

TEXTBOOK

UNIT 7

UNIT 07

MAKING SENSE OF RANDOMNESS

TEXTBOOK

UNIT OBJECTIVES

- Mathematical consideration and understanding of chance took a curiously long time to arise.
- The outcome of any particular event can be unpredictable, but the distribution of outcomes of many independent events that are trials of a single experiment can be predicted with great accuracy.
- The simple probability of a particular ("favorable") outcome is the ratio of outcomes that are favorable to the total number of outcomes.
- The Galton board is a useful model for understanding many concepts in probability.
- The Law of Large Numbers says that theoretical and experimental probabilities agree with increasing precision as one examines the results of repeated independent events.
- The distribution that results from repeated binary events is known as a binomial distribution.
- A standard deviation is a measurement of the spread of the data points around the average (mean).
- A normal distribution is determined solely by its mean and standard deviation.
- The Central Limit Theorem says that the average of repeated independent events is, in the long run, normally distributed.
- Conditional probability and Markov chains provide a way to deal with events that are not independent of one another.
- The BML Traffic model represents the frontier of modern probabilistic understanding.

“ ”

The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of unreason. Whenever a large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

SIR FRANCIS GALTON

SECTION 7.1

INTRODUCTION

Mathematics is often thought of as an exact discipline. In fact, many people who practice math are drawn to it because it tackles situations in which there are clear and predictable answers. There is a certain comfort in the idea that we can use mathematics to make exact predictions about what happens in the future. For example, we could use the mathematical formulation of physical laws to predict the outcome of a coin flip if we knew enough about its size, weight, shape, initial velocity, initial angle, and its other initial conditions. In practice, however, we have a very hard time knowing all of the conditions that contribute to the outcome of a coin flip. In the face of such complexity, we call the flip a “random” event, one in which the outcome is based solely on chance and not on any immediate knowable cause. Nonetheless, mathematics has a broad set of tools to explain and describe events that appear, like the coin toss, to be random. This set of tools makes up the mathematics of probability.

Does the past determine the future? If an event is truly random, the answer must be “no.” There would be no way to predict the outcome of a specific event given knowledge about its previous outcomes. Although it might seem that situations like this are beyond the reach of mathematics, the truth is that random events behave quite predictably, as long as one has no interest in the outcome of any single event. Taken on average, random events are highly predictable.

Probability theory manifests itself in many ways in our daily lives. Most of us have insurance of some form or another—house, car, life, etc. These are products that we purchase to help mitigate risk in our lives. We often associate risk with unpredictable outcomes. This could be in the context of a small business opening in an up-and-coming neighborhood, a commodities trader making decisions based on how global political situations affect prices, or a teenager getting behind the wheel for the first time. All of these situations involve a certain amount of complexity that is functionally unpredictable on a case-by-case basis. Probability theory, however, shows that there is paradoxically a large amount of structure and predictability when these individual situations are examined on a larger scale.

Probability theory shows that we can indeed make useful analyses and predictions of events that are unpredictable on a case-by-case basis, provided we look at the bigger picture of what happens when these events are repeated many times. Concepts such as the Law of Large Numbers and the Central

SECTION

7.1

Limit Theorem provide the machinery to make predictions about these types of situations with confidence.

INTRODUCTION

CONTINUED

One of the most ubiquitous, and familiar, uses of probability is in gambling. Casinos are the ultimate “players” in using mathematics to foresee the results of a series of events that, taken individually, are functionally random. Indeed, the mathematics of probability ensures that while an individual gambler may have a good night or a lucky streak, in the long run, “the house always wins.” Have you ever wondered how Las Vegas seems to have vast amounts of money to spend on glitzy hotels and golf courses in the middle of the desert? Gambling is a large, lucrative business, and its success is due, in part, to the laws of probability.

In this unit we will see how probability, the mathematical study of the seemingly unpredictable, has developed over a period of time to become an extremely valuable tool in our modern world. We will see its relatively late origins in European games of chance and its most recent applications in modeling and understanding our increasingly complex and unpredictable world. We will ponder how it is that news networks are able to predict the winners of elections before all the votes have been counted. By the end of this unit, we will have a sense of how mathematics can be used to make accurate predictions about unpredictable events.

SECTION 7.2

HISTORY

- Roll Dem Bones
- A Rather Late Development

ROLL DEM' BONES

- The mathematical study of probability probably was delayed for centuries because of mysticism.

Throughout the ages, people have responded to the problem of what to do about the future in different ways. For many ancient societies, the unknown future was considered to be the province of the gods. Understanding and making predictions about this future was left to religious figures and oracles. These people employed a number of methods and devices with which they supposedly divined the will of the gods.

Some of the most common tools of the ancient religious diviner were astragali. These were bones, taken from the ankles of sheep, that would be cast and interpreted. Astragali commonly had six sides, but they were very asymmetrical. Often they were cast in groups, with the specific combinations of values revealing the name of the god who could be expected to affect the future affairs of the people. For example, if the bones said that Zeus was at work, there would be reason for hope. If the bones said that Cronos was in charge, then the people knew to prepare for the worst.

SECTION 7.2

HISTORY CONTINUED



Item 2242 / Kathleen Cohen, KNUCKLE-BONES (2008). Courtesy of Kathleen Cohen.



Item 2241 / Kathleen Cohen, DICE; CHECKERS; AND ASTRAGALUS (FOR KNUCKLE-BONES). (2008). Courtesy of Kathleen Cohen.

SECTION 7.2

HISTORY CONTINUED



Item 2240 / Kathleen Cohen, DICE (2008). Courtesy of Kathleen Cohen.

Gradually, technology enabled the development of more-regularly-shaped “prediction” devices. The first dice, made of pottery, are thought to have appeared in ancient Egypt. By the time of the flowering of Greek culture, dice were quite common, both for fortune-telling and for gaming. Dice have always been popular tools in recreational gaming, or gambling, precisely because they are thought to be random-event generators. “The roll of the dice” is thought to be the ultimate unknown, so dice are thought to be somewhat fair arbiters. This assumes of course that the dice are perfectly symmetrical and evenly weighted, which early dice often were not. Discoveries of ancient loaded dice reveal that, even though ancient people did not have a mathematical understanding of probability, they knew how to weight games in their favor.

One might think that the Greeks, who embraced a central role for mathematics in the world of the mind, would have discovered the features of probabilistic thinking. Evidence shows that they did not. It is thought that the Greeks deemed matters of chance to be the explicit purview of the gods. According to this view, they believed that any attempt to understand what happens and what should happen was a trespass into the territory of the gods. It was not of human concern.

SECTION 7.2

HISTORY CONTINUED

Additionally, the Greeks favored understanding based on logical reasoning over understanding based on empirical observations. One of the concepts at the heart of our modern understanding of probability is concerned with how actual results compare with theoretical predictions. This type of empirical thinking often took a back seat to logical axiomatic arguments in the mathematics of ancient Greece.

The mathematics of probability went undiscovered for centuries, but gambling, especially with dice, flourished. It seems that dice, in some form or another, have been a constant feature of civilization from the time of the Greeks onward. The Romans were fond of them, as were the knights of the Middle Ages, who played a game called Hazard, an early forerunner of the modern game of craps, thought to have been brought back from the Crusades.

A RATHER LATE DEVELOPMENT

- Renaissance mathematicians took the first strides toward understanding chance in an abstract way.
- Pascal's and Fermat's solutions to the "Problem of the Points" provided an early glimpse of how to use mathematics to say definite things about unknown future events.
- Tree diagrams are useful for keeping track of possible outcomes.

It was not until the Renaissance that fascination with dice as an instrument of gambling led to the first recorded abstract ideas about probability. The man most responsible for this new way of thinking was a quintessential Renaissance man, an accomplished doctor and mathematician by the name of Girolamo Cardano.

Cardano was famous in the mathematical world for many things, most notably his general solutions to cubic equations. As a doctor, he was among the best of his day. His passion, however, was to be found at the dice table. He was a fanatic and compulsive gambler, once selling all of his wife's possessions for gambling stakes. Out of his obsession grew an interest in understanding analytically, and, thus, mathematically, the odds of rolling certain numbers with dice. In particular, he figured out how to express the chances of something happening as a ratio of the number of ways in which the event could happen to the total number of outcomes.

SECTION 7.2

HISTORY CONTINUED

For example, what's the probability of rolling a 4 with two regular dice?

There are thirty-six possible equally likely outcomes when a pair of dice is rolled, and of these only three combinations (1 and 3, 2 and 2, and 3 and 1) produce a total value of 4. So, the probability of rolling a 4 is $\frac{3}{36}$ or $\frac{1}{12}$. This seemingly straightforward observation was the first step toward a robust mathematical understanding of the laws of chance. Dr. Cardano penned his thoughts on the matter around 1525, but his discoveries were to go unpublished until 1663. By that time, two Frenchmen had already made significant progress of their own.

In the mid-1600s, the Chevalier de Méré, a wealthy French nobleman and avid gambler, wrote a letter to one of the most prominent French mathematicians of the day, Blaise Pascal. In his letter to Pascal, he asked how to divide the stakes of an unfinished game. This so-called “problem of the points” was framed as follows:

Suppose that two men are playing a game in which the first to win six points takes all the money. How should the stakes be divided if the game is interrupted when one man has five points and the other three?

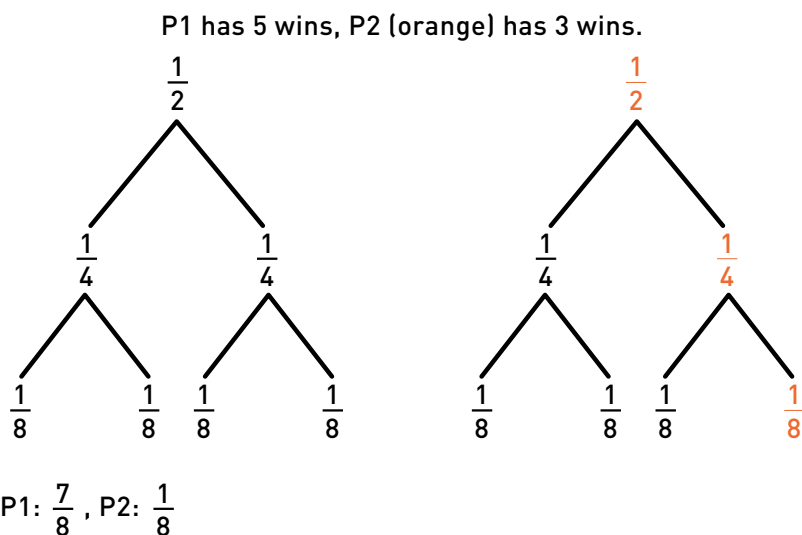
Pascal consulted with Pierre de Fermat, another very prominent mathematician of the day, in a series of letters that would become the basis for much of modern probability theory. Fermat and Pascal approached the problem in different ways. Fermat tended to use algebraic methods, while Pascal favored geometric arguments. Both were concerned basically with counting. They figured that in order to divide the stakes properly, they could not simply divide them in half, because that would be unfair to the man who was in the lead at the time of the game's cessation. A proper division of the stakes would be based on how likely it was that each player would have won had the game continued to completion.

The player in the lead could win in one more round. The trailing player would need at least three more rounds to win. Furthermore, he must win in each of those three rounds. Therefore, the two Frenchmen reasoned, the proper division of the stakes should be based on how likely it was that the trailing player would win three rounds in a row.

SECTION 7.2

HISTORY CONTINUED

The trailing player has a one-in-two chance of winning the next round. Provided he wins, he then has another one-in-two chance of winning the following round. So, after two rounds, there are four possible outcomes, only one of which is favorable to the trailing player. If he should happen to win the first two rounds, he then again has a one-in-two chance of winning the third round. Let's take a look at how all of this information can be represented in a tree diagram.



As we see in the tree diagram above, only one of the eight possible outcomes results in the trailing player winning the stakes. Therefore, the trailing player should be awarded $\frac{1}{8}$ of the pot, with the remaining $\frac{7}{8}$ going to the player who was winning at the time of the interruption. This method of enumerating and examining the possible outcomes of random events was a crucial link in the mathematical conquest of the unpredictable.

SECTION 7.3

SIMPLE PROBABILITY AND COUNTING

- A Coin Toss
- The Galton Board
- Enter Pascal's Triangle

A COIN TOSS

- Simple probability is the ratio of favorable outcomes to the total number of possible outcomes.

Probability theory enables us to use mathematics to characterize and predict the behavior of random events. By “random” we mean “unpredictable” in the sense that in a given specific situation, our knowledge of current conditions gives us no way to say what will happen next. It may seem pointless to try to predict the behavior of something that we fundamentally characterize as unpredictable, but this is exactly what makes the mathematics of probability so powerful. Let's think about a coin toss. There is no way to predict the outcome of a single coin toss. In this sense it is a random event.¹ But we can say some definite things about it.

The first thing that we can say is that the outcome will definitely be either heads or tails. Putting this in mathematical terms, we say that the probability of the coin landing heads up or tails up is 1, or absolutely certain. An event of probability zero is effectively impossible. In the case of equally likely outcomes, such as in the dice example above, determining the probability that a particular outcome occurs basically involves counting. So, we do as Cardano did and compare the number of ways a specific outcome can happen to how many total outcomes are possible. The probability of the coin landing heads up would then be 1 out of 2, or $\frac{1}{2}$. There is, of course, the same probability that it will land tails up. To see this we could start with the probability that the coin will be heads or tails, 1, and subtract the probability that it will be heads, $\frac{1}{2}$. This leaves us with $\frac{1}{2}$ as the probability that the coin will not land heads up, in other words, the probability that it will land tails up.

¹ Now, we have to be a bit careful here because a coin toss would not be random if we were able to know all of the initial conditions of the toss, but since we can't know all of the conditions that affect the outcome, we can treat it as random.

SECTION 7.3

SIMPLE PROBABILITY AND COUNTING CONTINUED

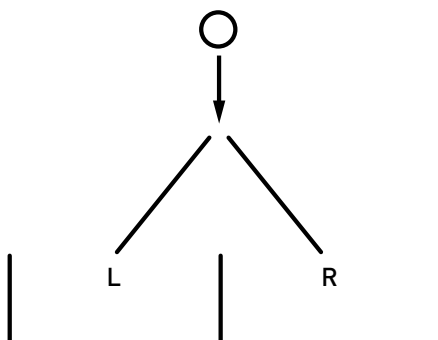
Determining the probability of the optional outcomes of a single coin flip may not seem that interesting, but it provides a good starting point for understanding the probabilities associated with any event or series of events that have only binary outcomes, (e.g., heads or tails, win or lose, on or off, left or right).

We can abstract any such situation in the form of a simple machine known as a Galton board.

THE GALTON BOARD

- The Galton board is a model of a sequence of random events.
- Each marble that passes through the system represents a trial consisting of as many random events as there are rows in the system.

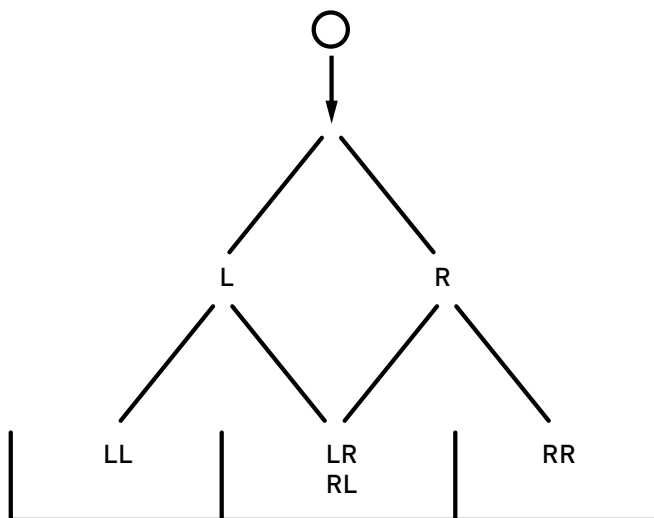
Imagine a peg fixed in the middle of an inclined board, with the base of the board divided into two bins of equal size. If we drop a marble towards the peg, it will hit it and deflect either into the right bin or the left bin. In terms of probability, this is just like our coin toss from before.



The two bins represent the possible outcomes for the dropped marble, and because there is only one way to get into either bin, the probability associated with a marble ending up in a particular bin is $\frac{1}{2}$. The advantage of viewing binary systems in this way is that it is very easy to build complexity into the experiments by adding rows of pegs. Let's add a row of two pegs below the initial peg, one to the right and one to the left.

SECTION 7.3

SIMPLE PROBABILITY AND COUNTING CONTINUED

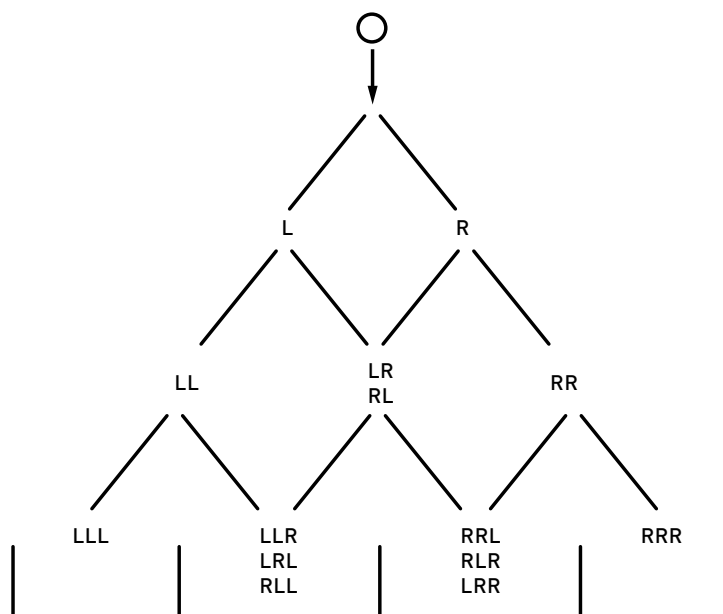


Notice that there are now three bins at the base. We can, as before, figure out the probability of the marble ending up in any particular bin. We may be tempted to think that all three bins are equally likely destinations, which would make the probability for any individual bin $\frac{1}{3}$. This ignores the fact that if we look at the paths a marble can take through the machine, we find that there are two possible paths to the middle bin, whereas there is only one path leading to each of the side bins. This suggests to us that we need to count the paths to bins instead of the bins themselves. With such a simple system, enumerating the paths is straightforward: LL, LR, RL, RR. With four possible paths, two of which end up in the middle bin, the probability of the marble ending up in the middle bin is 2 out of 4, or $\frac{1}{2}$. Because each side bin has only one path associated with it, the probability of the marble ending up in one particular side bin is 1 in 4 or $\frac{1}{4}$. If we add all the probabilities together, we get the probability that the marble will end up in one of the three bins: $\frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1$.

SECTION 7.3

SIMPLE PROBABILITY AND COUNTING CONTINUED

This is what we would expect, because the marble cannot disappear and must, therefore, end up in one of the three bins. To represent more-involved binary outcome systems, we can continue adding rows to our machine:



Here we have shown all possible ways that the marble can traverse three rows of pegs. In each row the marble hits one peg and deflects either right or left. This is a good model for any collection of three binary decisions, such as our problem of points from the last section. If instead of left and right, you imagine each peg represents win or lose, you have a nice model of the three rounds that the two players might hypothetically play to finish their game. Let's call each deflection to the left a win for Player 1 and each deflection to the right a win for Player 2. As we said earlier, Player 2 needs three consecutive wins, so the only path that would lead him to victory would be the RRR path. Because this is just one of eight possible paths, his chances of winning are 1 in 8. This means that if the game is interrupted at this point, Player 2, the one who is behind, should get $\frac{1}{8}$ of the pot.

To see why, let's imagine that someone wishes to take Player 2's place in the game, even though he needs an unlikely three consecutive points to win. How much should this newcomer pay to get into the game (which is the same as asking how much Player 2 must accept to get out of the game)?

SECTION 7.3

SIMPLE PROBABILITY AND COUNTING CONTINUED

Using the language of the Galton board, the newcomer would need the sequence RRR to win the entire pot—any other sequence results in a loss of however much she paid Player 2 to get into the game. Let's say that this newcomer is rather cautious and shrewd and, knowing that she is at a disadvantage, wishes to hedge her main wager (her payment to Player 2) with a series of side bets with onlookers at the contest. Because there are eight possible outcomes, and she can win on only one of them, she should place seven side bets to cover every possible outcome. Each side bet is a 50/50 bet on whether a certain sequence of events will happen. If the pot is \$8, the newcomer should make the following side bets:

Onlooker A pays \$1 if LLL happens and gets \$1 if RRR happens.

Onlooker B pays \$1 if LLR happens and gets \$1 if RRR happens.

Onlooker C pays \$1 if LRL happens and gets \$1 if RRR happens.

Onlooker D pays \$1 if RLL happens and gets \$1 if RRR happens.

Onlooker E pays \$1 if LRR happens and gets \$1 if RRR happens.

Onlooker F pays \$1 if RLR happens and gets \$1 if RRR happens.

Onlooker G pays \$1 if RRL happens and gets \$1 if RRR happens.

In the event that RRR happens, the newcomer would win the \$8 pot and owe a total of \$7 on all of her side bets, resulting in a gain of \$1. If any sequence other than RRR happens, the newcomer would get \$1 from one of her side-bets (the rest would be ties) and the pot would go to Player 1. No matter what happens, the newcomer ends up with \$1, so to enter the contest she should pay Player 2 no more than \$1.

The newcomer, of course, does not actually have to make all of the side bets. In fact, if she hopes to gain anything from a fair bet gamble, she shouldn't, because she would be guaranteed to break even. However, considering these bets, also known as hedges, helps in figuring out what the fair price is for entrance to the game. Besides, if someone offered her the opportunity to play for less than a dollar, then using these side bets she is guaranteed to make a profit. Such a guarantee of profit is called an arbitrage opportunity, and this view of probability as the hedge-able fair price plays a fundamental role in applying probability theory to finance. It also happens to be the way that early thinkers such as Fermat and Pascal viewed probability.

SECTION 7.3

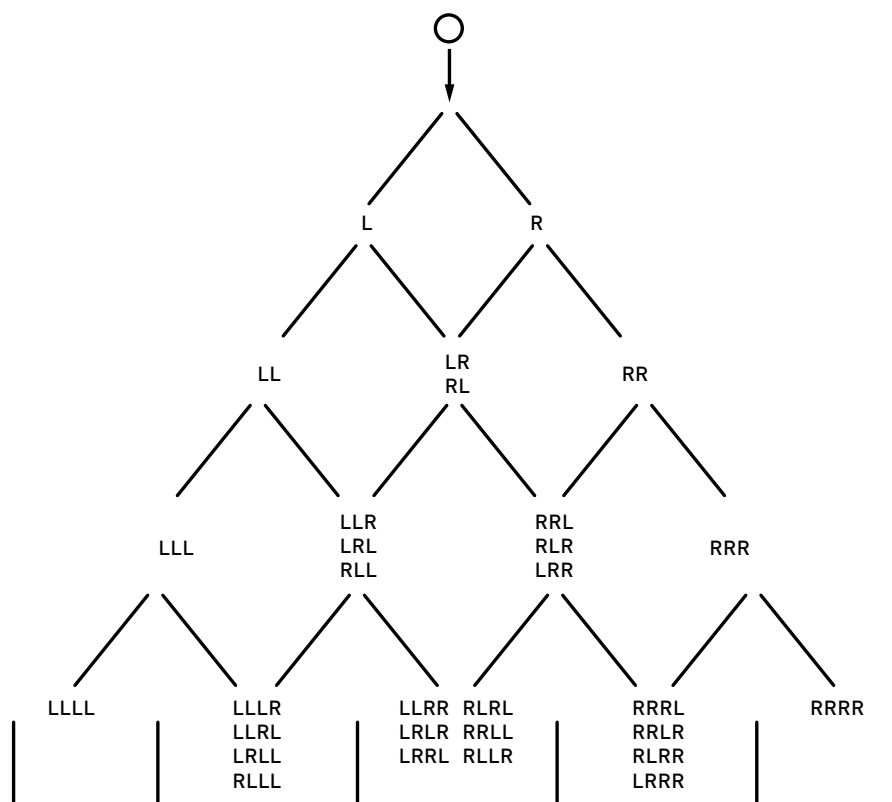
SIMPLE PROBABILITY AND COUNTING CONTINUED

The newcomer should pay \$1 to Player 2, which means that Player 2 could walk away from the game at this point with \$1, or $\frac{1}{8}$ of the total pot. So, this is also the amount Player 2 should get if the game is interrupted, because in both cases he is leaving the unresolved game.

ENTER PASCAL'S TRIANGLE

- For Galton boards with many rows, the task of enumerating paths is greatly facilitated by using Pascal's Triangle.

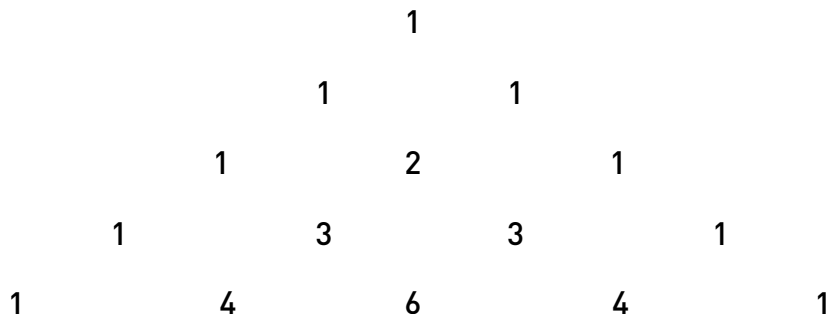
Let's now return to the Galton board and see what happens as we continue to add rows.



We notice that it quickly becomes unwieldy to enumerate every path. It would be nice to have some easy way to find the number of paths to each bin, given how many rows, or rounds, or decisions there are. Fortunately we can model this situation, as we did in our discussion of combinatorics in Unit 2, using Pascal's Triangle:

SECTION 7.3

SIMPLE PROBABILITY AND COUNTING CONTINUED



In Unit 2, we found the generalization that the number of paths from the top of Pascal's Triangle to the k^{th} "bin" in the n^{th} row is given by:

$$\frac{N!}{k!(n-k)!}$$

Let's verify that there are indeed six paths to the middle bin ($k=2$) of the fifth row ($n=4$). $\frac{4!}{2!(4-2)!} = \frac{4*3*2*1}{(2*1)(2*1)} = \frac{24}{4} = 6$.

Adding the path totals for each bin in the n^{th} row gives the total number of paths available to the marble, 16. The probability that the marble will end up in the middle bin is, therefore, $\frac{6}{16}$, or $\frac{3}{8}$.

Our discussion up until now has been quite theoretical. We have used the power of combinatorics to enumerate all the paths available to a marble traveling through the Galton box, and we have calculated probabilities associated with those paths, assuming each individual path has an equal probability. If we would actually perform such an experiment, however, we would have no way of knowing in which bin a single marble will end up. We can speak only generally. However, even this general view has great power, as we shall see in the next section.

SECTION 7.4

LAW OF LARGE NUMBERS

- A Change of Perspective
- What Did You Expect?

A CHANGE IN PERSPECTIVE

- Bernoulli's Law of Large Numbers shifted the thinking about probability from determining short-term payoffs to predicting long-term behavior.

In the preceding section, with the use of the Galton board, we found the theoretical probability that a marble will end up in any specific bin. Now let's turn our attention to what actually happens when we let a marble go through the board; furthermore, let's see what happens when many marbles go through it!

Each path is equally likely, and we have to assume that marbles dropped randomly into our machine are not predestined to follow any particular path. Because the number of paths to each of the bins varies, we should expect that over time, bins that have more paths leading to them will end up with more marbles than bins that have fewer paths leading to them. Thus, the distribution of a large number of marbles through the machine will not be even. The Law of Large Numbers will help us to predict roughly the distribution that we would find were we to run such an experiment.

The Law of Large Numbers says that when a random process, such as dropping marbles through a Galton board, is repeated many times, the frequencies of the observed outcomes get increasingly closer to the theoretical probabilities. Jacob Bernoulli, the man who is credited with discovering this law around the beginning of the 18th century, is said to have claimed that this observation was so simple that even the dumbest person knows it to be true. Despite this pronouncement, it took him over 20 years to develop a rigorous mathematical proof of the concept.

SECTION 7.4

Let's look at the Law of Large Numbers in terms of the Galton board:

LAW OF LARGE NUMBERS CONTINUED

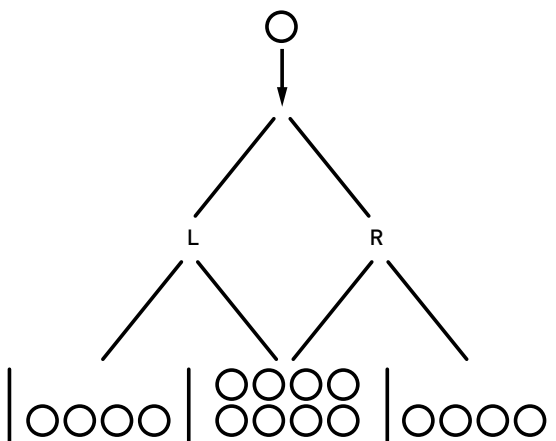


Diagram Showing Lots of Balls Going Through the 2 Row Galton Board

Recall that we found the probability of a ball ending up in the middle bin to be $\frac{1}{2}$. According to the Law of Large Numbers, if we ran 100 marbles through this setup, about 50 of them would end up in the middle bin. If we ran 1000 marbles, about 500 would end up in the middle. Furthermore, as we run more marbles through the board, the proportion in the middle bin will get closer and closer to $\frac{1}{2}$.

Bernoulli may have thought that this concept is self-evident, but it nevertheless is striking. Recall that we can't say with any certainty at all where one particular marble will end up. Still, we can say with very high accuracy how 1000 marbles will end up. Better yet, the more marbles we run, the better our prediction will be. The Law of Large Numbers is a powerful tool for taming randomness.

SECTION 7.4

LAW OF LARGE NUMBERS CONTINUED

WHAT DID YOU EXPECT?

- Expected value is the average result of a probabilistic event computed by taking into account possible results and their respective probabilities.
- Expected value is a key concept in both gambling and insurance.

The notion of expected value, or expectation, codifies the “average behavior” of a random event and is a key concept in the application of probability. For example, imagine that you are a door-to-door salesperson. Your experience tells you that the probability of making a sale, and thus a commission, on each try is as follows: $\frac{8}{10}$ that you make no sale and make no commission, $\frac{3}{20}$ that you make a small sale that leads to a \$100 commission, and $\frac{1}{20}$ that you make a large sale that leads to a \$500 commission.² How much can you expect to make, on average, per appointment? That is, what do you “expect” to be the value of the total sales divided by the number of appointments? This will be an expected value.

The expectation or expected value of a random process in which each outcome has a particular payoff is simply the sum of the individual probabilities multiplied by their corresponding payoffs. If P_1 is the probability of outcome #1 and V_1 is the payoff value of that outcome, and so on (P_j and V_j for the j^{th} outcome and j th payoff value respectively), then the expected value can be represented by the expression:

$$P_1V_1 + P_2V_2 + \dots + P_NV_N \quad \text{where } n = \text{number of possible outcomes}$$

In our sales example, the individual terms are as follows: for a no-sale, $\frac{8}{10} \times \$0 = \0 ; for the small commission, $\frac{3}{20} \times \$100 = \15 ; and for the large commission, $\frac{1}{20} \times \$500 = \25 . Thus, the expected value of the sales call payoff, per appointment, is $0 + \$15 + \$25 = \$40$. This, of course, does not mean that you will make \$40 for every appointment, but it is what you can “expect” to make on average over a period of time (assuming that your probabilities are correct!).

² Note that, colloquially, it might seem more natural to speak in terms of “odds,” as in the “odds of making a sale.” Odds are computed as ratios of alternative probabilities. If the probability of not making a sale is $\frac{8}{10}$, then the probability of making a sale is $\frac{2}{10}$, and the odds of not making a sale are $(\frac{8}{10})/(\frac{2}{10})$, more commonly expressed as “4 to 1.”

SECTION 7.4

LAW OF LARGE NUMBERS CONTINUED

The Law of Large Numbers ensures that the more sales calls you make, the closer your average payoff, per appointment, will be to \$40.

The concept of expected value, in conjunction with the Law of Large Numbers, help form the operating principle of businesses that are based on risk. Two prominent examples are casinos and insurance companies. Let's look a little more closely at each.

Any single casino game carries a certain risk for both the player and the house. A player's loss is the house's gain and vice versa. It would seem that no business could thrive in such a zero-sum situation, yet generally, the casino business is quite lucrative. This is possible because while the individual player's risk is concentrated in a small number of hands or rounds of a game, the casino's risk is spread out among all the games and all the bets going on. In short, casinos have the Law of Large Numbers working in their favor. Owners and managers of casinos know that while the outcome of any single game is unpredictable, the outcome of many rounds of that same game is entirely predictable. For instance, they know that the probability of rolling a seven at the craps table is $\frac{1}{6}$. Averaging this over many rolls means that a player will roll a seven $\frac{1}{6}$ of the time. In other words, in a group of six players, only one, on average, will be a winner. The casinos then structure their payoffs or "odds" slightly in their favor so that the money paid out to any player who wins will be more than offset by the money taken in from the five players who, on average, don't win. Note that this does not require any sort of rigging or cheating as far as actual game play. Casinos don't need to cheat the individual gambler-as long as they keep their doors open, the odds settle in their favor. They've structured their payoffs to guarantee it in the long run and because they generally have more working capital than any of the players, they can take advantage of the long-term "guarantees."

Insurance companies use similar principles to set premiums. They spend a great deal of effort and resources calculating the odds of certain catastrophes, such as a house fire, then multiply this value by the payoff they would give in such an event. This amount is how much the company can expect to have to pay, on average, for each person that they cover. They then set their rates at levels that cover this "expense" in addition to providing their profit. The policyholder gets peace of mind because the insurance company has effectively mitigated the risk of potential loss in a given catastrophe.

SECTION 7.4

LAW OF LARGE NUMBERS CONTINUED

The insurance company gets a flow of regular payments in exchange for a massive payoff in the unlikely event of a big claim.

The Law of Large Numbers is a powerful tool that enables us to say definite things about the real-world results of accumulated instances of unpredictable events. This useful tool represents just one example of how mathematics can be used to deal with randomness. The Law of Large Numbers applies to specific outcomes and their probabilities, but what about the entire range of possible outcomes and their associated probabilities? Just as the frequency of a specific event will tend toward its probability over the long run, the full set of possible outcomes will each tend to their own probabilities. Studying the distribution of possible outcomes and probabilities will give us even more powerful tools with which to predict long-term average behavior.

SECTION 7.5

THE GALTON BOARD REVISITED

- Distribution
- Tilting the Board
- 68-95-97.5

DISTRIBUTION

- Each bin of the Galton board has an associated probability, and looking at all of the bins simultaneously gives a distribution.
- Because each peg represents a right-or-left, or binary, decision/option, the distribution of probabilities is called a binomial distribution.

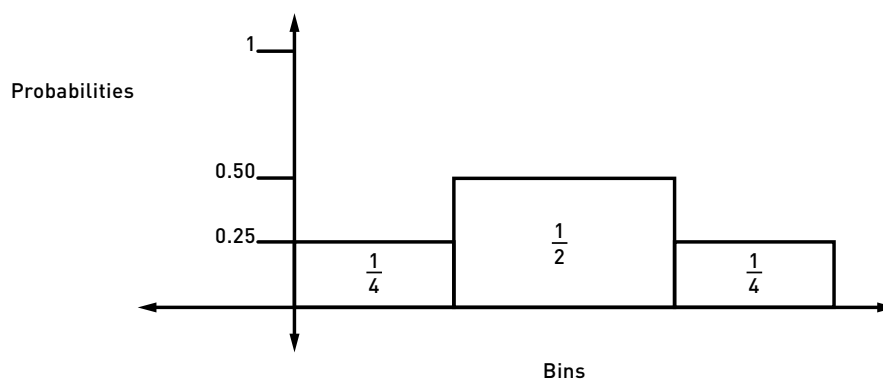
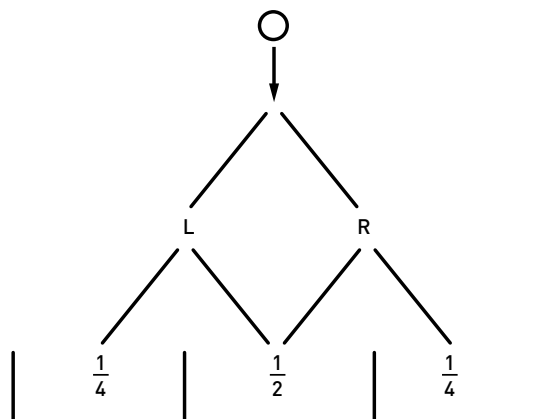
Let's return to our Galton board for further exploration. In the previous section, we used the Law of Large Numbers to see that, for each bin, the theoretical and experimental probabilities get closer and closer to one another as we put more and more marbles through the process. We're going to return to theoretical probabilities now and examine how the probabilities are distributed across all of the bins.

Recall that in our earlier examples the bins in the middle had higher probabilities than did the bins at the sides. This was because there are more paths that terminate in the middle bins than terminate in the side bins. Let's make a histogram that correlates to the bins and their probabilities.

SECTION 7.5

First a simple 2 row machine:

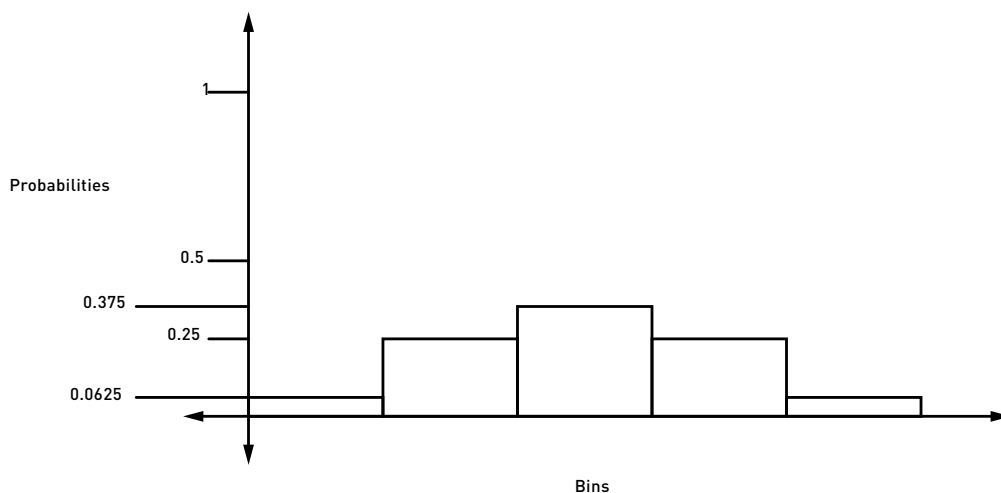
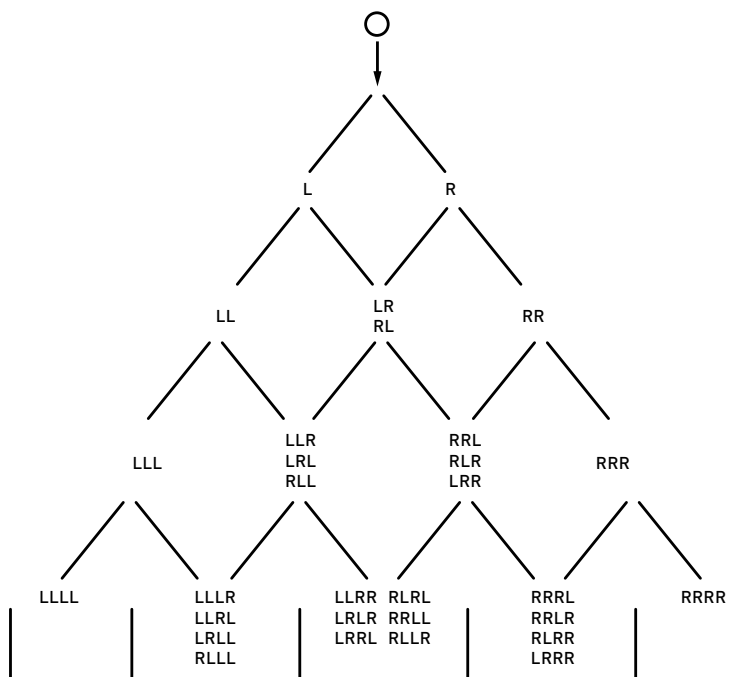
THE GALTON BOARD REVISITED CONTINUED



Notice that the distribution is symmetric; the probabilities for both the far right and far left bins are equal. Let's look at the histogram for a Galton board with four rows.

SECTION 7.5

THE GALTON BOARD REVISITED CONTINUED

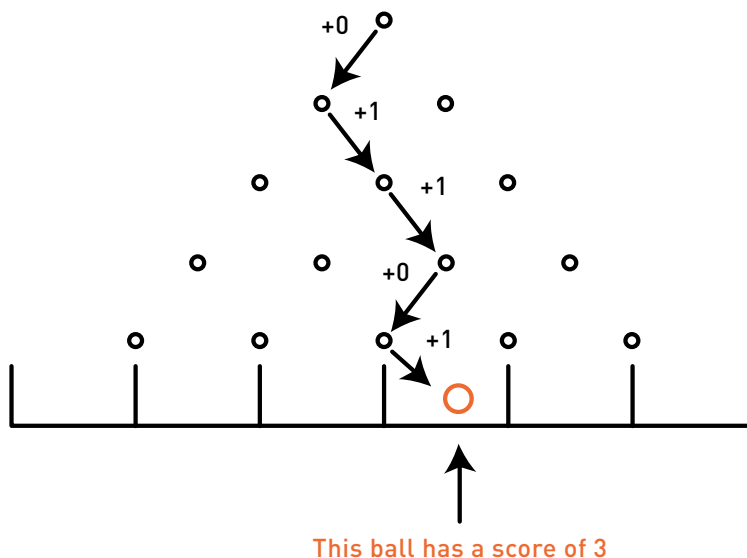


Notice that the probabilities for each version of the system are distributed across all of the bins and that even though the individual probabilities change as we add more bins, they always sum to 1. This is in keeping with our intuition that any marble must indeed end up in exactly one of the bins. At this point, we're going to need to label the bins so that we can discuss results in more detail.

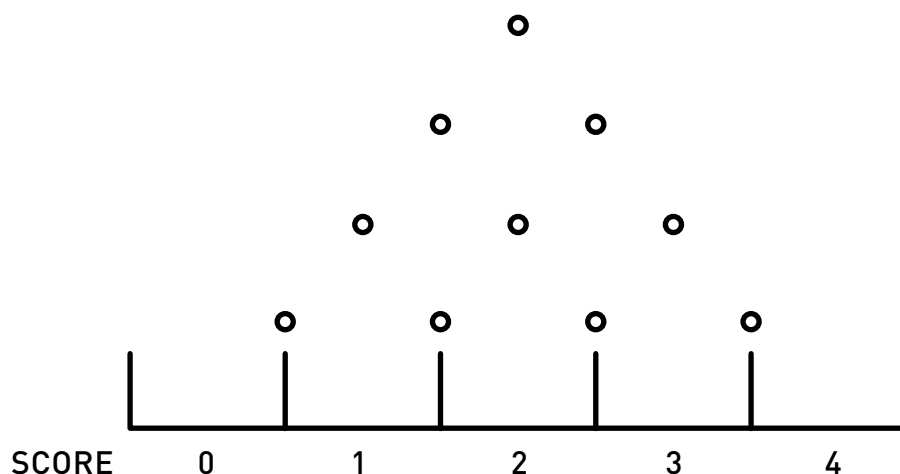
SECTION 7.5

THE GALTON BOARD REVISITED CONTINUED

To do this, let's say that each marble, as it progresses through the system, gets 1 point for each movement to the right and 0 points for each movement to the left. Each bin can then be represented by the summative "score" of a marble that ends up there.



For example, a marble passing through a four-row system would have a maximum possible score of 4, corresponding to the right-most bin, and a minimum possible score of zero, corresponding to the left-most bin. The remaining bins would have scores as follows:



SECTION 7.5

THE GALTON BOARD REVISITED CONTINUED

The average result, which would be the expected score for one average marble, can be found as before by multiplying the value of each bin by the probability of landing there and summing the results. This would give us

$$\left(\frac{1}{16} \times 0\right) + \left(\frac{4}{16} \times 1\right) + \left(\frac{6}{16} \times 2\right) + \left(\frac{4}{16} \times 3\right) + \left(\frac{1}{16} \times 4\right) = \frac{32}{16} = 2.$$

Interestingly, we can also find this by looking at the number of rows and multiplying by the probability of going right. This would be 4 rows times $\frac{1}{2} = 2$. Recall that the number of rows corresponds to the maximum score that a marble can get. Multiplying this maximum score by the average probability at each peg gives an expected average value. We'll need this method in just a bit.

Furthermore, we want to mathematically describe how the values of the scores of each marble are spread out around the mean. In other words, we need a way to describe how random results vary from their expected value. This would give us a sort of sensible ruler that we can use. To do this, it makes sense to look at the difference between the expected value of each marble and the mean, so we subtract the two quantities.

$$x - m \quad \text{where } x = \text{expected value and } m = \text{mean}$$

Because this quantity is something like a notion of distance, we square it to ensure that the value won't be negative.

$$(x - m)^2$$

We then multiply this squared difference by the probability of ending up in that bin.

$$P(x - m)^2$$

Finally, if we add all of these terms together, we will get a number that describes how the expected values of the bins are distributed around the mean. This is known as the variance.

$$\sum_{i=1}^n (x_i - m)^2 P_i$$

Because the variance is based on the square of the difference between a result and its expected value, it scales somewhat awkwardly.

SECTION 7.5

THE GALTON BOARD REVISITED CONTINUED

For example, if the difference between the expected value and the mean changes by a factor of three, the variance would change by a factor of 3^2 , or 9. To mitigate this so that our ruler scales in a more sensible manner, we can take the square root of our final result. Taking the square root of the variance gives us a measure of the average difference between a marble's score and the mean. This is known as the standard deviation.

The number of bins corresponds to the number of rows in the system. Let's call this number n . Notice that the maximum score is also n . Remember that this setup can represent many different situations, such as the result of n coin tosses, or any other situation, regardless of whether or not the odds of each individual event are 1 to 1, sometimes referred to as a "50-50 chance."

TILTING THE BOARD

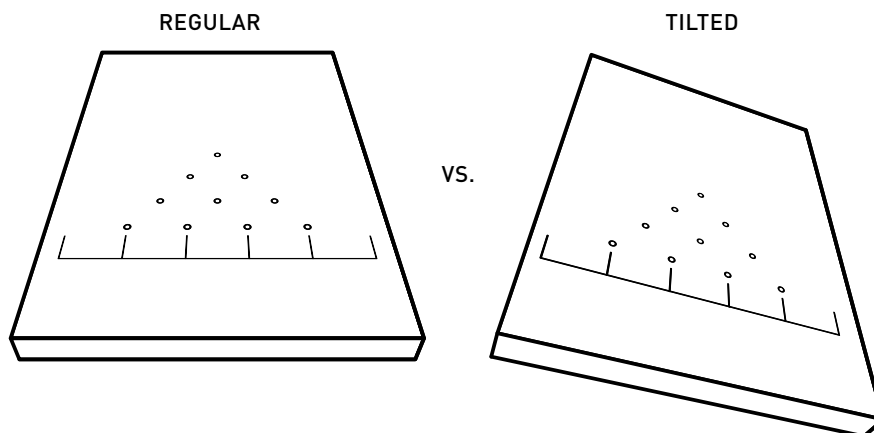
- If we assign a probability other than $\frac{1}{2}$ to each peg of the Galton board, the mean of the distribution will shift.

What would happen in a situation in which each individual event has a probability other than $\frac{1}{2}$? Let's return to dice for a moment, and then see how we can model this on our Galton board. For instance, let's say we want to roll a 5 with one die. We either roll a 5 or we don't, so this situation is binary, but unlike the coin toss, the odds are not 1 to 1. The probability of rolling a 5 with one die is $\frac{1}{6}$ —there is only one way to roll a 5, whereas there are five possible results other than a 5. We can model this using a Galton board by equating right deflections with rolling a 5 and left deflections with rolling anything else. If we then tilt the board such that each marble has a $\frac{1}{6}$ chance of going to the right at each peg and a $\frac{5}{6}$ chance of going to the left, we have a great model for our problem—it's like having a "biased coin," one in which the probability of getting a head is only $\frac{1}{6}$ and the probability of getting a tail is $1 - (\frac{1}{6})$, or $\frac{5}{6}$.

With this model, it is easy to answer questions such as, what is the probability of rolling a 5 exactly once in four rolls? In terms of our modified, tilted system, this correlates to a marble going through four rows and deflecting to the right only once, ending up in bin 1. To find the probability that a marble will end up in bin 1, which is the same probability of rolling one 5 in four rolls, we can no longer simply count paths as we did before, because not all paths are equally likely. Nonetheless, we can use our path count as a starting point.

SECTION 7.5

THE GALTON BOARD REVISITED CONTINUED



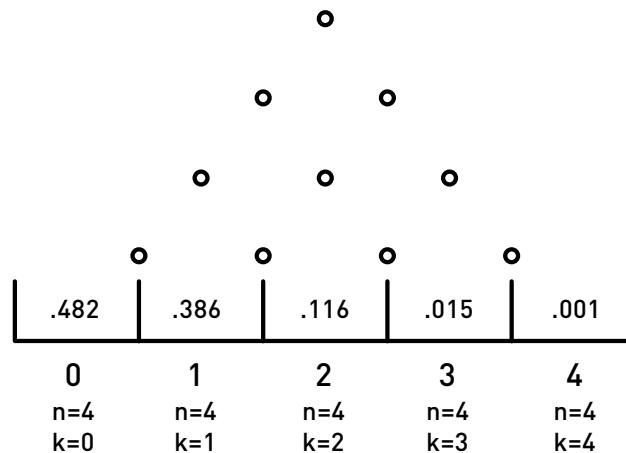
In our four-row system, we know that there are four possible paths to bin 1. Now, instead of looking at the ratio of the number of paths ending in bin 1 to the total number of paths to find the probability of ending up in bin 1, we can think about the probability of each specific path occurring. Each path is a sequence of four events, and each event is either a left (L) or right (R) shift in direction. The four paths to bin 1 are thus, LLLR, LLRL, LRLL, RLLL. The probability for each of these paths is the product of the probabilities of the individual events in the sequence. For example: the path LLLR has a probability of $\left(\frac{5}{6}\right) \times \left(\frac{5}{6}\right) \times \left(\frac{5}{6}\right) \times \left(\frac{1}{6}\right)$. The path RLLL has the probability $\left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right) \times \left(\frac{5}{6}\right) \times \left(\frac{5}{6}\right)$. Notice that all the paths to bin 1 have the same probability. Therefore, to find the probability of ending up in bin 1, we can just add the probabilities of taking the specific paths that end in bin 1. Since all of these probabilities are the same, we can simply multiply the probability for one path, $\frac{125}{1296}$, times the number of paths (4) to get $\frac{500}{1296}$.

SECTION 7.5

THE GALTON BOARD REVISITED CONTINUED

We can generalize this thinking to arrive at an expression that will tell us the probability of landing in the k^{th} bin of a system with n rows, in which the probability of going to the right at each peg is p and the probability of going to the left is $1-p$. We multiply the number of paths, $\frac{n!}{k!(n-k)!}$, times the probability of going right, p , to the k^{th} power, times the probability of going left, $(1-p)$ to the $(n-k)^{\text{th}}$ power (because if you go right k times, you necessarily go left the rest of the time). The probability of landing in the k^{th} bin is then:

$$\frac{n!}{k!(n-k)!} \times p^k \times (1-p)^{(n-k)}$$



Using $p = \frac{1}{6}$ and $(1-p) = \frac{5}{6}$, from our dice example, we see that the distribution of probabilities after four rows on the Galton board has shifted to the left somewhat from what it was for the $p = (1-p) = \frac{1}{2}$ situation of the fair coin toss. Intuitively it makes sense that, if a marble has a greater chance of going left than right at each peg, then there is a greater chance that it will end up in the left bins.

Let's look at how this affects the average marble's score. We'll need to find the mean again, and we can do this, as we did before, by multiplying the number of rows by the probability of deflecting to the right. ($4 \text{ rows} \times \frac{1}{6} = \frac{2}{3}$).

So, shifting the probability at each peg from $\frac{1}{2}$ to $\frac{1}{6}$ both moves the entire distribution of probabilities to the left and shifts the mean value from 2 to $\frac{2}{3}$. We now see how the probability of each event (turn) determines the overall distribution of outcomes of repeated events (sequences of turns).

SECTION 7.5

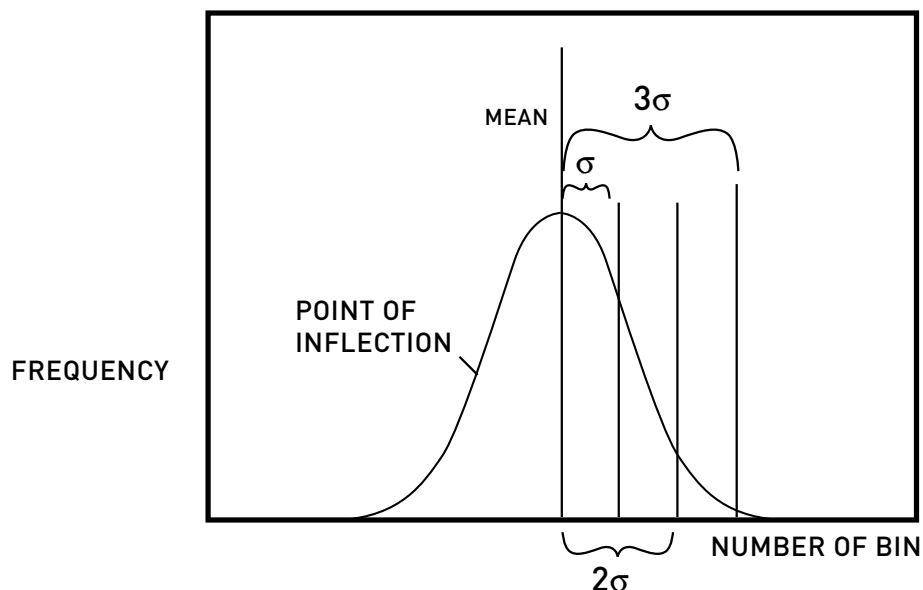
THE GALTON BOARD REVISITED CONTINUED

Not only does the mean shift, but the variance and standard deviation shift as well. Recall that these values have to do with how the outcomes are distributed around the mean. This distribution of probabilities, or outcomes, is called the binomial distribution, and it is a commonly occurring distribution in sequences of repeated events in which there are only two possible outcomes for each event.

68-95-97.5

- The normal distribution is an ideal distribution that is determined by only its mean and standard deviation.

The binomial distribution is useful, but it can take a long time to calculate, especially in situations in which n , the number of events, or the number of rows in the system, is large. There is an approximation to this distribution, however, that is much more easily calculated and that provides a reasonably good model for the probability distribution. It can be found using only the mean and the standard deviation, and it is known as the normal distribution, familiar to many of us as the “bell curve.”



The normal distribution is related to a model of the distribution of the probabilities of outcomes of repeated independent events—also called “Bernoulli trials.” As we can see, it is a bell-shaped curve, and it turns out that it is characterized by two properties. One distinguishing characteristic is its mean,

SECTION 7.5

THE GALTON BOARD REVISITED CONTINUED

which correlates to the central position of the bell around which it is symmetric. The other characteristic is the standard deviation. In the graph above, this corresponds to the position where we see a point of inflection on the graph (there is one on either side of the mean, indicated in the figure above). One standard deviation is the average difference between an outcome and the mean. In terms of percentages, the standard deviation, marked on either side of the mean, defines the range within which 68% of the results fall (on average). In other words, if scores on a test were normally distributed, about 68% of the students would fall within one standard deviation of the mean. For example, if the mean were 65 and the standard deviation 7, then 68% of students would score between 58 and 72. What's more, about 95% of students would have scores within two standard deviations of the mean, and about 97.5% of students would have scores within three standard deviations of the mean. For this example, only 2.5% of students would have scores higher than 86 or lower than 44. This is commonly known as the 68-95-97.5 rule for normal distributions.

The normal distribution approximation provides a powerful tool for predicting how the results of repeated independent experiments will be distributed. Furthermore, the more events in sequence that we look at, the better the normal distribution is at describing our results. Of course, there can always be outliers, such as a string of all heads or tails, that momentarily will skew the distribution one way or the other. However, on average, the normal distribution is fairly representative of the real world. In terms of our 50-50 Galton board, which can model a variety of binary situations, this means that the more rows we have, the closer our distribution will be to the normal distribution. The underlying reason for this involves the Central Limit Theorem, and it is to this concept that we will now turn.

SECTION 7.6

CENTRAL LIMIT THEOREM

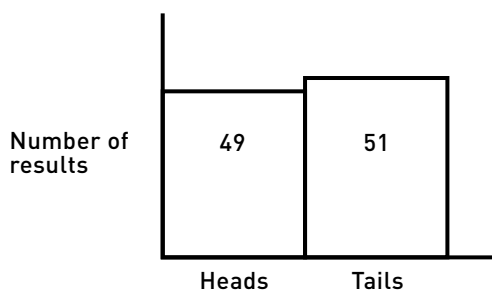
- An Average of Averages
- An Example from Politics

AN AVERAGE OF AVERAGES

- According to the Central Limit Theorem, the distribution of averages of many trials is always normal, even if the distribution of each trial is not.

Let's return to thinking about coin flips. If our coin is fair, the probability that the result will be heads is $\frac{1}{2}$, and the probability that the result will be tails is the same. If we flip the coin 100 times, the Law of Large Numbers says that we should have about 50 heads and about 50 tails. Furthermore, the more times we flip the coin, the closer we get to this ratio.

Result of 100 coin flip:
1 ball traversing a 100-row machine



The Law of Large Numbers says that this ratio goes to 50/50 as the number flips increase.

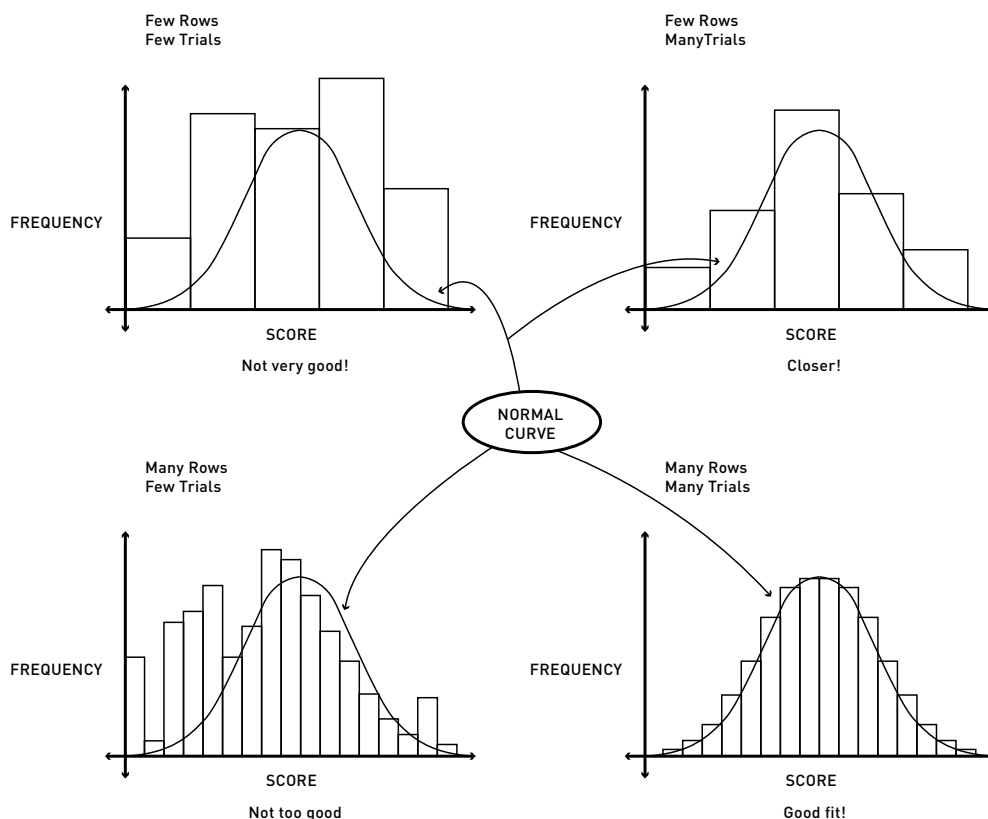
Let's now shift our thinking to consider sets of 100 coin flips. Flipping a coin 100 times is like running one marble through a 100-row version of our Galton board. Running many marbles through this system is like doing the 100-coin flip experiment many times, one for each marble. Instead of being concerned with each flip, or each left or right deflection of a marble, we are only concerned with the total result of 100 such individual events. According to the Law of Large Numbers, the more times we flip the coin, the closer our overall results will come to a 1-to-1 ratio of heads to tails.

SECTION 7.6

CENTRAL LIMIT THEOREM CONTINUED

However, if we cap our number of events at 100, and do multiple sets of 100 events, we will find that not all of the sets end up being an exact 50-50 split between heads and tails. Some will have more heads than tails and vice versa. It's also possible that a very few sets might come out all heads or all tails. To explain these results, we are going to need something a bit more powerful than the Law of Large Numbers.

What is amazing is that we can predict with a fair level of accuracy how many of these 100-flip tests should come out all heads, or all tails, or any mixture in between. In fact, the distribution of outcomes of our 100-flip tests will follow a normal distribution very closely. The guiding principle behind this reality is the Central Limit Theorem.



The Central Limit Theorem was developed shortly after Bernoulli's work on the Law of Large Numbers, first by Abraham De Moivre. De Moivre's work sat relatively unnoticed until Pierre-Simon Laplace continued its development decades later. Still, the Central Limit Theorem did not receive much recognition until the beginning of the 20th century. It is one of the jewels of probability theory.

SECTION 7.6

CENTRAL LIMIT THEOREM CONTINUED

The Central Limit Theorem can be quite useful in making predictions about a large group of results from a small sampling. For instance, in our sets of 100 coin flips, we don't actually have to do numerous rounds of 100 flips in order to be able to say with a fair amount of confidence what would happen were we to do so. We can, for instance, complete just one round of 100 flips, look at the outcome, say perhaps 75 heads and 25 tails, and ask, "how closely does this one experiment represent the whole?" This is essentially what happens during elections when television networks conduct exit polling.

AN EXAMPLE FROM POLITICS

- Exit poll results can be compared with a normal distribution to make predictions about the results of an election based on a relatively small sample of voters.

In an exit polling situation, voters are asked if they voted for a particular candidate or not. If you ask 100 voters and you find that 75 voted for Candidate A and 25 voted for Candidate B, how representative of the overall tally is this? The mean of this sample is 75% for Candidate A. This is calculated by assigning a score of 1 to a vote for Candidate A and a score of 0 to a vote for Candidate B, multiplying the votes by the scores, adding these results, and dividing by the total number of votes.

$$\text{Mean} = \frac{(75 \times 1) + (25 \times 0)}{100}$$

Intuition tells us that it would be unwise to assume that the final tally of all the votes will exhibit exactly the same ratio as this one sampling. That would be akin to flipping a coin 100 times, getting 75 heads, and assuming that this is what would happen more or less every time. In other words, we can't assume that the mean value of this one sample of 100 voters is the same as the true mean value of the election at large. Even so, we can say something about how the mean we found in the exit poll relates to the true mean.

We can use the Central Limit Theorem to realize that the distribution of all possible 100-voter samples will be approximately normal, and, therefore, the 68-95-97.5 rule applies. Recall that this rule says that 68% of sample means will fall within one standard deviation of the true mean (the actual vote breakdown of the whole election). However, this rule is useful only if we know the standard deviation and the true mean, and if we knew the true mean, why would we need to conduct an exit poll in the first place?

SECTION 7.6

CENTRAL LIMIT THEOREM CONTINUED

To find an approximation of the standard deviation, we must first find the variance. Recall from the previous section that the variance is related to the difference between how each person voted and the mean. Because the possible votes are only A or B, and A is assigned a score of 1 whereas B gets a score of 0, then the possible differences are “1 minus the mean,” which corresponds to the people who voted for A, and just the mean, which corresponds to the people who voted for B. The total number of voters multiplied by the mean is the total number of voters who voted for A. The total number of voters multiplied by “one minus the mean” is the total number of voters who voted for B. To find the variance, we square the differences, multiply by the vote proportions, add, and divide by the total number of votes. If the total number of votes is V , then the variance is:

$$\text{Var} = \frac{(1-\text{mean})^2 \times V \times \text{mean} + \text{mean}^2 \times V \times (1-\text{mean})}{V}$$

$$\frac{V \times \text{mean} - 2 \times V \times \text{mean}^2 + V \times \text{mean}^3 + V \times \text{mean}^2 - V \times \text{mean}^3}{V}$$

The V s cancel out and with a bit of algebra, we find:

$$\text{Var} = \text{mean} (1 - \text{mean})$$

The standard deviation is thus $\sqrt{\text{mean}(1-\text{mean})}$.

The mean in which we are interested here is the true mean, but as yet we have only a sample mean. Luckily, sample means and true means usually give standard deviations that are pretty close to one another, so we can use the standard deviation given by the sample mean to help us find approximately where the true mean lies.

We have now seen how probability theory can be used to make powerful predictions about certain situations. Up until this point, however, we have been chiefly concerned with simple, idealistic examples such as coin tosses, the rolling of dice, and quincunx machines. Let’s now turn our attention to probabilities that are more in line with what happens in the real world.

SECTION 7.7

OTHER TYPES OF PROBABILITY

- Let's Make a Deal
- Off the Chain

LET'S MAKE A DEAL

- Conditional probability applies to events that are not independent of one another.

Can you use what you know about the past to predict the future? When does past performance tell you about future returns? In roulette, the fact that the wheel lands on a red space eight times in succession has no bearing on the next spin of the wheel—even though we might be tempted to think that it does! Each spin of the wheel is an independent event. Many other situations in life do not exhibit such perfect independence, however. For instance, your chances of winning the lottery are greatly increased by purchasing a ticket, and your chances of being eaten by a shark are greatly reduced by staying on the beach. More realistically, what the weather will be doing in an hour depends to a large degree on what it is doing now. These examples and others like them come from the world of conditional probability.

A classic example of conditional probability is what is often referred to as the “Monty Hall Problem.” This is a situation in which a game show contestant is faced with three doors, one of which conceals a new car, and the other two of which conceal less desirable prizes, such as a donkey or a pile of sand. The contestant chooses a door, door number 2 let's say. Suppose that the host then opens door number 1 to reveal a pile of sand. Now, with two closed doors remaining, the host offers the contestant a chance to switch his/her selection to door number 3. Should the person switch?

The probability that switching one's selection will result in winning the car depends on the probability that one's initial selection was either correct or incorrect. The probability that your initial guess is correct is $\frac{1}{3}$. After the host narrows the choice, the probability that you were initially correct is still the same, $\frac{1}{3}$, which means that your probability of being initially incorrect, and thus the probability that switching your choice will prove fruitful, is $\frac{2}{3}$. After the host reveals one of the klunkers, we are now considering a conditional probability: the probability that the remaining door has the grand prize, given that one klunker has been revealed, is $\frac{2}{3}$.

SECTION 7.7

OTHER TYPES OF PROBABILITY CONTINUED

This is a much different result than if the host would reveal one of the non-winning doors prior to your first choice. In this scenario, your first choice would have a $\frac{1}{2}$ probability of being correct. If then given the option to switch, the probability that switching will be advantageous is only $\frac{1}{2}$. The fact that our original situation leads to the switch strategy presenting a higher probability of success may seem counter-intuitive, but one of the great strengths of probability theory is that it allows us to quantify the randomness that we are facing and gives us a rational and logical way to make decisions, one that is helpful in situations in which our intuition is often wrong.

OFF THE CHAIN

- Markov Chains provide a way to talk about sequences of events in which the probability of each event is dependent on the results of prior events.

A concept from probability that is similar to conditional probability, yet different in some important ways, is the Markov Chain. In a Markov Chain, the probability of a future event depends on what is happening now. The probability of the next event depends on what happened in the previous event. The outcome of a given experiment can affect the outcome of the next experiment.

Let's say it is raining in our current location. There is a certain probability that in ten minutes it will still be raining. There is also a certain probability that in ten minutes it will be sunny. These two probabilities, the rain-rain transitional probability and the rain-sun transitional probability, depend on many factors. If we want to project what the weather will be like in an hour, we can model this as a succession of six 10-minute steps. Each state along the way will affect the probabilities for transitioning to another state. The rain-sun transition's probability will be different than the rain-rain transition's, and both will be different than the sun-sun transition. So, if it is raining right now, in order to use our model to figure out the likelihood that it will still be raining in an hour, we need to map out the various sequences of transitions and their probabilities.

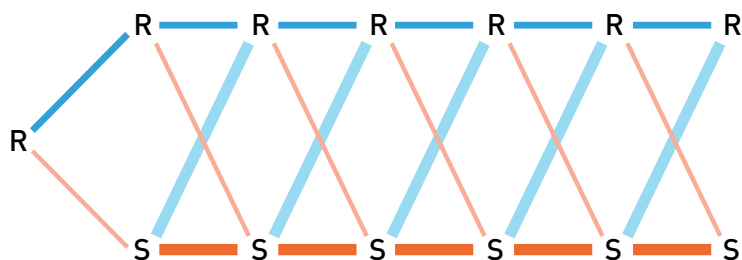
For example, let's say that the rain-rain transition has probability $\frac{2}{3}$. This leaves the rain-sun transition with a probability of $\frac{1}{3}$. Suppose the sun-sun transition has a probability of $\frac{4}{5}$, which makes the probability of the sun-rain transition $\frac{1}{5}$. We can organize these probabilities into a matrix to help us think through this exercise in weather forecasting.

SECTION 7.7

OTHER TYPES OF PROBABILITY CONTINUED

CURRENT STATE	NEXT STATE	
	R	S
	$\frac{2}{3}$	$\frac{1}{3}$
	R	S
S	$\frac{1}{5}$	$\frac{4}{5}$

We can also construct a branching diagram to show the possible ways that this model can develop over six steps:



R = Rain S = Sun
 Orange lines represent sun-sun transition
 Blue lines represent rain-rain transition
 Light orange lines represent rain-sun transition
 Light blue lines represent sun-rain transition

To find the probability that we end up either with rain or sun after six steps, we need an efficient way to consider all of the ways and probabilities that, after the sixth step, the weather will be sunny. For instance, after two steps the possible ways for it to be sunny, assuming we begin with rain, are: rain-rain-sun or rain-sun-sun. Each of these combinations has two transitions, and each transition has an associated probability. We can multiply the probabilities of each transition to find the overall probability of events developing according to each specific sequence. Because both sequences end up with the same result, we can add the probabilities of each sequence happening to obtain the overall probability of ending with sun after two steps.

SECTION 7.7

OTHER TYPES OF
PROBABILITY
CONTINUED

Multiplying and adding is okay for two steps, but for greater numbers of steps, this process can be quite unwieldy because as we consider more steps, we have to consider more specific sequences. Fortunately, we can find the probability of either rain or sun at any step by multiplying the entire matrix of probabilities by itself for however many steps we wish to consider. This is the same as raising the probability matrix to a power. While the details of why this works would be distractingly beyond the level of this discussion, it suffices to say that multiplying two matrices together accounts for all of the various ways in which we can go from a particular initial state to a particular final state in two steps.

Therefore, to find the probability that it will be sunny after six steps (i.e, in one hour), we take our original probability matrix and raise it to the sixth power, which gives us:

CURRENT STATE	STATE AFTER 6 STEPS		
		R	S
	R	0.38	0.62
	S	0.37	0.63

From this probability matrix, we can see that if it is currently raining, there is a 38% chance that it will be raining in an hour and a 62% chance that it will be sunny. This prediction of course is only as valid as the assumptions that went into our model. Often, these assumptions are quite reasonable and powerful. Because of this, Markov Chains form the heart of solving problems ranging from how we can have a computer recognize human speech to how we can identify a region of the human genome responsible for a genetic disease.

In this section we were introduced to two of the many ways in which probability is used in a modern context. We have also seen the important connection between probability and modeling. Our next section will bring us right to the forefront of both probability and mathematical modeling.

SECTION 7.8

MODERN PROBABILITY

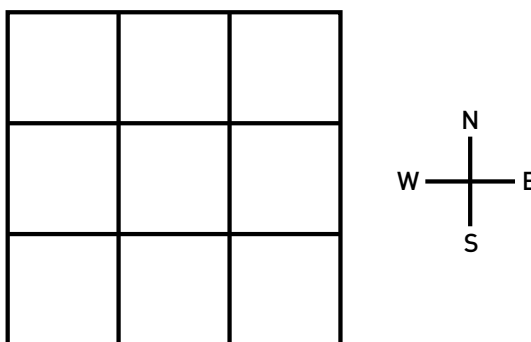
- The BML Traffic Model

THE BML TRAFFIC MODEL

- Although many of the principles of probability were established in centuries past, it is still a vibrant field of mathematical study.
- The BML Traffic Model represents the frontier of probabilistic understanding of complex systems.
- The BML Traffic Model applies not only to traffic congestion, but also to physical or chemical processes such as phase changes.

Most of the ideas that we have discussed so far in this unit were first developed in the 17th and 18th centuries. As in all of mathematics, there has been continuous further development of these ideas since then. (In fact, we've concentrated here on discrete probability and have not really said much regarding continuous probability, the situation in which there is a continuous possible range of outcomes, as with the height of an individual). An exciting case in point is in the modeling of theoretical traffic flows.

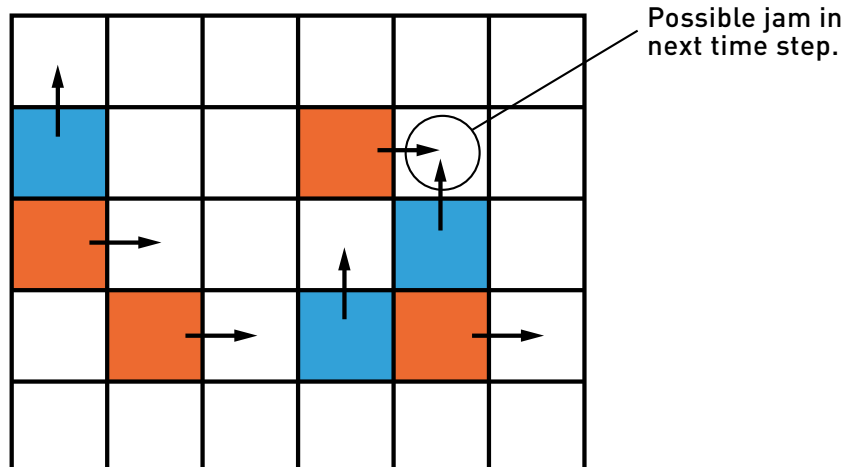
The Biham-Middleton-Levine (BML) Traffic Model, first proposed in 1992, provides a useful model to study how probability affects traffic flow and phase transitions, such as the transformation of liquid water into ice. To get an idea of how this model works, let's imagine an ideal grid of city blocks.



To make things easier, let's assume that the grid extends to infinity in all directions. That way we don't have to worry about any kind of boundary conditions or effects. Let's fill our grid with commuter cars, red ones trying to go east and blue ones trying to go north.

SECTION 7.8

MODERN PROBABILITY CONTINUED



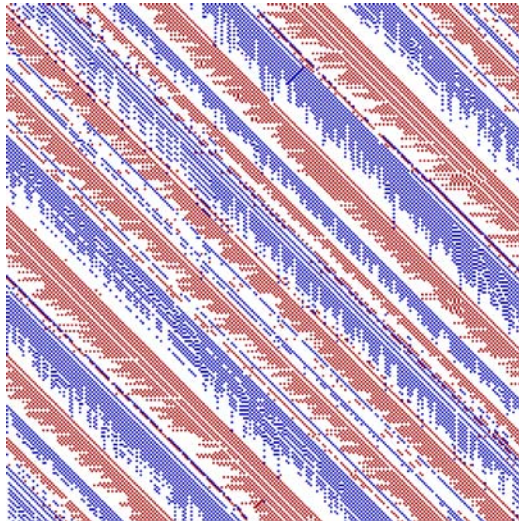
To simplify things further, assume that cars move only one space at a time and are allowed to move only as follows: Every odd second, red cars get to move if the space immediately east of them is vacant, whereas blue cars get to move every even second only if the space immediately north of them is vacant. This process goes on indefinitely.

To determine the starting configuration of cars, we can select a probability, p , that assigns whether or not a space is occupied by a car. We will be interested in the different behaviors that are associated with different values of p . If a particular space ends up being populated with a car, the car's color, and therefore its directional orientation, is determined by a method equivalent to flipping a coin.

SECTION 7.8

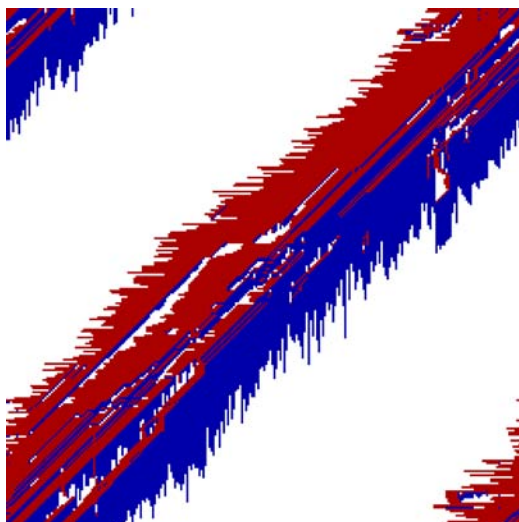
MODERN PROBABILITY CONTINUED

After the grid is populated, the simulation runs. After a period of time, patterns and structure begin to emerge.



Item 1733 / Raissa D'Souza, FREE FLOW $R \leftarrow R-C$ (1/2005). Courtesy of Raissa D'Souza.

Some initial probabilities lead to continuous flow. Cars can move freely forever. In this picture, both red and blue cars are able to move throughout the grid. If the cars were water molecules, these results would correspond to the liquid state.



Item 1734 / Raissa D'Souza, FULLY JAMMED $R \rightarrow R-C$ (1/2005). Courtesy of Raissa D'Souza.

SECTION 7.8

MODERN PROBABILITY CONTINUED

Other initial probabilities lead to traffic gridlock. Movement becomes impossible. Notice how the red and blue cars are stalemated along the center diagonal of the grid. In the water analogy, this would be ice. Note that the parts in the corners are due to the boundary conditions of the grid, so that if a car leaves the left part of the screen, it returns on the right side and vice versa. The same boundary conditions apply to the top and the bottom of the grid. Recalling a concept from our previous unit on topology, this is a flat torus.

The BML model is perplexing because, while at low initial densities traffic flows freely forever and at high initial densities traffic jams up rather quickly, the density at which this transition occurs is not known. Also, there are intermediate states of free flow mixed with periodic jams, depending on the initial population density. As of this writing, there is no detailed mathematical explanation for these behaviors, making this an area for continued exploration.

Rigorous attempts to address the issues involved in the BML Traffic Model and similar models play a huge role in modern probability. Mathematicians are truly just beginning to find ways of dealing with models that correspond to our physical world in meaningful ways. These sorts of results correspond to some of the deepest and most beautiful work in modern mathematics.

SECTION 7.2

HISTORY

- The mathematical study of probability probably was delayed for centuries because of mysticism.
- Renaissance mathematicians took the first strides toward understanding chance in an abstract way.
- Pascal's and Fermat's solutions to the "Problem of the Points" provided an early glimpse of how to use mathematics to say definite things about unknown future events.
- Tree diagrams are useful for keeping track of possible outcomes.

SECTION 7.3

SIMPLE PROBABILITY AND COUNTING

- Simple probability is the ratio of favorable outcomes to the total number of possible outcomes.
- The Galton Board is a model of a sequence of random events.
- Each marble that passes through the system represents a trial consisting of as many random events as there are rows in the system.
- For Galton boards with many rows, the task of enumerating paths is greatly facilitated by using Pascal's Triangle.

SECTION 7.4

LAW OF LARGE NUMBERS

- Bernoulli's Law of Large Numbers shifted the thinking about probability from determining short-term payoffs to predicting long-term behavior.
- Expected value is the average result of a probabilistic event computed by taking into account possible results and their respective probabilities.
- Expected value is a key concept in both gambling and insurance.

SECTION 7.5

THE GALTON BOARD REVISITED

- Each bin of the Galton board has an associated probability, and looking at all of the bins simultaneously gives a distribution.
- Because each peg represents a right-or-left, or binary, decision/option, the distribution of probabilities is called a binomial distribution.
- If we assign a probability other than $\frac{1}{2}$ to each peg of the Galton board, the mean of the distribution will shift.
- The normal distribution is an ideal distribution that is determined by only its mean and standard deviation.

SECTION 7.6

CENTRAL LIMIT THEOREM

- According to the Central Limit Theorem, the distribution of averages of many trials is always normal, even if the distribution of each trial is not.
- Exit poll results can be compared with a normal distribution to make predictions about the results of an election based on a relatively small sample of voters.

SECTION 7.7

OTHER TYPES OF PROBABILITY

- Conditional probability applies to events that are not independent of one another.
- Markov Chains provide a way to talk about sequences of events in which the probability of each event is dependent on the results of prior events.

SECTION 7.8

MODERN PROBABILITY

- Although many of the principles of probability were established in centuries past, it is still a vibrant field of mathematical study.
- The BML Traffic Model represents the frontier of probabilistic understanding of complex systems.
- The BML Traffic Model applies not only to traffic congestion, but also to physical or chemical processes such as phase changes.

BIBLIOGRAPHY

WEBSITES

<http://www.dartmouth.edu/~chance/>

PRINT

Barth, Mike. "Industry Loss and Expense Ratio Comparisons Between HMOs and Life/Health Insurers," NAIC Research Quarterly, vol. 3, no. 1 (January 1997).

Berlinghoff, William P and Fernando Q. Gouvea. Math Through the Ages: A Gentle History for Teachers and Others. Farmington, ME: Oxton House Publishers, 2002.

Berlinghoff, William P. and Kerry E. Grant. A Mathematics Sampler: Topics for the Liberal Arts, 3rd ed. New York: Ardsley House Publishers, Inc., 1992.

Bernstein, Peter L. Against the Gods: The Remarkable Story of Risk. New York: John Wiley and Sons, 1996.

Bogart, Kenneth, Clifford Stein, and Robert L. Drysdale. Discrete Mathematics for Computer Science (Mathematics Across the Curriculum). Emeryville, CA: Key College Press, 2006.

Burton, David M. History of Mathematics: An introduction, 4th ed. New York: WCB/McGraw-Hill, 1999.

Casti, John L. Five More Golden Rules: Knots, Codes, Chaos, and Other Great Theories of 20th-Century Mathematics. New York: John Wiley and Sons, Inc., 2000.

David, F.N. Games, Gods and Gambling; The Origins and History of Probability and Statistical Ideas from the Earliest Times to the Newtonian Era. New York: Hafner Publishing Company, 1962.

De Fermat, Pierre. "Fermat and Pascal on Probability" from Oeuvres de Fermat, Tannery and Henry, eds. University of York. <http://www.york.ac.uk/depts/maths/histstat/pascal.pdf> (accessed October 26, 2005).

Desrosieres, Alain. The Politics of Large Numbers (translated by Camille Naish). Cambridge, MA: Harvard University Press, 1998.

BIBLIOGRAPHY

D'Souza, R. M. "Coexisting phases and lattice dependence of a cellular automata model for traffic flow," *Physical Review E*, vol. 71 (2005).

Epstein, Richard A. *The Theory of Gambling and Statistical Logic*. New York: Academic Press, 1977.

Gigerenzer, Gerd et al. *The Empire of Chance: How Probability Changed Science and Everyday Life*. New York: Cambridge University Press, 1989.

Ghahramani, Saeed. *Fundamentals of Probability*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, Inc., 2000.

Gordon, Hugh. *Discrete Probability*. New York: Springer, 1997.

Grinstead, Charles M. and J. Laurie Snell. *Introduction to Probability*: 2nd rev. ed. Providence, RI: American Mathematical Society, 1997.

Griffiths, T.L. and Tenebaum, J.B. "Probability, Algorithmic Complexity, and Subjective Randomness," *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, (2003).

Gross, Benedict and Joe Harris. *The Magic of Numbers*. Upper Saddle River, NJ: Pearson Education, Inc/ Prentice Hall, 2004.

Kelly, J.L. Jr. "A New Interpretation of Information Rate," *Bell Systems Technical Journal*, vol. 35 (September 1956).

Larsen, Richard J and Morris L. Marx. *An Introduction to Probability and its Applications*. Englewood Cliffs, NJ: Prentice Hall, Inc., 1985.

Martin, Bruce. "Fate... or Blind Chance? What Seems Like Eerie Predestination Is Merely Coincidence," *The Washington Post*, September 9, 1998, final edition.

Relf, Simon and Dennis Almeida. "Exploring the 'Birthdays Problem' and Some of its Variants Through Computer Simulation," *International Journal of Mathematical Education in Science & Technology*, vol. 30, no. 1 (January/ February 1999).

BIBLIOGRAPHY

Smith, David Eugene. A Source Book in Mathematics. New York: McGraw-Hill Book Company, Inc., 1929.

Tannenbaum, Peter. Excursions in Modern Mathematics, 5th ed. Upper Saddle River, NJ: Pearson Education, Inc., 2004.

LECTURES

D'Souza, R.M. "The Science of Complex Networks." CSE Seminar at University of California - Davis, February 2006. HYPERLINK "<http://mae.ucdavis.edu/dsouza/talks.html>" <http://mae.ucdavis.edu/dsouza/talks.html> (accessed 2007).