no place is the role of probability clearer than in games of chance; thus, we will introduce some of the basic ideas of probability using cards, dice, and roulette. In fact, it was in the arena of gambling that the mathematical investigation of probability first arose. In the 17th century, a gambler by the name of Antoine Gombault, the Chevalier de Méré, sought the advice of leading mathematicians of the day with the goal of improving his ability to make good decisions when playing dice. In answering Gombault's questions, Pierre Fermat and Blaise Pascal developed the fundamental concepts of probability.

Probability is the study of events whose outcomes are random. But randomness is a subtle concept. Events with random outcomes have the property that no particular outcome is known in advance; however, in the aggregate, the outcomes occur with a specific frequency. For example, when we flip a fair coin, we do not know how it will land, but if we flip the coin millions of times, we know that it will land heads up very close to 50% of the time. The distinction between our ignorance about the outcome of a particular trial and our knowing the aggregate behavior of many trials is the peculiar domain of randomness and probability.

Probability has applications in many arenas. For example, randomness and probability are central to the concept of statistical inference. But surprisingly, probability is involved in the solutions to many questions that do not at first appear to contain any element of randomness. For example, there are methods by which one can take a very large number, such as one with several hundred digits, and test whether or not it is prime using methods that involve probability. It is certainly not obvious how randomness and probability could possibly play a role in such a situation, because ultimately, the number is prime or it's not—there is no randomness involved. Another application of randomness and probability occurs in psychology. If we want to train our dogs to respond to a signal and keep responding longest, the best method may be to reward them randomly rather than on any fixed pattern. In this way, the dog always has the hope that the next reward is just one more good deed away. Of course, applying these insights to the treatment of people is most suggestive. Other examples in which randomness and probability arise occur in game theory, the study of strategic decision-making. In game theory, often the optimal strategy is one that involves intentionally including randomness. Optimal business strategies or sports strategies often are probabilistic in nature rather than deterministic. This feature complicates the question of how to judge whether we have adopted a good strategy. When probability is involved, even the very best strategy can have a poor outcome by chance alone.

Einstein's famous quotation, "God does not play dice with the universe," expressed his philosophical resistance to the probabilistic nature of quantum mechanics. Quantum mechanics asserts that subatomic particles are not best described as being in a certain place at a certain time but, instead, are better described with probability distributions, suggesting that an electron has some chance of being at any location in the universe at any moment. In fact, randomness and probability lie at the heart of many of the scientific descriptions of the physical and biological worlds. The basic idea of genetic inheritance is that the parents randomly contribute different genetic material to offspring, which then determines many features of the children. Evolution relies entirely on probabilistic occurrences. But we do not need to look to grand scientific theories to find examples of probability. We see probability in the newspaper every day when we read a weather report that says there is a 30% chance of rain. We'll see what that statement actually means.

Probability is a fascinating study that has many real-world applications. But one of the most intriguing aspects of all is that the basic meaning of probability in the real world is not clearly agreed upon by probabilists. In a rough sense, some view probability as measuring an individual's assessment or belief of the likelihood of a future event, while others view the probability of a future event as a fact independent of any individual's opinions. Another kind of distinction is that some probabilists allow probability to be applied to statements that do not entail randomness, such as "There was life on Mars," whereas others feel that probability should refer only to repeatable events with random outcomes. The different views of probability are intriguing to consider and, in some cases, have practical implications. Probability presents us with a rich field of intriguing inquiry that contains questions and insights that are mathematical, practical, and philosophical. ■

Our Random World—Probability Defined

Lecture 1

It would be nice to say, "Well, our challenge in life is to get rid of uncertainty and be in complete control of everything." That is not going to happen. One of life's real challenges is to deal with the uncertain and the unknown in some sort of an effective way; and that is the realm of probability.

In many arenas, our understanding of our world involves processes and outcomes that we view as the result of random chance. We read in the newspaper that there is a 30% chance of rain. We talk about the chance of winning the lottery. Over the last century, scientific descriptions of the world have increasingly included probabilistic components. In quantum mechanics, the very location of subatomic particles is viewed as a matter of probability. The central concept of genetic inheritance and evolution is the random transmittal of genetic material from parents to offspring. Random happenings are those whose individual outcomes we do not or cannot know in advance but that will display regularity in the aggregate. The amazing accomplishment of probability is to put a meaningful numerical value on things we admit we do not know. Our challenges in this course are to understand what that numerical measure of chance is, to develop an intuition about probability in real-life situations, and to see a myriad of applications of probability in games, science, business, and many other aspects of life.

What are the chances? If you buy a lottery ticket, what are the chances that you will be rich? If you walk across a golf course on a stormy day, what are the chances that you'll be hit by lightning? If you bet on red in roulette, what are the chances you'll win? If you buy stocks and bonds, what are the chances those investments will pay off? If you have a fever and other symptoms, what are the chances you have a serious disease? A hurricane is spotted off the East Coast. What are the chances that it will cause great damage? What are the chances that a child brought up by a drug addict will become a criminal? What are the chances that an e-mail advertisement will lead to a sale? All these examples are real-life situations in which we are confronted with possibilities whose outcomes we do not know.

Dealing with the uncertain and the unknown is the realm of probability. One of life's challenges is to deal with the uncertain and unknown effectively. Probability accomplishes the amazing feat of giving a meaningful numerical description of the uncertain and unknown. It gives us information to act on. Probability decisions can be as inconsequential as deciding whether or not to take an umbrella if there is an 80% chance of rain. Making medical decisions based on probability, however, can have life-and-death consequences.

In many arenas, our understanding of our world involves processes and outcomes that we view as the result of random chance. Over the last two centuries, scientific descriptions of our world increasingly include probabilistic components. Physics, from thermodynamics to quantum mechanics, involves questions of probability—molecules moving randomly around and causing things to happen by the aggregate force of probabilistic occurrences. In biology, genetics and evolution are both based on random behavior. Often, underlying random behavior manifests itself in predictable, measurable observations. Scientific descriptions frequently are probabilistic analyses of random occurrences. The prevalence of probabilistic components of scientific descriptions represents a major paradigm shift in our concept of what scientific explanations are.

Probability describes what we would expect from random phenomena if they were repeated many times. But the concept of randomness is subtle. Outcomes of individual random events are unknown, but the aggregate behavior of random events is predictable. The amazing accomplishment of probability is to put a meaningful numerical value on things we admit we do not know. When we roll a fair die, we do not know which side will land uppermost on any individual throw. However, if we roll 60 dice, we expect that each side would land up about 1/6 of the time. One of the difficulties of probability is that we expect a certain result on average, but we also expect to be off by a little. When we roll 60 dice, we do not expect each number to appear exactly 10 times. One of the challenges of this course is to understand what to expect from randomness. A principal goal of probability is to give a numerical measure of chance. We will see a myriad of applications of probability in games, science, business, and many other parts of life.

The course is organized as follows: In this lecture, we will introduce the basic idea of probability. Lecture 2 explores the question: What is randomness? Lecture 3 is about expected value. Expected value is a numerical measure that assesses the value of various possible outcomes to a probabilistic occurrence. Expected value is useful in making decisions, such as those involving investments or other risks. Lecture 4 takes us on a random walk, in which the direction we take at each step is randomly selected. Random walks have applications in physics, biology, and finance. Lectures 5 and 6 show us that randomness and probability are central components of modern scientific descriptions of our world in physics and biology. Lecture 7 explores the world of finance, particularly probabilistic models of stock and option behavior. Probability can be used to find answers to questions that seem to have no random or probabilistic component to them. Lecture 8 explores unexpected applications of probability. Lectures 9 and 10 discuss conditional probability and some surprisingly counterintuitive examples of probabilities. One view of probability is that it can describe a level of belief. Lecture 11 explores this perspective and the Bayesian view of probability. In the final lecture, we will see some probabilistic conundrums that arise when there are infinitely many possible outcomes to a random trial. We end by reviewing how widely probability is applied in the world.

We begin our investigation of probability with gambling. Gambling presents some clear examples of randomness. It was in the arena of gambling that the mathematical investigation of probability first arose. In the 17th century, a gambler by the name of Antoine Gombault, the Chevalier de Méré, sought the advice of Pierre Fermat and Blaise Pascal, who developed the fundamental concepts of probability. A die has six sides. In a fair die, we presume that after rolling the die, any one of the sides is as likely to arise as any other. To give a numerical measure to the probability of a fair die coming up with a five, say, we note that there are six equally likely possible outcomes; a five is one of these outcomes, so its probability of arising is 1 out of 6, or 1/6. In general, if there are n equally likely outcomes, then the probability of one of those outcomes occurring is $1/n$. Gambling games present us with examples in which there are finitely many possible outcomes to the probabilistic occurrence, that is, *discrete probability*.

The concept of probability arising from dice and coin examples leads us to some basic definitions and observations about discrete probability. An *outcome* is a possible result of a single trial, observation, or experiment that we are considering. An *event* is a set of outcomes. For example, if we consider rolling a die, getting a five is an outcome. Rolling an even number is an event. Probability 1 (or, equivalently, 100%) means that the event is certain. Probability 0 means that the event will not happen. If we add up all the probabilities of all the possible outcomes of a trial, we get 1. If the probability of an event is p, then the probability of the event's not occurring is $1 - p$. For example, the probability of rolling a fair die and getting a 5 is 1/6, so the probability of rolling a fair die and getting something other than five is 1 – 1/6 = 5/6. In practice, it is often easier to measure the probability that an event does not happen; for this reason, we will use the $1 - p$ observation frequently.

Often, underlying random behavior manifests itself in predictable, measurable observations.

The basic principle of probability is simple when dealing with equally likely outcomes. Simply count how many total outcomes are possible, count how many are in the event you are considering, and divide. The problem is that "simply" counting is not simple. Let's think about poker. The value of hands is really an ordering of the probabilities of getting the hands. What is the probability of getting all four aces when dealt five cards? To compute the probability of being dealt all four aces, we need to count the total number of five-card hands and compute the total number of hands that contain all four aces. Here are the answers: The number of possible hands containing all four aces is 52 – 4, or 48. The 4 represents the four aces, leaving only 48 cards that could be the fifth card in a five-card hand. We can also calculate the number of possible hands: 52 × 51 × 50 × 49 × 48 = 311,875,200. But some of those hands will have the same cards, only in a different order; thus, we calculate the total number of different orderings of the five cards: 5 × 4 × 3 × 2 × 1 = 120.

The number of distinct five-card hands is (52 × 51 × 50 × 49 × 48)/(5 × 4 × 3 × 2 × 1) = 311,875,200/120 = 2,598,960. The probability of getting four aces is computed by dividing the total number of possible hands with four aces (48) by the total number of possible hands: 48/2,598,960 = 0.00002. To compute the probability of being dealt a straight (see Glossary for definition), we need to count the total number of five-card hands and compute the total number of hands that contain a straight. Here are the answers: The number of possible hands is 2,598,960. The number of possible hands containing a straight is 10,200. The probability of getting a straight is 10,200/2,598,960 = 0.004. To compute the probability of being dealt a flush, we need to count the total number of five-card hands and compute the total number of hands in which all the cards are in the same suit. (Again, straight flushes are not counted as flushes.) Here are the answers: The number of possible hands is 2,598,960. The number of possible hands containing a flush is 5108. The probability of getting a flush is 5108/2,598,960 = 0.002. Because the probability of being dealt a flush is less than the probability of being dealt a straight, a flush beats a straight in poker.

In summary, if you have an experiment or a trial that has equally likely outcomes, to compute the probability of some event, you count the number of outcomes in the event and divide by the total number of outcomes possible. That fraction is the probability of that event. ■

Suggested Reading

Edward B. Burger and Michael Starbird, *Coincidences, Chaos, and All That Math Jazz: Making Light of Weighty Ideas*.

———, *The Heart of Mathematics: An invitation to effective thinking*, 2^{nd} ed.

Ian Hacking, *The Taming of Chance*.

Sheldon Ross, *A First Course in Probability*.

Questions to Consider

1. Do you think that probability will play an increasing or decreasing role in explanations in science, business, social science, and other fields as they continue to develop?

2. Three couples, that is, six individuals, are seated randomly around a round table. What is the probability that the members of at least one couple are seated next to each other?

The Nature of Randomness

Lecture 2

The basic goal of probability is to describe what it is that we should expect from randomness, and so in this lecture we're going to try to undertake an understanding in some detail of the nature of random processes.

What is random? Can we ascertain whether phenomena in the world are best described by randomness or are better described by finding some underlying deterministic reason for what we observe? Questions about what is random arise in considerations of everything from a coin toss to dots on a page, stars in the sky, or the digits of π. Trying to produce lists of numbers that appear random is an unexpected challenge. If we look at a list of digits, can we determine whether or not they were generated by a random process? Many tests about randomness can ferret out the signature of nonrandom generation. One of the paradoxes of randomness is that within the random, we will find surprising instances of patterns that occur by chance alone.

One goal of probability is to describe what to expect from randomness. The challenge is to understand in some detail the nature of random processes. Surprisingly, clear order comes from random activities. Randomness refers to situations in which we don't know any individual result, but we have a sense of what will happen in the aggregate, that is, if an experiment or a trial is done over and over again. This idea is captured in a theorem called the *Law of Large Numbers*. We can illustrate this theorem ourselves by doing various experiments, such as rolling a die and calculating the percentage of times we roll a three. The more times we roll the die, the closer we come to the predicted probability of rolling a three, 1/6, or 0.1667. Throwing the die 6 times, we might get no threes, but in rolling the die 60,000 times, we come very close to the expected 0.1667.

The Law of Large Numbers works even when referring to relatively rare events If we draw three cards at random from each of three decks, the probability that the three cards will be identical is quite small: 1/52 × 1/52 = 1/2704, or 0.00037. After 2704 trials, we got no such matches, but after

2,704,000 trials, we got 1037 such matches: 1037/2,704,000 = 0.00038, very close to the probability.

There are counterintuitive aspects of what is produced by randomness. A visual example illustrates this phenomenon: Working with a square, we pick a place on the vertical axis at random and on the horizontal axis at random and put a dot there. We do this 12 times to produce 12 dots. We expect the dots to be more evenly distributed rather than the clusters and gaps we see. We can also see other patterns in random arrays. Look at the night sky, for example, and see the various constellations that have been identified for centuries. Flipping a coin also illustrates randomness. First, we flip a coin and record the results, heads (Hs) and tails (Ts), over 200 flips. Then, we ask a human being to write down a random list of 200 Hs and Ts. Strings of repeated Hs or Ts in the flips show up more often than in the human-generated list of HTs. Specifically, when you flip a coin 200 times, the probability of having a string of six Hs or six Ts is more than 96% and of having a string of five Hs or Ts is 99.9%. Our simulation shows that even if you have flipped 10 Hs in a row, the next flip is just as likely to be H again as it was the first time you flipped the coin. The coin has no memory.

One goal of probability is to describe what to expect from randomness.

Rare events are expected in probability. As we have seen, the probability of getting any particular five-card hand from a deck of cards, whether an ordinary hand or a royal flush, is 1/2,598,960. The probability of winning the Powerball lottery is 1/146,000,000, but someone is very likely to win. Even very rare events are almost certain to happen given enough opportunities. In 1929, the astronomer Sir Arthur Eddington wrote, “If an army of monkeys were strumming on typewriters, they might write all the books in the British Museum.” It is said, then, that if monkeys randomly type, they will eventually write *Hamlet*. Let’s look at this further. If, since the time of the Big Bang, a billion 18-character patterns were generated per second on a 100-key keyboard, chances are less than 1/1,000,000,000 that “To be or not to be” will be generated. An enterprising author made money with an observation a few years ago when he wrote *The Bible Code*. For example, he found that if he looked at every 1945$^{\text{th}}$ letter somewhere in the Bible, it spelled out “Atomic holocaust, Japan, 1945.”

Mathematicians found *mail* and *bomb* in Ted Kaczynski's manifesto.When we look retrospectively, things that appeared to be random can be explained. Stock movements can be explained in retrospect. Some psychics and stock analysts make correct predictions by chance alone.

How can we distinguish a set that was created from a random process versus some other method? The strategy is to analyze what patterns we would expect to occur by random chance. Suppose we consider flipping a coin. Roughly half the results should be Hs and half Ts. As we flip more coins, that fraction should get closer and closer to 50%. We can get more refined and determine what fraction of HHs or TTs we should expect and so forth. We can compute the probability of each pattern. By seeing whether the appropriate frequency of that pattern appears or does not appear, we gain evidence about the likelihood that the list of Hs and Ts was generated randomly.

Some examples bring up challenging philosophical questions about the meaning of randomness. Consider the first 10,000 digits of π. The digits look random from the point of view of the tests concerning the existence of patterns, yet we know they are completely determined.

Digit	Number of Appearances in the First 10,000 Digits of π
0	968
1	1026
2	1021
3	974
4	1012
5	1046
6	1021
7	970
8	948
9	1014

What kinds of events are actually random in the world and which are deterministic? These are issues that present us with a real philosophical challenge. ■

Suggested Reading

Ivars Peterson, *The Jungles of Randomness: A Mathematical Safari*.

Questions to Consider

1. Do you think that analyzing or modeling some phenomenon as if it were random devalues or depersonalizes the situation? Do you think that such an analysis skirts the actual meaning?

2. On learning that some girl in the neighborhood has committed a minor crime, how do you react to a statement such as: “Well, it was bound to happen; statistics show that about 20% of kids do that”?

Expected Value—You Can Bet on It

Lecture 3

There are consequences to different alternatives of the future, and we have to sort of weigh them.

When we bet money in a gambling game, such as roulette, we know the probability of winning, and we know what our winnings will be if we win. We do not know, however, the specific outcome. If we repeated that exact bet millions of times, we would win a predictable fraction of the time; thus, the average win or loss per bet is a predictable expectation over the long haul. That is to say, while we do not have deterministic regularity, we have statistical regularity. This average win or loss is called the *expected value*. As we saw in the last lecture, the Law of Large Numbers tells us that as random trials are repeated more and more, the fraction of times that a particular outcome occurs will more accurately reflect the probability of that outcome, and thus, the actual average win or loss per bet will become close to the expected value. The concept of expected value allows us to assess the wisdom of various random enterprises that have payoffs or consequences. Betting on red in roulette, buying insurance, or buying a lottery ticket are all susceptible to expected-value analysis. As is common with probability topics, expected-value considerations lead us to some interestingly paradoxical situations. Expected value is our first attempt to understand what kind of regularity these probabilistic experiments have.

Many daily-life decisions involve randomness.

Many daily-life decisions involve randomness. Buying stock, having surgery, studying for a test, and buying insurance all involve making such decisions. How do we make these decisions? We consider hypotheticals and perform a sort of "cost-benefit analysis" for each possible outcome. One math strategy is to start with ordinary thinking and abstract it. As Albert Einstein said, "The whole of mathematics is nothing more than a refinement of everyday thinking." We need to balance the likelihood of the various outcomes with the cost or benefit of each, which leads to the concept of *expected value*.

Let's use gambling, specifically roulette, to look further at this concept. There are 38 possible outcomes in American roulette. Betting \$10 on a single number will pay \$360 for a winning bet. The probability of winning is 1/38; thus, if we place a bet 38,000 times on 13, we should win about 1000 times (and lose 37,000 times). Therefore, we should win a total of \$360,000. However, we would have paid out \$380,000. Our loss is \$20,000; per bet, the average loss is –\$20,000 divided by 38,000 bets, or –\$0.53. Hence, the expected value of the \$10 roulette bet is –\$0.53. *On average*, the bettor will lose 53 cents per bet. Expected value is an average. We have a collection of outcomes, and we have a probability for each outcome's occurring. Each outcome has a value associated with it. In this case, for 13, the value is \$350, and for the other 37 numbers, it is –\$10 (the money bet on the non-winning number). Let O_1, O_2, O_3, … denote the possible outcomes. Let $P(O)$ denote the probability of an outcome and $V(O)$ denote the value of an outcome. Then, the expected value is: $P(O_1)V(O_1) + P(O_2)V(O_2) + P(O_3)V(O_4) + \dots$ and so on through however many possible outcomes you have.

If you bet \$10 on red, your chances of winning are 18/38 and of losing are 20/38. The payout of a \$10 bet on red is \$20, for a gain of \$10. Therefore, the expected value of the \$10 bet is:

$$\frac{18}{38}(\$10) + \frac{20}{38}(-\$10) = -\$0.53$$

Casinos count on the Law of Large Numbers to ensure their profits, as the table of roulette simulations illustrates.

Repetition	**Average Gain in 10,000 Bets**	**Average Gain in 1,000,000 Bets**
1	–0.41	–0.50
2	–0.66	–0.54
3	–0.56	–0.52
4	–0.65	–0.52

Repetition	Average Gain in 10,000 Bets	Average Gain in 1,000,000 Bets
5	–0.41	–0.51
6	–0.70	–0.55
7	–0.56	–0.52
8	–0.44	–0.52
9	–0.51	–0.53
10	–0.58	–0.54

When we made 10 repetitions of 10,000 bets on red, the average is very close to the predicted average loss of –\$0.53. When we made 10 repetitions of 1,000,000 bets on red, the average is even closer to the predicted loss of –\$0.53.

Let us look at unexpected instances of expected value. Suppose someone plays roulette 35 times, betting on one number each time. The expected value of each bet is –\$0.53. And the expected total value of the 35 rounds $= 35\left(\frac{1}{38}350+\frac{37}{38}(-10)\right)$, or –\$18.42. Surprisingly, the probability that a bettor would be ahead after 35 rounds is $1-\left(\frac{37}{38}\right)^{35}$, or 0.61.

However, the bettors who are ahead are only slightly ahead, and the people who are behind have lost \$350. Because the expected value gives weight, the expected value is negative.

Here is another example. Let's say you own a pub and you have a dart game with four rings. You wish to have the payoff be \$4 for hitting the inner circle, \$3 for the next largest ring, \$2 for the next largest, and \$1 for hitting the large outermost ring. You assume anyone who throws the dart has an equal chance to hit anywhere. You calculate the area of each ring and find that the largest has 44% of the area, 31% for the second largest, 19% for the third, and 6% of the area is in the small center.

You can calculate the average expected payoff:

$$0.06 \times \$4 + 0.19 \times \$3 + 0.31 \times \$2 + 0.44 \times \$1 = \$1.87$$

If you decided to make the game completely fair, you would charge \$1.87 per dart thrown, because a fair game is one where the expected value is 0.

Let us consider another unexpected surprise in dealing with the expected value. What is the expected number of rolls of a die until a five appears? If we roll the die 6000 times, we expect 1000 of those rolls to result in a five. The simulation results are very close to 1000. Now we ask what the average gap is between fives in that long list of 6000 numbers. The answer is 6. We have 6000 numbers, around 1000 of which are fives. But what if we take the long list of 6000 numbers and randomly choose any point on that list and ask ourselves what the gap is between fives? What is the expected value of the length of the gap (the number of spaces between two consecutive fives)? The answer is 11, not 6. The reason the answer comes out bigger than 6 is that we are more likely to choose long intervals than short intervals. Likewise, if we cut a string to represent the various lengths on the list between fives and mix the pieces in an urn, we are more likely to choose a longer piece from the urn than a shorter one. ■

Suggested Reading

Edward B. Burger and Michael Starbird, *The Heart of Mathematics: An invitation to effective thinking*, 2nd ed.

Sheldon Ross, *A First Course in Probability*.

Questions to Consider

1. Suppose you play a game with a weighted coin that lands heads up 2/3 of the time and tails up 1/3 of the time. If you are paid \$6 if it lands heads and \$4 if it lands tails, what is the expected value of playing the game once?

2. Expected value does not mean that the expected value is what will happen. When lotteries have very high prizes, the expected value of buying a \$1 lottery ticket can be \$2 or more. Even under those circumstances, why is it not a good investment for you to mortgage your house and buy lottery tickets?

Random Thoughts on Random Walks

Lecture 4

Life, of course, is the source of most mathematical ideas. We look at things that happen in the world, and then we try to abstract from those some principles that become the mathematics that we're trying to develop. This is certainly true in the case of talking about probability and randomness.

Suppose you want to go for a walk, but you feel in a particularly indecisive mood. You decide to walk along a straight north-south road while letting fate decide your direction at each block. You take out a coin and flip it. If it is heads, you walk one block north; if tails, one block south. At each block, you make that random choice. The path you take is called a *random walk*. Many intriguing questions arise in this indeterminate perambulation: Will you ever return home? Will you ever venture 100 blocks away? The analysis of random walks helps us to analyze real-life situations, such as counting ballots during an election, and it explains the sad fate of persistent bettors known as the *gambler's ruin*.

This lecture addresses the phenomenon of random fluctuations. Examples of random fluctuations include the stock market, ballots in an election, coin flipping, genetic drift, and Brownian motion. The simplest example is the random walk. As we leave home (position 0), if we flip a coin and get heads, we go one block north (position 1); if we flip tails, we go one block south (position –1). When we have walked one block, we then flip the coin again and go another block north or south, depending on the result, and so forth. We can see this walk recorded on a graph. How far away do you get? The answer is probabilistic because it depends on flips of a coin. You might also ask, when we take a random walk, what is the probability that, from position 1, we will return to where we started.

To answer that question, we can compute as follows: $P = (1/2) + (1/2)Q$, in which P is the probability that starting at 1, the walk eventually gets back to 0, and Q is the probability that starting at 2, the random walk eventually gets to 0. We can ask what the probability is that starting at position 2, we will return to where we started. To get from 2 to 0, the walk must first get to 1 (probability P), then eventually to 0 (probability P). Thus, $Q = P^2$, and we arrive at the equation:

$$P = \frac{1}{2} + \frac{1}{2}Q = \frac{1}{2} + \frac{1}{2}P^2$$

Working out the equation leads to the result $P = 1$; the probability is 100% that we will indeed return to 0. Although some random walks never return to where you began, the fraction of walks that have not returned becomes closer and closer to 0 as you take longer walks. Thus, the probability of never returning during an infinitely long walk is 0. Another question is: What is the probability that we will eventually get 100 blocks away from where we started? The surprising answer is again $P = 1$.

The *gambler's ruin* is a variation of a random walk. A gambler starts with $2000. Each bet is $200, with even odds. Let's say the game involves flipping a coin, with heads meaning the gambler wins and tails meaning the gambler loses. As we have seen in the random walk, the probability = 1 that you will eventually get back to 0. This means that the gambler will eventually lose everything, even in a fair casino.

Bertrand's *ballot theorem* deals with an election between two candidates in which the winner, A, receives a votes, and the loser, B, receives b votes, where a (52) is greater than b (47). Suppose the votes are tallied by drawing them out of the ballot box one by one, adding 1 to the proper person's score. What is the probability that the eventual winner will always be ahead, from the very first vote counted?

This problem can be rephrased as a graphical problem. Consider the graph whose horizontal axis is time (or ballot number) and whose vertical axis is the amount by which the eventual winner is ahead. The answer to the question of what the probability is that the eventual winner will always be ahead, from the very first vote counted, turns out to be $(a - b)/(a + b)$.

This discussion brings up the question of potential ties. Suppose you wish to hire a tennis pro. Two candidates have played one match against each other each day for the past year, keeping a running tally of how many matches each has won. The tally shows that one player was ahead for the entire last nine months, so that player seems to be better. By comparing this situation with randomness, we can test the strength of that conclusion. Knowing what to expect from randomness informs our interpretation of the results. Let us consider the case of randomness in which, for 366 days in a row, two people flip a coin to win or lose, and let's see where we might expect the last tie to occur. We find, in fact, a surprisingly high probability of one person being ahead for most of the year. In fact there is a 1/2 probability of one person being ahead for the entire last half of the year and a 1/3 probability of one person being ahead for the entire last nine months of the year—by luck alone.

Examples of random fluctuations include the stock market, ballots in an election, coin flipping, genetic drift, and Brownian motion.

If our case were expanded to north-south-east-west, then we would have a two-dimensional random walk. Such a random walk has some interesting properties. We can ask again: What is the chance of returning to the origin? As with a one-dimensional random walk, the probability of returning is 1. However, the rate at which we return is not so quick. In the 30 simulations we ran, it took anywhere from just 4 steps up to more than 100,000,000 steps before we returned to the origin. Peculiarly enough, in the case of a three-dimensional random walk, which allows up or down as an additional choice, we have only a 35% chance of returning to the origin. ■

Suggested Reading

John Haigh, *Taking Chances: Winning with Probability*.

Questions to Consider

1. Suppose as you finish grocery shopping there are two checkout counters open, and both seem to have an equally long line. You pick a line. Does it seem that more often than not you pick the slow line? How does the fact that ties are less frequent than our intuition would predict relate to this situation?

2. Suppose two people play a game where one flips a coin and the other guesses how it will land. If the person guesses correctly, the guesser gets $1 from the flipper; if the guesser is wrong, the flipper gets $1 from the guesser. Suppose the flipper starts out with $10 more than the guesser. What is the probability that at some time in the future, if they play forever, they will have equal amounts of money?

Probability Phenomena of Physics

Lecture 5

Today's lecture concerns the role of probability in descriptions of the physical world. The probabilistic analysis of random behavior lies at the very heart of how we understand physical phenomena, from everything from quantum mechanics to the weather.

Quantum mechanics describes the location of a subatomic particle of physics as a probability distribution. Our intuition would prefer elementary particles to be more like tiny round balls, each of which is at some specific place at each moment. But quantum mechanics suggests that an electron has some chance of being at any location in the universe at any moment. Interestingly enough, Einstein philosophically opposed the probabilistic nature of quantum mechanics. Weather predictions give us probabilistic descriptions of our world that have more obvious consequences. We might read, "There is a 30% chance of rain tomorrow." Then our question becomes: What exactly does that mean?

The probabilistic analysis of random behavior lies at the heart of physical phenomena, from quantum mechanics to the weather. One of the most basic features of understanding the world is that physical matter is made up of atoms and molecules. At the turn of the 20th century, the scientific community was not so clear that atoms and molecules actually existed. It turned out that strong evidence for their existence was an application of probability, and one of the major players in that analysis was Albert Einstein. It was Einstein's theoretical work on Brownian motion that allowed experimentalists to do actual measurements that helped confirm the reality of atoms and molecules. Brownian motion was discovered in the early 1800s by botanist Robert Brown, who made microscopic observations of grains of pollen on the surface of water and noticed that these grains appeared to constantly and randomly move in a jittery way on the surface of the water. In his 1905 paper, Einstein hypothesized that Brownian motion was caused by actual atoms and molecules hitting the grains of pollen, impelling them to take a "random walk" on the surface of the liquid. Einstein wrote down a formula that predicted what distance a piece of pollen would move on average per unit time. Experiments accorded with Einstein's predictions and, thus, were strong evidence for the actual existence of atoms. In a sense, Einstein's work encouraged the mode of reasoning that led to the inherently probabilistic nature of quantum mechanics. In quantum mechanics, the most fundamental objects that make up matter are to be viewed, not as being in one location at one time, but instead, as having a probability of being anywhere. Einstein never accepted the probabilistic nature of quantum mechanics. Probability plays a central role in physical theories, from quantum mechanics up the ladder of different sizes of interacting matter to chemistry and into macroscopic matters, such as the weather, which is where we now turn our attention.

Suppose you read that there is a 30% chance of rain in your region tomorrow. What should that statement mean? First, we need to dispose of the issue of threshold, that is, how much rain is rain. The answer is 0.01 of an inch. Second, what does it mean that there is a 30% chance of rain *at one spot*? It means that on about 30 out of 100 days in which the weather circumstances are like they are today, you would expect at least 0.01 inch of rain *in that spot*.

At the turn of the 20th century, the scientific community was not so clear that atoms and molecules actually existed.

The problem arises when we hear we have a 30% chance of rain *in a whole region*. Because there are different points in the region, we must deal with these variations. The simplest case is if the region is very small and very homogeneous in its character. In that case, the conditions are indeed the same throughout the region for the 30% probability of rain. In another case, though, you might have 50 acres out of a 100-acre region where the probability of rainfall is 40%, and in the other 50 acres, it is 20%. The expected value of the probability of rain for the whole region is 30%. In another case, though, you might have 30 acres out of a 100-acre region where the probability of rainfall is 100%, and in the other 70 acres, it is 0%. Again, the expected value of the probability of rain for the whole region is 30%. In another case, you might have 25 acres out of a 100-acre region where the probability of rainfall is 40%, 50 acres where the probability of rainfall is 30%, and in the other 25 acres, it is 20%. Again, the expected value of the probability of rain for the whole region is 30%.

The consequence to the above conclusions is that on average, the amount of the area that will get rain is 30%. In other words, suppose for each day of 10 days, we knew how many square inches were rained on; we would add all these square inches, then divide by 10 to get 30%. For example, let's use a familiar example where 30 acres out of a 100-acre region always get rain, and in the other 70 acres, it never rains. Again, on average, 30% of the region gets rain. Suppose now that every point in the region has a 30% chance of rain. We can look at each tiny square inch and record rainfall for 10 days there. We would expect to have rain on 3 of those 10 days. Expanding our region to 10 square inches over 10 days, we see that every square inch would expect to be rained on for 3 of the days. Thus, the number of square inches rained on averaged over the 10 days is 3 square inches, which is 30% of the total area. We would get the same result if we looked at a situation where half the region gets 40% chance of rain and the other half, 20%.

The definition of probability of precipitation is tricky. The official definition from the National Weather Service is, at best, misleading: "Technically, the probability of precipitation (PoP) is defined as the likelihood of occurrence (expressed as a percent) of a measurable amount (.01 inch or more) of liquid precipitation (or the water equivalent of frozen precipitation) during a specified period of time *at any given point* in the forecast area. Forecasts are normally issued for 12-hour time periods." The definition should be written: "Technically, the probability of precipitation (PoP) is defined as the likelihood of occurrence (expressed as a percent) of a measurable amount (.01 inch or more) of liquid precipitation (or the water equivalent of frozen precipitation) during a specified period of time *at a random point* in the forecast area. Forecasts are normally issued for 12-hour time periods." A multiple-choice question given to the public to determine if they understand the phrase "The chance of rain is 30%" proves that most Americans do not understand the definition of probability of precipitation. ■

Suggested Reading

Ian Stewart, *Does God Play Dice? The New Mathematics of Chaos*.

Questions to Consider

1. A consequence of quantum physics is that there is a non-zero probability that the moon will spontaneously fall on our heads tomorrow. Why shouldn't we be worried about that possibility?

2. Currently, weather prediction is viewed as a probabilistic enterprise. Do you think that with better knowledge of weather patterns, the randomness will be removed and weather prediction will become deterministic? The theory of mathematical chaos suggests not.

Probability Is in Our Genes
Lecture 6

The basic concept of genetics is that the genetic material from each of the parents is randomly combined; that is, part of the genetic material of the father and part of the genetic material of the mother are combined to become the genetic material for the offspring. And the offspring then have different traits according to which material was contributed by the two parents.

One of the most basic issues in biology is to describe how characteristics of parents are passed on to their offspring. The basic idea is that each parent randomly contributes part of that parent's genetic material to the offspring. The combination of genetic material received from the parents determines characteristics of the offspring. Because randomness is centrally involved in the passing down of genetic material, genetics, the science of inheritance of traits and characteristics, is modeled probabilistically. The simple Mendelian model of dominant and recessive genes provides a probabilistic answer to the question: What traits will the offspring of two specific parents have? Then, probability is used to show the distribution of traits over a whole population and to describe how the characteristics of the whole population will alter through a random process called *genetic drift*. Probability lies at the very core of biological descriptions of mutation and evolution.

Genetics, the science of inheritance of traits and characteristics, is modeled probabilistically. This lecture discusses three probabilistic aspects: the Mendelian model of genetics, genetic drift, and mutation and evolution.

The simple Mendelian model of dominant and recessive genes is the basic model of inheritance. For the sake of simplicity, we will use brown and blue eye color to illustrate this concept, and we will make the simplifying assumptions (though they are not true for real people) that a single gene determines eye color and that there are only two possible colors, blue and brown. The Mendelian model gives a probabilistic answer to the question: What traits will the offspring of two specific parents have? Different versions of a given gene are called *alleles*. In our example, these would be brown (B) and blue (b). People will have BB alleles, Bb alleles, or bb alleles. Each parent contributes one allele for a given gene, either B or b. If either of the alleles in the offspring is the dominant type (B), its trait will be expressed. Otherwise, the recessive trait (b) is expressed. Therefore, the probability of the recessive trait (b) being expressed is 1/4 if both parents carry one recessive allele, as shown in the chart that follows.

Parent	**B**	**b**
B	BB	Bb
b	Bb	bb

The chart below shows the percentage breakdown of the offspring (in the shaded area) if we imagine that 60% of the alleles in the parent population are for brown eyes and 40% are for blue eyes.

Parent Alleles	**B 60%**	**b 40%**
B	**BB**	**Bb**
60%	**36%**	**24%**
b	**Bb**	**bb**
40%	**24%**	**16%**

If we imagine a representative population of 100 offspring, each with two alleles (BB, Bb, or bb), note that the proportion of brown to blue alleles has not changed from the original: 36B + 36B + 24B + 24B = 120 B alleles (60% of 200) 24b + 24b + 16b +16b= 80 b alleles (40% of 200). The *Hardy-Weinberg equilibrium theorem* shows that even if you have a recessive characteristic, it will not disappear. Instead, there is a stable percentage that remains as generations pass. The Hardy-Weinberg equilibrium theorem applies to recessive disorders as long as those disorders do not have an impact on reproductive success.

Probability plays a central role in viewing genetics over the time scale of tens of thousands of years. Genetic drift alters the percentage of alleles that are dominant for a given trait. By random chance, the percentage of dominant alleles in the next generation is different. The expected value of the percentage in the next generation is the same as the percentage in the present generation. But the actual percentage is often a bit different by chance, as our simulations show. This changing percentage is called *genetic drift*, and it can be modeled using the idea of a random walk. Genetic drift is most prominent when the population is small. It happens much more slowly in larger populations. All of this assumes that no natural selection is going on that affects the proportion of the allele. In other words, no trait has an advantage in the number of offspring that a person with that trait can reproduce. If such an advantage exists—for example, if each blue-eyed parent has an extra child, that selective advantage quickly takes over.

The simple Mendelian model of dominant and recessive genes is the basic model of inheritance.

Another way that genetic material changes is through mutations. A mutation is a stable change in the genetic material, brought about by various means, transmitted to offspring. Mutations to nonessential portions of the DNA are useful for measuring time (the molecular clock). It is assumed that mutations to nonessential aspects occur with a uniform probability per unit of time in a particular portion of the DNA. If P is the probability that a single segment of nonessential DNA has no mutations in a year, then P^Y is the probability of no mutations in a segment of DNA happening over Y years. On the average, if you have two individuals who had a common ancestor many generations ago, you would expect them to have about the same percentage P^Y of segments of nonessential DNA that had no mutations. Assuming that mutations are so rare that it is very unlikely that a mutation in the same segment has occurred in the two individuals, the percentage of segments that are mutated in one or the other is, on average, $2(1 - P^Y)$. This is an estimate of the percentage of segments that would be found different if comparing two individuals with a common ancestor Y years ago. Using this kind of probabilistic inference, we can estimate that the most recent common female ancestor of all living humans lived about 150,000 years ago.

Let's look at a hypothetical situation that has a probabilistic aspect: universal HIV testing. About 1% of the time, HIV tests give a false-positive result. Of those who have HIV, their tests will come out positive 95% of the time. If someone has a positive result, what is the probability that that person has HIV? Let's look at the numbers: Let's say the population of the United States is about 300,000,000, of which about 500,000 people are HIV-positive. Of the 500,000 who actually have the disease, the test will come out positive 95% of the time, which equals 475,000 cases. There are 299,500,000 (that is, 300,000,000 – 500,000) people who do not have the disease. Of the 299,500,000 people who do not have the disease, the test will come out falsely positive 1% of the time, which equals 2,995,000 cases. Thus, the total number of people receiving a positive test result is: 475,000 + 2,995,000 = 3,470,000. But of the 3,470,000 who get positive test results, only 475,000 actually have the disease. Therefore, if you get a positive test result, your probability of having the disease is 475,000/3,470,000, which is less than 15%. This is an example of a probabilistic anomaly that is an artifact of giving universal testing for a rare disease when the tests have a significant possibility of giving false-positive results. ■

Suggested Reading

Brian Charlesworth and Deborah Charlesworth, *Evolution: A Very Short Introduction*.

Questions to Consider

1. Assume that for some gene, there are more dominant alleles than recessive alleles in the current population. How can you reconcile the following facts: First, that the expected value for the percentage of a recessive allele in the next generation's population is its current percentage, and second, the percentage of that allele is probabilistically expected to become only half as great as it is now or twice as great as it is now at some point in the future?

2. How could the rate of change in nonessential parts of DNA be used to disprove the theory of evolution if it were false?

Options and Our Financial Future

Lecture 7

Predicting the future prices of stocks can have a significant impact on how we view our whole future financial security. The question is: How are we going to model the behavior of stocks or other financial instruments so that we can have a guess as to whether or not our retirement fund is going to be adequate to keep us living in the lap of luxury?

We've already discussed several applications of probability to gambling, and it seems natural that probability theory would arise in an area where great gambles are made—Wall Street. Predicting the future prices of stocks can have a significant impact on our view of our future financial security. Starting in 1900, a Frenchman, Louis Bachelier, devised the first model of stock prices that involved probability, which is fundamental to modern finance. In this lecture, we will also discuss options contracts, which are fundamental to modern finance. In fact, more money is traded in options than in stocks.

Simply put, an option contract is an agreement between two people that gives one the right to buy or sell a stock at some future date for some preset price. Options are used as speculation, as well as a way to hedge risk, but it is a challenge to derive a rational price for such a contract. For quite some time, option pricing was viewed as a form of gambling. After the Black-Scholes theory was developed, the option price was viewed as an investment. As we will see from the example of Long-Term Capital Management, however, the application of sophisticated probability theory is not without its risks.

The world of finance is full of uncertainty, as is the world of gambling. Among many other financial issues, the future prices of stocks and options are definitely uncertain. If we want to evaluate whether our retirement fund is adequate, we need to consider what might happen to our investments and their values. We can take our financial portfolio and run probabilistic simulations. The probabilistic factors might include inflation or world events. Decisions about how much people are willing to pay for stocks are human decisions that are not predictable. Randomness and probability play central roles in the determination of what our financial future is going to be.

How are prices of stocks or options modeled by financial mathematicians? The model Bachelier devised was basically a starting price plus a random walk. In this model, the price varies purely randomly from its current price with equal likelihood of going upward or downward; underlying trends do not appear in the model. In reality, there may be some reason to believe that an asset will increase in value. For example, consider a cattle ranch that has lots of food and today has a small herd of cattle. We expect growth. The value of that asset will rise. Other assets, such as heating oil and corn, have cyclical trends. More robust, sophisticated models of future stock prices were developed that include a drift component. One model (Samuelson, 1960) incorporates three components: today's price, plus a function that relates to how the stock price is expected to change (the drift), plus a random walk feature (volatility).

If we want to evaluate whether our retirement fund is adequate, we need to consider what might happen to our investments and their values.

An option is a contract that gives the holder of the option certain specified rights. This might be the right to buy or sell a security or a commodity at a specified price on a specified future date. We will talk about the simplest kind of options, namely, a piece of paper that says I can buy one share of XYZ stock on April 30 for $100, even if at that time, XYZ is trading for a higher price. The possibility that XYZ will be worth more than $100 is what gives the option its value. If XYZ is trading for less than $100 at that time, the option is worthless. Options can be used as speculation and as a method to hedge risks. Options used as speculation: If I contract the right to buy stock at a future time at $100, I am betting that the stock will actually exceed that price, so I can resell it at a profit. Options used as a hedge against risks: Let's say I need copper for my business. I have a business plan, and I know I need a certain amount of copper at a certain price. I can buy an option to ensure that, at a future time, I can buy copper at today's price.

How much should someone pay for an option? The idea of finding a rational price for options was developed in the late 1960s and early 1970s and allowed the options market to be created. Let's take an example: I have bought an option that states that if XYZ stock, which now sells for $87, gets to $100 in the future, you pay me $1. To determine how much I should pay to acquire that option, we can work out an expected-value analysis: If I believe the probability of the stock reaching $100 is 90%, then the option would be worth 90 cents. But someone else might feel that the option would be worth only 50 cents. The rational price is one that enables the seller of the option to eliminate the risk and to ensure that he has the resources to pay out the $1 if the stock reaches $100. If another person buys 1/100 share today, then he owns 1/100 of the stock. And if the share reaches $100, the seller of the option can pay the $1 by selling the 1/100 share. Thus, the rational price for the option is the cost of 1/100 share of our $87 XYZ stock today, or 87 cents.

Let us look at another example: Suppose an option is associated with a stock that today is selling for $100 per share, and we are talking about the option to buy a share at $100 one month from today. We make a simplifying assumption: The price will be either $110 or $95 one month from today. This concept of looking at a finite collection of possible future values at discrete moments of time is called the *Cox, Ross, and Rubenstein* (*CRR*) *tree*. The CRR tree can be used to price options. Here, we try to replicate the risk of the option. We are going to buy a certain number of shares of stock and have a certain (negative) amount of cash in our portfolio. The value of our portfolio will be equal to the value of the option in one month's time. In other words, we are trying to quantify the risk itself.

Here's the math:

x = number of shares in the portfolio
d = amount of cash in the portfolio
If the price goes to $110, the option is worth $10.
If the price goes to $95, the option is worth $0.
$110x + d = 10 \qquad 95x + d = 0$

Our solution is:

$$x = \frac{2}{3} \text{ and } d = \$-63\frac{1}{3}.$$

In other words, a portfolio containing 2/3share and owing $63.33 will have the same value as the option one month from now. Thus, the rational value of the option is the cost of 2/3 share ($66.67) minus $63.33, or $3.33.

This type of analysis leads to the *Black-Scholes model*. Before the Black-Scholes model, these contracts were viewed as a pure gamble. The main result of the Black-Scholes theory is that the option price can be viewed as an investment, which led to the establishment of trading houses, such as the Chicago Board Options Exchange, created in 1973.

The application of sophisticated probability theory is, however, not without its risks. In 1994, the hedge fund Long-Term Capital Management (LTCM) began its historic money-making run, using advanced mathematics from top mathematicians. The man in charge was John Meriwether, a legendary head of bond trading of Salomon Brothers in the 1980s. He brought Myron Scholes and Robert Merton to serve on the Board of directors of LTCM. They later won the Nobel Prize in Economics for their work on options pricing. LTCM used complicated mathematical strategies and sophisticated models to trade bond products. In its first three years, to take full advantage of the bond mispricings their models found, LTCM borrowed heavily. In 1998, LTCM collapsed. The Federal Reserve Bank of New York arranged a bailout of several billion dollars by 14 investment banks. ■

Suggested Reading

Roger Lowenstein, *When Genius Failed*.

Questions to Consider

1. Why do the prices of a given stock go down as well as up even when the company is doing well?

2. The future prices of stocks are uncertain. What option and stock portfolio could you purchase to guarantee that you will not lose more money than the price of the option even if the stock price falls dramatically, yet you still reap the benefits of substantial gains in the price of the stock? This is an example in which options are used to hedge against stock decline.

Probability Where We Don't Expect It

Lecture 8

[I]n today's lecture we're going to talk about finding probability in unexpected places—places where you wouldn't expect probability to play a role at all.

Sometimes, probability is centrally involved in solving problems that seem to have no random or probabilistic component to them at all. In mathematics, an example occurs in some methods of determining whether a number is prime or not. Any number is either prime or it's not—there is no randomness involved—yet probabilistic methods can essentially determine whether a number is prime even when the number is far too large for any computer to factor. Randomness and probability are involved in psychology when talking about conditioned behavior. Pigeons rewarded randomly rather than on any fixed pattern will retain their training longest. Strategic decision-making, or game theory, often finds that optimal strategies involve taking one action or another with a certain probability rather than finding one best move. Optimal business strategies or sports strategies often are probabilistic rather than deterministic. Probability pops up in many unlikely places.

Probability can be used to determine to any desired degree of certainty the primality of a natural number with hundreds of digits.

In this lecture, we will talk about finding probability in unexpected places. We start with the world of math. Probability can be used to determine to any desired degree of certainty the primality of a natural number with hundreds of digits. Whether a positive whole number is prime (that is, whether the number is not the product of natural numbers smaller than itself) is clearly not a question with any random or undeterministic feature, yet a method of determining whether it is prime uses randomness and probability.

One way to see if a number is prime is to try to divide into it all smaller numbers. Here is an example of this method of determining that the number 91 is not prime.

Divide 91 by	**Get remainder of**
2	1
3	1
4	3
5	1
6	1
7	0

When we arrive at 7, we see that 91 divided by 7 is 13, with no remainder; thus, 91 is not prime. This strategy, however, would be impossible to use for longer numbers, even with today's computers. The method is effective even when it might be impossible to determine whether or not the number is prime in any known deterministic way.

Another strategy for determining if a number is prime uses *Fermat's little theorem*: Start with a number that is prime, take any number less than that number and raise it to the power of 1 less than the prime, then divide by the prime; you get a remainder of 1. This remainder formula is written $n^{p-1} \equiv 1 \bmod p$. For example, if you start with the prime number 5, then you take any number less than 5 (for example, 2) and raise it to the fourth power (5 – 1), you get 16, and 16/5 = 3, *with a remainder of 1.* Likewise, if you start with the prime number 5, then you take 3 (instead of 2) and raise it to the fourth power (5 – 1), you get 81, and 81/5 = 16, *with a remainder of 1.* If you start with the prime number 5, then you take 4 (instead of 2) and raise it to the fourth power, you get 256, and 256/5 = 51, *with a remainder of 1.*

Let's take a different prime, 7. If we choose 2 as the smaller number, then we find 2^6 is 64, and 64/7 = 9, *with a remainder of 1.* No matter what smaller number we choose, we always have *a remainder of 1.*

This theorem then gives us a way to see if a number is *not* prime. For example, we can prove 9 is not prime:

$$2^8 = 256$$

$$\frac{256}{9} = 28, \text{ with a remainder of } 4$$

Because the remainder is not 1, 9 is *not* prime. In addition, there is a computational simplification using just remainders that speeds up the calculation. We can also use this theorem to test if a huge number is not prime.

The question must be asked, though: Even if we use the number 2, how do we raise it to the required power and find the remainder after dividing? For large values of p, $2^{p-1} \bmod p$ can be cleverly computed by simplifying:

$$2 \times 2 = 2^2$$
$$2^2 \times 2^2 = 2^4$$
$$2^4 \times 2^4 = 2^8$$
$$2^8 \times 2^8 = 2^{16}$$
$$2^{16} \times 2^{16} = 2^{32}, \text{ etc.}$$

If p has 300 digits, it takes only on the order of 1000 such doublings to calculate $2^{p-1} \bmod p$. This is a probabilistic test, however, because some numbers fool it. For example, 341 is a product of 11×31, yet 2^{340} divided by 341 does give a remainder of 1. However, for a randomly chosen 13-digit number, there is a 99.9999985% chance that a number that satisfies this test is prime. Of the 308,457,624,821 thirteen-digit primes, only 132,640 will fool this test!

Probability arises in game theory. Game theory is the mathematical model of strategic decision-making. It is used in economics, business, games, sports, war, and other areas where strategic decisions must be made. Game theory uses the concept of a payoff matrix, which describes the payoffs for each player for each combination of options that the players could choose.

We will study game theory as it applies to football. In football, on the third down with many yards to go for a first down, the usual options are a pass play or a run play. The defending team, then, can defend against the pass or defend against the run. Below is a possible payoff matrix. Each number represents expected yards gained by the offense. The defensive payoffs are understood to be the negative of the numbers:

		Defense Options	
		Defend against pass	*Defend against run*
Offense Options	*Pass*	5	7
	Run	6	1

If the offense always passes, the defense will learn to always defend against the pass. That combination gives an expected value of 5 yards for the offense. But if the offense always runs, the defense will learn to always defend against the run. That combination gives 1 for the offense. Game theory confirms that once in a while, at random, making the unobvious play is the best long-run strategy. According to our calculations, the expected number of yards gained if the offense passes with probability p and the defense defends against the pass is $p \times 5 + (1 - p) \times 6$. The expected number of yards gained if the offense passes with probability p and the defense defends against the run is: $p \times 7 + (1 - p) \times 1$. Our probability of passing is a max/min strategy:

$$p \times 5 + (1 - p) \times 6 = p \times 7 + (1 - p) \times 1.$$

Our conclusion is that the offense should pass 71% of the time (randomly). That combination gives an expected value of 5.3 for the offense, which is a higher value than either of the two pure strategies. Likewise, using the payoff matrix figures again, we find that the defense should defend against the pass 86% of the time (randomly). This is called a *Nash equilibrium*, that is, a strategy whereby no player can get an advantage by unilaterally changing strategy. It was named for John Nash, who won the Nobel Prize for his work on game theory.

Let's turn now to risk management in business, studying how a large NASA project estimates its budget. The project lists all the risks that might incur a cost, with an estimate of both the possible cost and the probability of occurrence. The expected value is the probability of occurrence times the cost. As risks are retired or reevaluated or as new risks are added to the list, the expected value is recomputed. In this way, the project can estimate how much money it should keep in reserve.

Psychologists have learned that randomness can play a valuable role in reinforcing a desired behavior. Giving rewards is an ingredient in training an animal, for instance, a pigeon, to behave in a desired way. The question is, how frequently should you reward the instances of the desired behavior in order to have the conditioning last the longest? If you give a reward for a certain behavior (pecking) every time, at first the pigeon learns but quits rather soon when the reward ceases to appear. The best strategy is to randomly reinforce the behavior. Changing the frequency of rewards in an unpredictable, random way leads to behaviors that are retained for long periods even in the absence of rewards. Applied to humans, this observation may help explain the compulsiveness of some gamblers. ■

Suggested Reading

Edward B. Burger and Michael Starbird, *The Heart of Mathematics: An invitation to effective thinking*, 2nd ed.

Oskar Morgenstern and John von Neumann, *Theory of Games and Economic Behavior* (commemorative edition).

Questions to Consider

1. Here is a payoff matrix in which each player is a driver who can choose to drive on the right or on the left of the street.

	Drive on right	**Drive on left**
Drive on right	70, 70	–100, –100
Drive on left	–100, –100	70, 70

The value (by some measure) is 70 for each player if the two agree to drive on the same side of the street, thus avoiding a crash when going in opposite directions. The value is –100 if they choose different sides and, thus, crash. What are the Nash equilibriums for this payoff matrix?

2. Can you think of an example in your own life where random reinforcement has had a lasting impact on your behavior or attitudes?

Probability Surprises

Lecture 9

No course on probability could possibly be complete without a discussion of two of the most famous examples of counterintuitive probabilistic scenarios. The first one we're going to do is the birthday problem, and then we're going to do the *Let's Make a Deal*® Monty Hall question.

Probability is full of surprises. The birthday problem is a famously counterintuitive result; namely, if 50 random people are in a room, there is a 97% chance that at least two of them have the same birthday. The analysis of that probability illustrates strategies of combining probabilities. This counterintuitive result can be confirmed by looking at various groups of people, such as presidents of the United States or Oscar winners, and finding birthday coincidences as predicted. The Monty Hall problem is equally baffling to most of us. It is a challenge for all of us to hone our sense of probability so that our intuition more closely accords with reality. Tricky probability problems arise in issues from birthdays to game shows to tennis to choosing socks from a drawer!

Let's start this lecture with the famous birthday problem, mandatory for any probability course. If 50 random people are in a room, what is the probability that two of them will have the same birthday? In fact, the surprising answer is that there is a 97% chance that two of them will have the same birthday. It is easier to compute the probability that all 50 birthdays are different.

To compute the probability that all the people have different birthdays, you would multiply as follows:

$$\frac{365}{366}\times\frac{364}{366}\times\frac{363}{366}\ \cdots\ \frac{319}{366}\times\frac{318}{366}\times\frac{317}{366}=0.03$$

The product of all the fractions is about 0.03. Thus, the probability that no two people have the same birthday is only about 3%. Hence, the chance that at least two people *do* have the same birthday is about 97%.

The surprisingly high probability for birthday coincidences can be tested in reality by looking at some collections of about 50 people. Of the first 42 different presidents, one pair has the same birthday: Polk and Harding: November 2 (1795, 1865). One pair and one triple of presidents have the same death day: Fillmore and Taft: March 8 (1874, 1930); J. Adams, Jefferson, Monroe: July 4 (1826, 1826, 1831). Of the first 46 vice presidents, three share a birthday: Hannibal Hamlin: August 27, 1809; Charles G. Dawes: August 27, 1865; and Lyndon B. Johnson: August 27, 1908. Of the Oscar winners for best actor, two have the same birthday: Ben Kingsley, who won in 1983: December 31, 1943; and Anthony Hopkins, who won in 1992: December 31, 1937. Two winners have the same death day: Humphrey Bogart, who won in 1952: January 14, 1957; and Peter Finch, who won posthumously in 1977: January 14, 1977. If you have 90 random people in a room, chances are .999993 that at least two will have the same birthday. And if you have only 23 people in the room, the chances are even that at least two will have the same birthday.

If you have 90 random people in a room, chances are .999993 that at least two will have the same birthday.

Another famously non-intuitive problem is the Monty Hall problem from the TV show *Let's Make a Deal*®. Here is how it works: A contestant in a game show gets to pick one of three doors and keep whatever prize is behind the door. One of the doors has a desirable prize; the two others don't. At this stage, no matter what door the contestant chooses, the probability is 1/3 that she will pick the door with the desirable prize. Having announced her choice, but before the door is opened to disclose the prize, Monty Hall, the host of the game show, opens one of the two doors she did not choose, revealing an undesirable prize, and offers her the chance to change her choice. Should she change? Yes, she should change: The probability that her original choice is the desirable prize is only 1/3, while the probability is 2/3 that the other unopened door has the good prize. The validity of the above answer assumes that the host knows which door conceals the desirable prize and never opens it.

Here is a variation on the Monty Hall problem. If there are 1,000,000 doors, the contestant's initial guess has a 1/1,000,000 chance of being right and a 999,999/1,000,000 chance of being wrong. Let's say Monty Hall opens 999,998 other doors and leaves one closed besides the one the contestant selected. The probability is 999,999/1,000,000 that the prize is behind the other remaining closed door, so the contestant should definitely switch. And here is another variation of the Monty Hall problem: Let's say there are five doors; the contestant's initial guess has a 4/5 chance of being wrong. Suppose Monty Hall then opens two of the losing doors and offers the contestant the chance to pick one of the other two remaining closed doors. The probability is 4/5 that the prize is behind one of the two closed doors other than the door originally selected. Switching to one of the other doors gives the contestant a 2/5 chance of winning, while sticking with the original choice gives her a 1/5 chance.

Our next example is a problem from tennis: If the score in a tennis game gets to deuce, what is the probability of the server winning the game? Deuce occurs when the game is tied and one player has to get ahead by two points to win. It appears that this is an infinite problem because there is no theoretical limit to the number of deuces in a game. In fact, this problem can be resolved by a clever strategy. Suppose the server has a 0.6 probability of winning each point and the receiver, a probability of 0.4 of winning. The probability of the server winning the next two points is $0.6 \times 0.6 = 0.36$. The probability of returning to deuce is $0.6 \times 0.4 + 0.4 \times 0.6 = 0.48$. Let p be the probability that the server eventually wins. Either the server could win in two points (0.36) or the game could return to deuce (0.48), followed by the server's eventually winning. We get the following equation: $p = 0.36 + 0.48p$. Solving the equation, we see that the server will win with a 0.69 probability.

Here is a final problem to ponder: Suppose you have three sock drawers. In one drawer, you have two blue socks. In a second drawer, you have two red socks. In a third drawer, you have one red and one blue sock. You randomly choose a drawer, reach in, and pick out a sock without looking. You see it is red. What is the probability that the other sock in the drawer is also red? Answer: You are equally likely to have chosen any one of the three red socks. For two of them, the other sock is red; for the third, the other sock is blue. Thus, the probability of the other sock being red is 2/3. ■

Suggested Reading

Edward B. Burger and Michael Starbird, *Coincidences, Chaos, and All That Math Jazz: Making Light of Weighty Ideas*.

———, *The Heart of Mathematics: An invitation to effective thinking*, 2^{nd} ed.

Questions to Consider

1. Suppose I am in a room with 49 other people. What is the probability that someone in the room has the same birthday as I do? Hint: This question requires a different calculation from the one presented in the lecture. To see why, suppose that my birthday is, for example, July 10.

2. Suppose in the *Let's Make a Deal*® show that Monty Hall did not know the location of the big prize, and he sometimes would open the big prize door by accident. Now analyze the situation in which the contestant selects a door, Monty Hall opens another door, and it happens to reveal a worthless prize. Is the contestant better off switching, or in this case, are the probabilities for switching and sticking the same?

Conundrums of Conditional Probability

Lecture 10

In this lecture we're going to introduce a very basic concept of probability that's associated with what happens when we're asked a probabilistic question, but then we're given more information. It changes the probability because we put ourselves in a more restricted arena of possibilities.

An important concept used to help us find our way through probabilistic complexity is the idea of *conditional probability*. Conditional probability refers to a situation in which we begin with a clear probabilistic scenario but are then told more information. The additional information alters the probabilities, but frequently, the change is challenging to analyze. Principles of dealing correctly with conditional probability can guide us to correct answers, but these are tricky and highly non-intuitive issues. The famous *Bayes' theorem* describes the relationships among related conditional probabilities. The ideas of conditional probability are introduced via some probabilistic conundrums that delightfully puzzle us.

To introduce conditional probability, we will consider a collection of 27 cards that have been chosen to illustrate the idea. There are 21 black cards, of which 9 are face cards, and 6 red cards, of which 3 are face cards. We can answer questions about the probability of choosing a certain type of card from this group of cards. What is the probability of choosing a face card? Because we have 12 face cards, the answer is 12/27. What is the probability of choosing a red card? Because we have 6 red cards, the answer is 6/27. Conditional probability enters the picture when we are told one of the characteristics that cuts down the population. For example, what is the probability of getting a red card given that we have chosen a face card? There are 3 red cards out of the total 12 face cards; thus, the conditional probability of choosing a red card given that we have chosen a face card is 3/12.

Let's look at another question that relates two different conditional probabilities. What is the probability of getting a face card that is red? This question involves two probabilities: the probability of choosing a face card and the conditional probability of choosing a red card given that we have a face card. The answer is the product of two probabilities: the probability of choosing a face card from among 27 cards 12/27 times the conditional probability of choosing a red card given that we have a face card 3/12, or (12/27) × (3/12) = 1/9. We can look at the same situation backward. What is the probability of getting a red card that is a face card? The analysis is the same: the probability of choosing a red card from among 27 cards 6/27 times the conditional probability of choosing a face card given that we have a red card 3/6, or (6/27) × (3/6) = 1/9.

Bayes' theorem is a principal tool that is used to deal with conditional probability. Suppose A represents one characteristic (such as "red card") and B represents another characteristic (such as "face card"). Bayes' theorem relates two conditional probabilities, the probability of A given B and the probability of B given A. It can be presented in two ways:

$$P[B] \times P[A|B] = P[A] \times P[B|A] \text{ or } P[A|B] = (P[B|A]P[A])/P[B]$$

Conditional probability can surprise us. Consider the following scenario: Suppose you meet a man and learn that he has exactly two children. Suppose that you learn that his older child is a boy. Therefore, we know that two of four possibilities are eliminated (two girls [GG] or an older girl and a younger boy [GB]), leaving the possibility that he has two boys (BB) or an older boy and a younger girl (BG). Of the remaining two equally likely possibilities, one is boy-boy. Thus, the probability that both children are boys given that the older child is a boy is 1/2.

This is called conditional probability. But suppose you ask the man instead, "Do you have a son?" and he answers, "Yes." The GG possibility is eliminated, and three possibilities remain, GB, BG, and BB. Thus, the probability that both of his children are boys given the knowledge (or "condition") that at least one is a boy is 1/3. Notice that the answer is not 1/2. The information that at least one child is a boy affects the probability differently than the information that the older child is a boy. Suppose you had asked the following question of the man instead: "Do you have a son who was born on a Tuesday?" and he answers, "Yes." Most people's intuition is that this birthday information is irrelevant and should yield the same probability as the previous version of the problem. To do this calculation, we begin by writing down all the possible day-of-the week and gender combinations, and we find that there are 196 in all. We then narrow down the possibilities by focusing on the pairs for which at least one child is a boy born on a Tuesday. We find we have 13 BB possibilities, 7 BG possibilities, 7 GB possibilities, and of course, no GG possibilities, for a total of 27. The probability that both children are boys given that at least one is a boy born on a Tuesday is 13/27, which is between 1/3 and 1/2.

As our knowledge and information about possibilities in a situation change, the probabilities of events change.

Let's look at another problem: Suppose we have two urns, each containing 10 balls. In one urn, we have 7 blue and 3 red balls, and in the other, we have 3 blue and 7 red balls. We can't tell which urn is which. I select an urn at random and draw a red ball from it; then I put the ball back in the urn and choose a ball again from the same urn, and it is red. I choose a third time and get a red ball and a fourth time and get a red ball. What is the probability that the urn I chose was the one with 7 red and 3 blue balls? One strategy might be to imagine having 20,000 people performing the same experiment, randomly choosing one of the two urns and randomly drawing out 1 of the 10 balls four different times. Logically, about half the people (10,000) would choose the blue-heavy urn and half, the red-heavy urn.

Out of the 10,000 people who chose the red-heavy urn, how many would we expect to choose red balls four times in a row? Each of the four times one of the people reaches into the red-heavy urn, he or she has a 70% chance of getting a red ball. Therefore, we arrive at this equation: $0.7 \times 0.7 \times 0.7 \times 0.7 = 0.2401$, or 2401 of the 10,000 people. However, for the blue-heavy urn, people have only a 30% chance of getting a red ball each of the four times a ball is chosen. The equation is: $0.3 \times 0.3 \times 0.3 \times 0.3 = 0.0081$, or 81 of the 10,000 people. We know, then, that 2482 people would draw four red balls. Therefore, the probability that the person choosing four reds is drawing from the red-heavy urn is 2401/2482, or 97%.

Let's change the scenario a little and draw a ball out of an urn 50 times instead of 4 times. Let's say we find that 27 times, we choose a red ball, and 23 times, we choose blue. What is the probability that we are choosing from the red-heavy urn? Surprisingly, our calculations show us that the probability is again 97%! ■

Suggested Reading

Edward B. Burger and Michael Starbird, *The Heart of Mathematics: An invitation to effective thinking*, 2nd ed.

Peter G. Moore, *The Business of Risk*.

Jeffrey S. Rosenthal, *Struck by Lightning: The Curious World of Probabilities*.

Questions to Consider

1. Someone tells you the following: "I met a man who told me that he has exactly two children. I asked him one question, but I can't remember what question I asked. It was either 'Is your older child a boy?' or 'Is your younger child a boy?' I remember that he answered yes." What is the probability that both of the man's children are boys?

2. Suppose you have two urns, one of which contains 10 red balls and the other, 5 red balls and 5 blue balls. You select an urn at random and draw a red ball from it; then you put the ball back in the urn and choose a ball again from the same urn, and it is red. You choose a third time and get a red ball and a fourth time and get a red ball. What is the probability that you are reaching into the red urn?

Believe It or Not—Bayesian Probability

Lecture 11

[T]his basic view of probability is called the frequentist probability, because it's talking about the frequency with which a repeated event happens. But there's another sense in which we often think in terms of probability that measures really a quite different kind of phenomenon, and so we sometimes wish to use probability to express in some sort of a quantitative way the degree to which we believe something.

What does probability mean in the real world? Probabilists do not agree. Mostly in these lectures, we've focused on the frequentist view of probability; namely, that if we repeat an experiment in question many times, the percentage of successful outcomes is the probability. However, another view of probability is that it measures a person's belief in the likelihood of the item in question. "Did Shakespeare write *Hamlet*?" We can't do a repeatable experiment pertinent to this question. A frequentist holds the view that probability applies only to experiments whose outcomes are random and, therefore, would not discuss the *Hamlet* question as one susceptible to probabilistic comment. Bayesian probability concerns itself with describing a weighted assessment of possibilities, then develops a method for revising that assessment as more evidence is amassed. The different views of probability are intriguing to consider, and in some cases, adopting one philosophy or another has practical implications.

Another use of probability is to express quantitatively our degree of belief in some statement.

In most of the examples in earlier lectures, probability could be interpreted as the fraction of successes in a series of identical experiments or trials. An example would be saying that the probability of rolling a die and coming up with a four is 1/6. That is, if you rolled the die many times, about 1/6 of those times would show four. This view of probability is called *frequentist probability*.

Another use of probability is to express quantitatively our degree of belief in some statement. For example, if I say that the probability is 98% that Shakespeare wrote *Hamlet*, you'll know that I believe very strongly that Shakespeare was the one who indeed wrote Hamlet, but that there is some small possibility that someone else was the actual author of the play. Saying that we believe there is 98% probability that Shakespeare wrote *Hamlet* makes a statement about the strength in our belief. But this does not mean that if 100 Shakespeares were born, 98 of them would have written *Hamlet*. We have two kinds of situations in which we use the same word—*probability*. As a measure of the strength of belief, *probability* expresses our uncertainty, but the two kinds of probability are different kinds of things.

When probability is used to express quantitatively a degree of belief, it must be clear what all the possibilities are. Among the various potential states of the world, we express the relative probabilities of those different states being the correct one. And we assign to each such possible state of the world a probability that that one is the correct one The sum of the probabilities is 100%.

To ground our discussion, let's take an example of fish in a stream. We're interested in what fraction of fish in the stream are trout, from the possibilities of 5%, 15%, 25%, …, 95%. Before any data are collected, we assume that we have no bias; we establish the 10 possibilities and give each the same probability.

Hypothesis: This percentage of fish in the stream are trout	**Probability of this hypothesis**
5%	0.10
15%	0.10
25%	0.10
35%	0.10
45%	0.10
55%	0.10

Hypothesis: This percentage of fish in the stream are trout	Probability of this hypothesis
65%	0.10
75%	0.10
85%	0.10
95%	0.10

Suppose we catch three fish—trout/trout/non-trout. We would naturally believe that it is more likely that the percentage of trout is high. We can update our probabilities for the various potential percentages of trout in a stream by doing a thought experiment in which we imagine 10,000,000 fishermen—1,000,000 fishing in each of 10 different universes (one for each hypothesis). We can calculate how many of those fishermen would catch a trout/trout/non-trout combination in their respective streams. The following table shows our calculations:

Hypothesis: If this percentage of fish in the stream are trout	Then of 1,000,000 fishermen, this many catch two trout and one non-trout
5%	2375
15%	19,125
25%	46,875
35%	79,625
45%	111,375
55%	136,125
65%	147,875
75%	140,625
85%	108,375
95%	45,125
Total	**837,500 in all streams**

We can now update our belief system. Our *a priori* assumption, before catching any fish, was that each hypothesis is equally likely. Then, we caught two trout, then one non-trout. We can now recalculate the probability that the stream contains 5% trout by dividing 2375 by 837,500. Likewise, we can recalculate the probability that it is a 15% stream by dividing 19,125 by 837,500, and so forth.

By applying Bayes' theorem to make an update to our previous resulting distribution, we get a new distribution that has most of the probability concentrated in the choices 35%, 45%, 55%, 65%, and 75%. If we catch another trout and another non-trout, we can perform the same type of calculations using our new, updated distribution. Now we have evidence that changes our sense of the possibility; we have, for example, many more fishermen in the 65% stream than in the 5% stream. As we catch more fish, the evidence will dominate over our initial estimate, thus reflecting the Law of Large Numbers. After catching 100 fish, we have a very strong belief that we have a 65% stream, but about a 10% chance that it is a 55% stream or a 10% chance that it is a 75% stream.

Thus, we have two views of probability. The frequentist probability is the view in which probability is defined in terms of long-run frequency or proportion in outcomes of repeated experiments. Bayesian probability is the view in which probability is interpreted as a measure of degree of belief. In this view, the concept of probability distribution is applied to a feature of a population to indicate one's belief about possible values of that feature.

Let's look at another example of updating our probability distribution in the field of medicine. A doctor narrows a patient's illness to three possibilities: A, B, or C. After assessing the patient, the doctor assigns probabilities of the patient having the diseases as follows: A: 50%, B: 40%, C: 10%. After a more thorough exam, a symptom, S, is discovered, and the doctor knows what the probability is of a patient with each of the diseases exhibiting this symptom.

Therefore, we update our initial probability distribution:

Hypothesis: Patient has this disease	Probability of this hypothesis	Number of imagined patients[1]	Probability of showing S	Number of patients showing S[2]	Updated probability of this hypothesis
A	50%	5000	10%	500	500/2500 = 20%
B	40%	4000	30%	1200	1200/2500 = 48%
C	10%	1000	80%	800	800/2500 = 32%

[1]Out of 10,000
[2]Note that the total is 2500.

We see that B is now the most probable disease, replacing disease A. ■

Suggested Reading

Donald A. Berry, *Statistics: A Bayesian Perspective*.

E. T. Jaynes, *Probability Theory: The Logic of Science*.

Questions to Consider

1. Bayesian probability involves having an a priori distribution and updating it in light of evidence. What is the influence of different a priori beliefs after a great deal of evidence is accumulated? Why?

2. Suppose your a priori belief about a coin is that you are 100% certain that it will always land heads. You flip the coin and it lands tails. Then you cannot update your probability distribution because you ascribed 0 to the probability of ever getting a tail. What went wrong?

Probability Everywhere
Lecture 12

In this lecture we're going to follow a road that often leads to interesting ideas, and that is the road of trying to understand what appears to be a paradoxical kind of situation; and then in thinking it through, we develop an idea.

One of the strengths of mathematics is its strategy of generalizing and abstracting ideas. In the case of probability, we have mostly considered situations for which a finite number of possible outcomes was possible for a given situation; then, we investigated issues of probability associated with that situation. The techniques we developed can be extended to situations in which infinitely many outcomes are imagined as possible. The two envelopes problem and the St. Petersburg paradox each force us to confront new challenges that arise when infinitely many outcomes are possible.

Probability is a fascinating study that has many real-world applications. It presents us with a rich field of intriguing inquiry that contains questions and insights that are mathematical, practical, and philosophical. Often, mathematical ideas are born by trying to tackle a specific problem. In thinking through how to deal with the specific problem, new ideas are created.

Here is a conundrum known as the two envelopes problem: You are given two envelopes and told that each envelope contains a check for a certain amount of money, and one of the checks is for exactly twice as much money as the other. You randomly select one of the envelopes and open it. The enclosed check is for a certain amount of money, say d dollars. Now you can either keep that money, or you can take the contents of the other envelope. You know that you are as likely to have chosen the lesser amount as you are likely to have chosen the greater amount. But now you do an expected-value analysis and find a paradoxical situation.

There is a 0.5 probability that you have the higher amount and a 0.5 probability that you have the lower amount; thus, the expected value of switching is:

$$\frac{1}{2}(2d)+\frac{1}{2}\frac{d}{2}=d+\frac{d}{4}=\frac{5}{4}d\ .$$

The result is greater than d, so this analysis seems to suggest that you should switch. But this makes no sense, because it is clearly as likely that you have the lower amount as the greater amount. What is wrong? The two envelopes problem brings up a situation we have not dealt with much, namely, one in which the experiment has infinitely many possible outcomes. How can we revise our thinking to cope with infinitely many alternatives?

If any amount of money is possible, then there are infinitely many possibilities theoretically. But in reality, huge numbers are not possible. Actually, we have an a priori sense—an expectation—of a probability distribution. We can hearken back to the Bayesian strategy and realize that we have an a priori sense of the probabilities of various amounts. Depending on our a priori beliefs, we are forced to confront reality and realize that we don't have an infinite number of possible amounts of money in the envelopes. We can describe the probabilities by a graph based on the expected-value analysis using the probabilities according to our a priori distribution. The expected-value analysis using the probability distribution that takes into account the infinite number of possible outcomes will give us good guidance about whether to switch envelopes. In addition, we must point out that even when dealing with an infinite number of possible outcomes, we must assign probabilities that total 1.

Another famous paradox involved with gambling is the St. Petersburg paradox. Suppose a gambler plays a coin-flipping game, winning \$2 for flipping heads. If the gambler flips tails, then heads, he wins \$4. If the gambler flips two tails in a row, then heads, he wins \$8. If the gambler is very lucky, he might flip five tails in a row, followed by heads, to win \$64. How much would you pay to play this game?

We can calculate the expected value:

$$\frac{1}{2}2+\frac{1}{4}4+\frac{1}{8}8+\frac{1}{16}16+\ldots= \ 1+1+1+\ldots=\infty$$

The expected payout is infinity, so it appears you should pay any amount of money to play this game. The paradox is that you would not make a great deal of money in this game. Simulations show that average payoffs are quite low.

Probability predictions are the basis of statistical inference. Statistical inferences boil down to comparing expectations from probability with collected data. If your expectation from probabilistic analysis differs greatly from what you see in the data, you can make the deduction that the concept you had about how the data were being produced must be wrong. Statistics is an important application of probability and is covered in The Teaching Company course *Meaning from Data: Statistics Made Clear*.

Probability is a fascinating field that plays a fundamental role in how we understand our world, from games to science to finance. One recurring theme of the course was that randomness and probability often confront us with situations that are counterintuitive. Probability offers intriguing and sometimes subtle puzzles, such as the birthday problem, the Monty Hall *Let's Make a Deal*® puzzle, and the two-boys puzzle. All these examples seem wrong, but when our intuition and reality are not in accord, one of them has to give, and it has to be our intuition. After we have adjusted our understanding to see the truth of these counterintuitive examples, then the probability results are ones that we can make reliable decisions on.

Probability gives us a logically sound way of quantifying uncertainty. Many of the ideas about probability in this course were illustrated in the realm of gambling, because gambling games are fundamentally based on probability. Casinos count on probability to ensure their success. Casinos are the modern world's testament to the Law of Large Numbers. Random behavior that results in regularity in the aggregate is a central feature of our serious, scientific understanding and descriptions of nature. The directions and speeds of molecules are far too numerous to count and describe. Instead, we describe the interactions as the result of a probabilistic description of random motion, with appropriate constraints that describe the molecular behavior.

The role of randomness is central to the science of genetics because the whole premise of the subject is that parts of the genetic material from each parent are randomly donated to the offspring. Of course, probability plays a central role in descriptions of our financial world and investments. Investments are viewed as having a probability of rising or falling. Devising an optimal portfolio involves optimizing the probability of success.

Often, mathematical ideas are born by trying to tackle a specific problem. In thinking through how to deal with the specific problem, new ideas are created.

One of the fundamental sources of our uncertainty about the world is that often, we don't know what is really true among several possibilities. When we sit on a jury, we may not know whether the accused is innocent or guilty. Instead, we have a sense that there is some likelihood of guilt and some likelihood of innocence. As evidence is adduced at the trial, our relative confidence in guilt or innocence shifts. The strategy of Bayesian probability describes the relative strengths of our beliefs and how they are altered by evidence.

Randomness and uncertainty are fundamental parts of reality. Probability describes what we should expect from randomness. Probability is a basic tool for making sense of and coping with the reality of randomness and uncertainty in our world. ■

Suggested Reading

Ivars Peterson, *The Jungles of Randomness: A Mathematical Safari*.

Sheldon Ross, *A First Course in Probability*.

Questions to Consider

1. Suppose in the St. Petersburg game, the rule was changed so that you received $64 as soon as you flipped five tails in a row and the game then ended. How much should you pay to make it a fair game? Would you play such a game?

2. Some people believe that everything that happens in life happens for a reason. To what extent do you believe that the occurrences of everyday life are random?

Timeline

1563... Girolamo Cardan writes (but doesn't publish) *Liber de Ludo Aleae*, a book on games of chance. He was the first to venture into studying probability.

1654... Blaise Pascal and Pierre de Fermat, through a series of five letters, discuss probabilistic solutions to a number of mathematical questions raised in the analysis of dice games.

1655... Christiaan Huygens publishes *De Ratiociniis in Ludo Aleae*, on the calculus of probabilities, the first printed work on the subject.

1689... Jacob Bernoulli publishes the concept of the Law of Large Numbers, a mathematical statement of the fact that when an experiment is repeated a large number of times, the relative frequency with which an event occurs will equal the probability of the event.

1713... Nicholas Bernoulli edits and publishes *Ars Conjectandi* (*The Art of Conjecture*), written by his uncle, Jacob Bernoulli, in which the work of others in the field of probability is reviewed and thoughts on what probability really is are presented.

1728 .... Sir Isaac Newton publishes *The Chronology of Ancient Kingdoms Amended*, in which he gives a 65% confidence interval for the length of a king's reign.

1733 .... Abraham de Moivre publishes an account of the normal approximation for the binomial distribution for a large number of trials. This improves upon Jacob Bernoulli's Law of Large Numbers. This account will be included in the 1756 edition of De Moivre's *The Doctrine of Chances*, a treatise on probability first published in 1718.

1738 .... Daniel Bernoulli publishes *Exposition of a New Theory on the Measurement of Risk*, an early look at probability theory and economic decision making.

1820 .... Pierre-Simon Marquis de Laplace publishes a seminal work on probability.

1827 .... Robert Brown, a botanist, while observing the motion of pollen grains, hypothesizes underlying mechanics for erratic movements. This later led Bachelier and Einstein to study and make rigorous Brown's work. The mechanics are now known as Brownian motion in his honor.

1837 .... Simeon Denis Poisson publishes *Recherches sur la probabilité des jugements en matière criminelle et matière civile*, which introduces the expression *Law of Large Numbers* and in which the Poisson distribution first appears.

1853.. Augustin-Louis Cauchy presents an outline of the first rigorous proof of the central limit theorem, which is a generalization of the Law of Large Numbers.

1867.. Pafnutii Lvovich Chebyshev publishes a paper, *On Mean Values*, which uses Irenée-Jules Bienaymé's inequality to give a generalized Law of Large Numbers.

1887.. Pafnutii Lvovich Chebyshev publishes *On Two Theorems*, which gives the basis for applying the theory of probability to statistical data, generalizing the central limit theorem of de Moivre and Laplace.

1900.. Louis Bachelier publishes the first mathematical approach to Brownian motion in his Ph.D. thesis, *Théorie de la Spéculation*.

1905.. Einstein publishes three groundbreaking scientific papers. The third and least famous of the three (the first won the Nobel Prize for Physics and the second was on special relativity) detailed a mathematical treatment of Brownian motion.

1919.. Paul Levy delivers three lectures at the École Polytechnique, highlighting entirely new areas of research in probability theory.

1938.. Kolmogorov publishes the influential *Analytic Methods in Probability Theory.*

1942	Kiyosi Ito publishes *On Stochastic Processes (Infinitely Divisible Laws of Probability)*, a groundbreaking paper.
1944	Von Neumann and Morgenstern publish *Theory of Games and Economic Behavior*, the first text on the new field of game theory.
1950	William Feller writes the first volume of his famous *Introduction to Probability Theory and Applications*.
1953	Joseph Leo Doob publishes *Stochastic Processes*, a now classic text on stochastic (probabilistic) analysis and martingale theory.
1966	Norbert Wiener publishes *Nonlinear Problems in Random Theory*.
1966	MIT mathematician Ed Thorp publishes *Beat the Dealer*, a popular work on applying probabilistic thinking in the game of blackjack in Las Vegas casinos.
1969	Fischer Black and Myron Scholes write their seminal paper on a mathematical and probabilistic approach to pricing options.
1973	Robert C. Merton publishes *Theory of Rational Option Pricing*.
1994–1998	Long-Term Capital Management experiences its strong profitable run, then collapses.
1997	Robert Merton and Myron Scholes, applied mathematicians, win the Nobel Prize for Economics for their work in options-pricing theory.

Glossary

Bayes' theorem: A mathematical equation relating two conditional probabilities: $P[A|B] = (P[B|A]P[A])/P[B]$.

Bayesian probability: The view in which probability is interpreted as a measure of degree of belief. In this view, the concept of probability distribution is applied to a feature of a population to indicate one's belief about possible values of that feature. The principal result of experiments or more evidence is to update such a probability distribution, indicating a change in belief. The Bayesian viewpoint is in contrast to the frequentist view.

Bell's theorem: A theorem asserting that a particular inequality of certain probabilities would be true if intuitive concepts of local realism were true of particle physics. The theory of quantum physics violates the inequality. Quantum theory implies that when one particle of an entangled pair of particles is observed, the other particle in the pair, which could be distant, instantaneously undergoes a state change. Bell's theorem implies that this aspect of quantum theory cannot be explained by hidden local variables.

chance: An informal term that tries to capture the same notion as the term *probability*.

complementary event: The event complementary to a given event is the set of all possible outcomes that are not in (or do not satisfy or do not represent) the given event. For example, in rolling two dice, one event is: "The sum of the dice is 8." Its complementary event is: "The sum of the dice is not 8."

conditional probability: The probability of an event under the assumption of the existence (or happening or satisfaction) of another event. For example, in rolling a blue fair die and a red fair die, the conditional probability of the event "the sum of the dice is 8," given the event "the blue die is 3 or 6," is 2/12 = 1/6, because there are 12 possible outcomes with the blue die being

3 or 6, and two of those (blue 3 and red 5; blue 6 and red 2) sum to 8. We would say, "The probability of the sum being 8 given that the blue die is 3 or 6 is 1/6."

deterministic model: A mathematical description of a phenomenon or mechanism that does not depend on randomness. Every time the model is executed with the same initial conditions, the result (prediction) will be the same. Contrast with **probabilistic model**.

disjoint events: Two (or more) events that cannot both happen (for one experiment). Each possible outcome of the experiment is in (or satisfies or represents), at most, one of the events. For example, in rolling two dice, the event "the sum is 8" is disjoint from the event "there is a 1."

event: A set of possible outcomes of an experiment, trial, or observation. For example, for the trial of rolling a blue die and a red die, a possible event is: "The sum of the dice is 8." This event consists of the following five outcomes: blue 2 and red 6, blue 3 and red 5, blue 4 and red 4, blue 5 and red 3, blue 6 and red 2. Compare to **outcome**.

expected value: Assuming a numerical value is associated with each possible outcome of an experiment (or a trial or an observation), the expected value of the experiment is the weighted average of the values, where each weight is the probability of the associated outcome. The expected value is a number that summarizes the possible values. The term can be misleading, because often the expected value as a number is not associated with any possible outcome. For example, in the experiment of flipping a fair coin, if the value 2 is associated with heads and the value 5 with tails, then the expected value is 3.5 (which is neither 2 nor 5 and, hence, hardly to be "expected"). More formally, it is the expected value of a random variable that is defined, rather than the expected value of an experiment.

fair: When used in such phrases as "a fair coin" or "a fair die," this term indicates the ideal situation in which the probability of any of the possible outcomes is the same.

flush: In poker, a hand of five cards in which all the cards are of the same suit but cannot be placed in sequential order. See **straight** for examples of cards in sequential order. Compare to **straight flush**.

frequentist probability: The view in which probability is defined in terms of long-run frequency or proportion in outcomes of repeated experiments. This concept of probability is applied to outcomes of actual or hypothetical experiments that have an element of randomness. But in the frequentist view, probability is not used as a measure of knowledge or belief of the possible values of a quantity that does not have a random element. The frequentist viewpoint is in contrast to the Bayesian view.

independent events: Two events are independent if one event's occurring does not affect the probability that the other occurs. If A and B are independent events, then $P(AB) = P(A)P(B)$; that is, the probability that both A and B occur is the product of the probabilities that each occurs. For example, in flipping two coins, assuming that the results of one flip don't affect the results of the other, then the probability of both coins landing on heads is the product of the probability that the first coin lands on heads times the probability that the second coin lands on heads.

Law of Large Numbers: The theorem that the ratio of successes to trials in a random process will converge to the probability of success as increasingly many trials are undertaken.

mutation: A change in a gene of an organism. Some mutations are inherited by offspring of the organism that suffered the mutation. Mutations are often modeled as occurring randomly. Probabilistic models make assumptions on the rate of mutations that are passed to offspring. From these models, conclusions are drawn about the evolutionary history of species.

odds: An alternative way of expressing the probability of an event by stating the ratio: the probability that the event happens divided by the probability that the event does not happen. For example, if the probability of an event is 20%, the odds are 20/80, or 1/4. This is sometimes stated, "four to one against."

option: In the financial markets, a contract giving the holder the right to buy a prescribed asset (such as a certain number of shares of a specific stock) at a prescribed time in the future for a prescribed amount of money, or a contract giving the holder the right to sell a prescribed asset at a prescribed time in the future for a prescribed amount of money, or other related contracts.

outcome: A possible specific result of an experiment, trial, or observation. For example, for the trial of rolling a blue die and a red die, one possible outcome is blue 3 and red 5. Compare to **event**.

permutation: An ordering of distinct objects. For example, there are 24 permutations of the four cards ace of spades, king of diamonds, queen of diamonds, and eight of hearts because there are 24 different ways to order those four cards.

poker: A card game (with several variations) played with an ordinary deck of 52 cards, in which five-card sets are compared to see which is "better." The ordering is based on the probabilities of various possible features of a five-card set; rarer features win.

prime number: A whole number (an integer) bigger than 1 that is not evenly divisible by any positive whole number except itself and 1.

probabilistic model: A mathematical description, with random aspects, of a phenomenon or mechanism. The model could consist of mathematical formulas that refer to random numbers. Thus, one execution of the model will generally give different results than another execution. Contrast with **deterministic model**.

probability: As the term is used in mathematics, a number between 0 and 1 (or 0% and 100%) applied to a possible future event that quantifies the likelihood of the event's occurring, or that number applied to a statement that quantifies our degree of belief in the truth of the statement.

probability distribution: A discrete probability distribution is a table, function, or graph that assigns a probability to each possible outcome. For the continuous case, in which any real value is a possible outcome, the probability distribution can be viewed as a graphed curve that has an area of 1 under the curve and above the horizontal x axis. The probability of an outcome being between value a and b is equal to the area under the part of the curve between $x = a$ and $x = b$.

random variable: The assignment of a number to each possible outcome of an experiment. The term *random variable* is an unusually poorly chosen term, because it denotes something that is neither random nor a variable. We avoided using this term in this course.

random walk: A sequence of positions of an object that takes one step each second (or other unit of time), in which the direction of each step is random. The direction of each step is randomly chosen independent of any other step. An example of a one-dimensional random walk is formed by flipping a coin to determine whether the next step should be forward or backward.

randomness: The aspect of life, or a system, or a pattern, or a mathematical model that is unpredictable even in theory or unpredictable because of lack of detailed knowledge. Randomness in a system implies that the behavior of the system can be different even if the system is subjected to identical circumstances. Although random occurrences are not predictable, they exhibit regularity in the aggregate after many repetitions.

roulette: A gambling game in which a small ball settles into one of 38 slots in a wheel as the wheel is spun and slows. The slots are numbered 0, 00, 1, 2, … , 36. Presumably, each slot is equally likely on any given spin of the wheel to be the stopping point for the ball. Note: European roulette wheels have only 37 slots (no 00).

stochastic model: Synonym for **probabilistic model**.

straight: In poker, a hand of five cards that can be put in sequential order, with not all five cards being of the same suit. Examples include ace, 2, 3, 4, 5; 9, 10, jack, queen, king; and 10, jack, queen, king, ace; but not jack, queen, king, ace, 2. Compare to **straight flush**.

straight flush: In poker, a hand of five cards that can be put in sequential order and in which all five cards are of the same suit. See **straight** for examples of sequential order.

uniform distribution: A probability distribution in which every possible value is equally likely.

weighted average: Given a set of numbers $\{a, b, c, d, ...\}$ (thought of as values of some quantity) and a weight for each number $(w_a, w_b, w_c, w_d, ...)$, the weighted average is the value $aw_a + bw_b + cw_c + dw_d + ...$. The weights must add up to 1 and must be non-negative.

Biographical Notes

Bayes, Thomas (1701–1761). British nonconformist minister. Little is known about Bayes's life, save that he was educated at Edinburgh University and was a member of the Royal Society. His major contribution to the field of probability was the work he did on the inverse probability problem. At the time, the calculation of the probability of a number of successes out of a given number of trials of a binomial event was well known. Bayes worked on the problem of estimating the probability of the individual outcome from a sample of outcomes and discovered the theorem for such a calculation that now bears his name.

Bernoulli, Jacques (often called Jacob or James, 1654–1705). Professor of mathematics at Basel and a student of Leibniz. He formulated the Law of Large Numbers in probability theory and wrote an influential treatise on the subject.

Black, Fischer (1938–1995). Applied mathematician and economist. Worked both in academia and on Wall Street. Pioneer in the field of options pricing and among the first to bring higher mathematics to the financial sector. Held long-standing beliefs about the inherent uncertainties in the markets. Most famous for coauthoring the Black-Scholes formula, for which his coauthor, Myron Scholes, received the Nobel Prize in 1997.

Cardano, Girolamo (1501–1576). Italian mathematician. An avid gambler, he was the first to explore the mathematics of probability in order to improve his game play. He also recorded the first calculations with imaginary numbers. Cardano was the first to understand that there are fundamental scientific and mathematical principles guiding events previously only describable by chance.

Cauchy, Augustin-Louis (1789–1857). French mathematician and engineer. Professor in the École Polytechnique and professor of mathematical physics at Turin. He worked in number theory, algebra, astronomy, mechanics, optics, and analysis. His contribution to probability and statistics was the production of the outline of the first rigorous proof of the central limit theorem, which is a generalization of the Law of Large Numbers.

Chebyshev, Pafnutii Lvovich (1821–1894). Russian mathematician, founder of the St. Petersburg School of Mathematics. He made fundamental contributions to the theory of probability and statistics, including generalizations of the central limit theorem, which is itself a generalization of the Law of Large Numbers.

de Moivre, Abraham (1667–1754). French-English mathematician. Born in France and educated at the Sorbonne in mathematics and physics, de Moivre, a Protestant, emigrated to London in 1688 to avoid further religious persecution. A future fellow of the Royal Society of London, de Moivre supported himself in England as a traveling mathematics teacher and by selling advice in coffee houses to gamblers, underwriters, and annuity brokers. De Moivre is recognized in statistics as the first to publish an account of the normal approximation to the binomial distribution. In fact, some of de Moivre's methods are so ingenious as to be shorter than modern demonstrations of solutions to the same problems.

Doob, Joseph Leo (1910–2004). American mathematician. Produced substantial work on probability theory, stochastic processes, potential theory, and much more. Also authored several seminal texts on probability theory.

Einstein, Albert (1879–1955). Probably the most famous scientist of all time. In addition to his well-known work in several areas of physics, in 1905, he presented one of the first mathematical treatments of Brownian motion. It was Einstein's interest in statistical mechanics that led him to explore Brownian motion.

Fermat, Pierre de (1601–1665). French lawyer and mathematician. Through an interest in games of chance, Fermat used his mathematical prowess to study the mathematics of chance. Following a brief correspondence with Pascal, the two came to be considered joint founders of mathematical probability.

Huygens, Christiaan (1629–1695). Dutch astronomer and mathematician. While most famous for his discoveries about the planet Saturn and his invention of the pendulum clock, Huygens was also an early pioneer of the mathematics of probability. Following a meeting with Fermat, he presented the first printed work on probability theory.

Ito, Kiyosi (b. 1915). Japanese mathematician and statistician. His contribution to probability theory was to develop the notion of stochastic (probabilistic) differential equations.

Kolmogorov, Andrei Nikolaevich (1903–1987). Russian mathematician who ranks among the greatest of the 20th century. A formalist who helped axiomatize probability.

Laplace, Pierre-Simon Marquis de (1749–1827). French mathematician and astronomer. Professor at the École Normale and École Polytechnique, known for his contributions to calculus, analysis, probability theory, and physics. One of the earliest mathematicians to formalize the theory of probability.

Levy, Paul Pierre (1886–1971). French mathematician. A pioneer in modern probability theory. Not a formalist like his contemporary, Kolmogorov; an important class of stochastic processes bears his name.

Markov, Andre Andreevich (1856–1922). Russian mathematician. Member of the St. Petersburg Academy of Science. Markov worked on the Law of Large Numbers and random walks.

Merton, Robert Carhart (b. 1944). Applied mathematician. Student of Nobel laureate Paul Samuelson. Credited with being among the first to bring stochastic calculus and other sophisticated probabilistic tools to finance. Helped develop the Black-Scholes pricing formula (also called Merton-Black-Scholes). He developed probabilistic and analytic theorems that paved the way for the now-high-profile field of financial engineering. Recipient of the 1997 Nobel Memorial Prize in Economics.

Neumann, John von (1903–1957). Hungarian mathematician and one of the original members of the Institute of Advanced Study at Princeton University (along with Albert Einstein). A genius who contributed to many areas of mathematics and physics, he is most popularly known as the inventor of game theory. He authored a celebrated text, *Theory of Games and Economic Behavior*.

Newton, Sir Isaac (1642–1727). English mathematician and scientist known for the discovery of the law of gravity and as one of the fathers of calculus. Within the field of probability, he is known for his proof of the binomial theorem. There is also evidence that he gave thought to the variability of the sample mean, the basis for the central limit theorem. In his last work, *The Chronology of Ancient Kingdoms Amended*, published posthumously in 1728, Newton estimated the mean length of a king's reign to be between 18 and 20 years.

Pascal, Blaise (1623–1662). French mathematician and philosopher. In the summer of 1654, he exchanged a series of five letters with Fermat, in which they explored a dice game. The first question they considered was how many times one must throw a pair of dice before one expects a double six, as well as how to divide the stakes if a game is incomplete. Because of this correspondence, they are usually considered the cofounders of probability.

Poisson, Simeon Denis (1781–1840). French mathematician. He published *Recherches sur la probabilité des jugements en matière criminelle et matière civile* in 1837, marking the first appearance of the Poisson distribution, originally found by de Moivre, which describes the probability that a random event will occur in a time or space interval under the conditions that the probability of the event's occurring is very small. Poisson also introduced the expression *Law of Large Numbers*, by which he meant that, for a larger number of trials, the proportion of successful outcomes exhibits statistical regularity. Although we now rate this work as of great importance, it found little favor at the time, the exception being in Russia, where Chebyshev developed his ideas.

Scholes, Myron (b. 1941). Applied mathematician and economist. Coauthor of the Black-Scholes options-pricing formula. Recipient of the 1997 Nobel Prize in Economics. Scholes laid down fundamental mathematical assumptions that still dominate derivatives pricing in the financial markets today. He was a partner at the famously ill-fated hedge fund Long-Term Capital Management.

Wiener, Norbert (1894–1964). Applied mathematician. He mathematically extended the work done by Einstein on Brownian motion (hence, the results are often called Wiener processes). In addition, he generalized and abstracted several fundamental notions and definitions in probability theory, laying the foundation for Ito's work on stochastic analysis.

Bibliography

Albert, Jim, and Jay Bennett. *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*. New York: Copernicus Books, 2003. Introduces the fundamental concepts of statistics through applications to historical baseball. Tackles in detail such questions as the best hitter and hitting streaks.

Berry, Donald A. *Statistics: A Bayesian Perspective*. Belmont, CA: Duxbury Press at Wadsworth Publishing Co., 1996. An excellent elementary introduction to statistics that includes some probability, with many interesting real examples, most from medicine. The Bayesian approach is used.

Burger, Edward B., and Michael Starbird. *Coincidences, Chaos, and All That Math Jazz: Making Light of Weighty Ideas.* New York: W.W. Norton, 2005. A lighthearted but authentic presentation of some intriguing ideas of mathematics, including coincidences and other probabilistic phenomena.

———. *The Heart of Mathematics: An invitation to effective thinking*, 2nd ed. Emeryville, CA: Key College Publishing, 2005. This award-winning book presents deep and fascinating mathematical ideas in a lively, accessible, and readable way. The review in the June-July 2001 issue of the *American Mathematical Monthly* said, "This is very possibly the best 'mathematics for the non-mathematician' book that I have seen—and that includes popular (non-textbook) books that one would find in a general bookstore."

Charlesworth, Brian, and Deborah Charlesworth. *Evolution: A Very Short Introduction*. New York: Oxford University Press, 2003. This is one of the books in the "Very Short Introduction" series of the Oxford University Press. It is a brief but thorough account of the current status of evolution. It includes discussion of mutations, natural selection, and genetic drift.

Derman, Emanuel. *My Life as a Quant*. Hoboken, NJ: John Wiley and Sons, 2004. An autobiography of a successful probabilist working on Wall Street.

Feller, William. *An Introduction to Probability Theory and Its Applications*. Hoboken, NJ: John Wiley and Sons, 1968. Technical but classic text on probability.

Gordon, Hugh. *Discrete Probability*. New York: Springer, 1997. A mathematically rigorous book that starts from the beginning. Treats such topics as how to count combinatorially and has many exercises. Sticks to mathematical examples rather than applications.

Hacking, Ian. *The Emergence of Probability*. Cambridge, UK: Cambridge University Press, 1975. A history of probability, emphasizing the philosophical transitions and early applications that represent breakthroughs in how our world is perceived.

———. *The Taming of Chance*. New York: Cambridge University Press, 1990. Explores the changes in how chance, probability, and statistics were perceived and employed during the 18th, 19th, and 20th centuries.

Haigh, John. *Taking Chances; Winning with Probability*. New York: Oxford University Press, 2003. Illustrates ideas and results of probability using lotteries and games of all sorts, including casino games, dice, football pools, and sports. Includes discussion of the non-intuitive behavior of ties in coin flipping and random walks.

Heyde, C. C., and E. Seneta, eds. *Statisticians of the Centuries*. New York: Springer-Verlag New York, 2001. This book contains short biographies of statisticians from the 16th to the 20th centuries, many of whom made important contributions to probability and its uses.

Huff, Darrell. *How to Lie with Statistics*, New York: W.W. Norton, 1954. This charming little book has been in continuous publication since 1954. Although it is more about statistics than probability, it is eminently readable and cheerfully describes methods to mislead with statistics.

Jaynes, E. T. *Probability Theory: The Logic of Science.* Edited by G. Larry Bretthorst. Cambridge, UK: Cambridge University Press, 2003. This book makes the case for the Bayesian approach to probability and statistics, pointing out the difficulties in the orthodox approach. The book is technical.

Johnson, Norman L., and Samuel Kotz, eds. *Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present.* New York: John Wiley and Sons, 1997. Presents biographies of more than 100 statisticians and probabilists of the last four centuries.

Laplace, Pierre-Simon Marquis de. *A Philosophical Essay on Probabilities*. Mineola, NY: Dover Publications, 1951 (actually written about 1820). An early and famous work on inductive reasoning, using probability in what we now would call the Bayesian style. The terminology and wording is very old-fashioned and non-standard by today's conventions. Laplace wrestles with fundamental questions about making conclusions based on partial evidence. The most famous section is the computation of the probability that the Sun will rise tomorrow.

Levinson, Horace C. *Chance, Luck and Statistics*. Mineola, NY: Dover Publications, 1963. A down-to-earth, readable, elementary treatment of basic ideas and computations of probability, using mostly concrete settings, often gambling games such as poker as examples.

Lewis, Michael. *Moneyball: The Art of Winning an Unfair Game*, New York: W.W. Norton, 2003. A delightful story of applying probability and statistics in baseball management.

Lowenstein, Roger. *When Genius Failed*, New York: Random House, 2000. The story of the Nobel laureates' foray onto Wall Street.

Moore, Peter G. *The Business of Risk*. New York: Cambridge University Press 1983. Applications of probability to actual business situations. Gives a user's point of view of probability rather than a mathematician's. Has the flavor of a fairly sophisticated economics book, with some formulas and graphs that are essential to the content.

Morgenstern, Oskar, and John von Neumann. *Theory of Games and Economic Behavior*, commemorative edition. Princeton, NJ: Princeton University Press, 2004. The famous book from the inventors of the subject.

Mosteller, Frederick. *Fifty Challenging Problems in Probability with Solutions*. Mineola, NY: Dover Publications, 1965. A nice collection of probability problems, complete with carefully worded solutions.

Paulos, John A. *A Mathematician Plays the Stock Market*. New York: Basic Books, 2003. Engaging, witty stories about what mathematical thinking can disclose about the stock market.

Peterson, Ivars. *The Jungles of Randomness: A Mathematical Safari*. New York: John Wiley and Sons, 1998. Ivars Peterson is a popular author of general mathematics books for the public. This book discusses the subtleties of the concept of randomness.

Rapoport, Anatol. *Two-Person Game Theory*. Mineola, NY: Dover Publications, 1966. A readable classic on game theory. Includes treatment of cases in which the best strategy for a player is a probabilistic one.

Rosenthal, Jeffrey S. *Struck by Lightning: The Curious World of Probabilities*. Washington, DC: Joseph Henry Press, 2006. This book explains the ideas of probability and its appearance in everyday life.

Ross, Sheldon. *A First Course in Probability*. Upper Saddle River, NJ: Prentice Hall, 2001. Good introductory text to college-level probability theory.

Salsburg, David. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. New York: Henry Holt & Co., 2001. Presents the history of statistics during the 20th century with humor and insight. Historical anecdotes bring the topics of statistics to life. This book is very readable and enjoyable, though mostly about statistics.

Stewart, Ian. *Does God Play Dice? The New Mathematics of Chaos*. Malden, MA: Blackwell Publishing, 2002. Explores the topics of mathematical chaos and randomness in the world.

Taleb, Nassim N. *Fooled by Randomness.* New York: Random House, 2004. A wonderful book on the hidden nature of chance and luck in our lives.

Thorp, Ed. *Beat the Dealer: A Winning Strategy for the Game of Twenty-One*. New York: Vintage, 1966. The original MIT-versus-Las Vegas book. Genuine mathematical strategies that casinos would never actually let you use.

Yaglom, A. M., and I. M Yaglom. *Challenging Mathematical Problems with Elementary Solutions*, Volume 1. Mineola, NY: Dover Publications, 1964. Contains challenging and fairly sophisticated probability and other mathematical puzzles. The solutions often involve clever ways of looking at problems.

Internet Resources:

Chance Magazine. Accessible articles about statistics and probability and their applications. www.amstat.org/publications/chance/index.html.

Index of Biographies. School of Mathematics and Statistics, University of St. Andrews, Scotland. This website gives biographical information about thousands of noted mathematicians. Both chronological and alphabetical indexes are presented, as well as such categories as female mathematicians, famous curves, history topics, and so forth. www.groups.dcs.st-andrews.ac.uk/~history/BiogIndex.html.

Probability. Cut-the-Knot. This site includes a list of delightful probability puzzles. www.cut-the-knot.org/probability.shtml.

The R Project for Statistical Computing. Department of Statistics and Mathematics, Vienna University of Economics and Business Administration. R is a language and environment for statistical computing. This website contains a downloadable statistical and probabilistic software tool for computing and graphing statistical probabilistic calculations. www.R-project.org.

Wizard of Odds. Excellent site run by a mathematician who works in Las Vegas. He includes probabilistic analysis of hundreds of different gambling games and scenarios. www.wizardofodds.com.