CONTENTS

CHAPTER 1

Introduction

I suppose it is tempting, if the only tool you have is a hammer,
to treat everything as if it were a nail.
— ABRAHAM MASLOW, *The Psychology of Science* (1966)

Probability is a vast subject. There's a wealth of applications, from the purest parts of mathematics to the sometimes seedy world of professional gambling. It's impossible for any one book to cover all of these topics. That isn't the goal of any book, neither this one nor one you're using for a class. Usually textbooks are written to introduce the general theory, some of the techniques, and describe some of the many applications and further reading. They often have a lot of advanced chapters at the end to help the instructor fashion a course related to their interests.

This book is designed to both supplement any standard introductory text and serve as a primary text by explaining the subject through numerous worked out problems as well as discussions on the general theory. We'll analyze a few fantastic problems and extract from them some general techniques, perspectives, and methods. The goal is to get you past the point where you can write down a model and solve it yourself, to where you can figure out what questions are worth asking to start the ball rolling on research of your own.

First, similar to Adrian Banner's *The Calculus Lifesaver* [Ba], the material is motivated through a rich collection of worked out exercises. It's best to read the problem and spend some time trying them first *before* reading the solutions, but complete solutions are included in the text. Unlike many books, we don't leave the proofs or examples to the reader without providing details; I urge you to try to do the problems first, but if you have trouble the details are there.

Second, it shouldn't come as a surprise to you that there are a lot more proofs in a probability class than in a calculus class. Often students find this to be a theoretically challenging course; a major goal of this book is to help you through the transition. The entire first appendix is devoted to proof techniques, and is a great way to refresh, practice, and expand your proof expertise. Also, you'll find fairly complete proofs for most of the results you would typically see in a course in this book. If you (or your class) are not that concerned with proofs, you can skip many of the arguments, but you should still scan it at the very least. While proofs are often very hard, it's not nearly as bad following a proof as it is coming up with one. Further, just reading a proof is often enough to get a good sense of what the theorem is saying, or how to use it. My goal is not to give you the shortest proof of a result; it's deliberately wordy below to have a conversation with you on how to think about problems and how to go about proving results. Further, before proving a result we'll often spend a lot of time looking at special cases to build intuition; this is an incredibly valuable skill and will help you in many classes to come. Finally, we frequently discuss how to write and execute code to check our calculations or to get a sense of the answer. If you are going to be competitive in the twenty-first-century workforce, you need to be able to program and simulate. It's enormously useful to be able to write a simple program to simulate one million examples of a problem, and frequently the results will alert you to missing factors or other errors.

In this introductory chapter we describe three entertaining problems from various parts of probability. In addition to being fun, these examples are a wonderful springboard which we can use to introduce many of the key concepts of probability. For the rest of this chapter, we'll assume you're familiar with a lot of the basic notions of probability. Don't worry; we'll define everything in great detail later. The point is to chat a bit about some fun problems and get a sense of the subject. We won't worry about defining everything precisely; your everyday experiences are more than enough background. I just want to give you a general flavor for the subject, show you some nice math, and motivate spending the next few months of your life reading and working intently in your course and also with this book. There's plenty of time in the later chapters to dot every "i" and cross each "t".

So, without further fanfare, let's dive in and look at the first problem!

1.1 Birthday Problem

One of my favorite probability exercises is the **Birthday Problem**, which is a great way for professors of large classes to supplement their income by betting with students. We'll discuss several formulations of the problem below. There's a good reason for spending so much time trying to state the problem. In the real world, you often have to figure out what the problem is; you want to be the person guiding the work, not just a technician doing the algebra. By discussing (at great lengths!) the subtleties, you'll see how easy it is to accidentally assume something. Further, it's possible for different people to arrive at different answers without making a mistake, simply because they interpreted a question differently. It's thus very important to always be clear about what you are doing, and why. We'll thus spend a lot of time stating and refining the question, and then we'll solve the problem in order to highlight many of the key concepts in probability. Our first solution is correct, but it's computationally painful. We'll thus conclude with a short description of how we can very easily approximate the answer *if* we know a little calculus.

1.1.1 Stating the Problem

Birthday Problem (first formulation): How many people do we need to have in a room before there's at least a 50% chance that two share a birthday?

This seems like a perfectly fine problem. You should be picturing in your mind lots of different events with different numbers of people, ranging from say the end of the

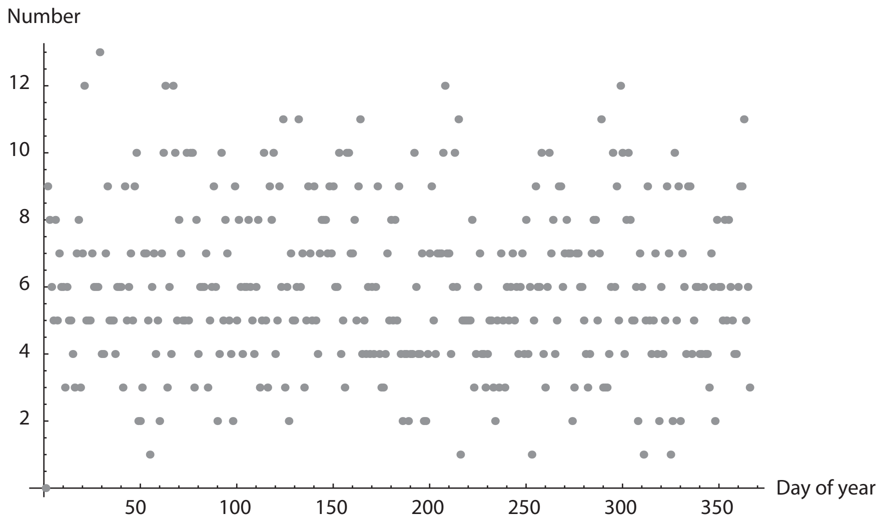

Figure 1.1. Distribution of birthdays of undergraduates at Williams College in Fall 2013.

year banquet for the chess team to a high school prom to a political fundraising dinner to a Thanksgiving celebration. For each event, we see how many people there are and see if there are two people who share a birthday. If we gather enough data, we should get a sense of how many people are needed.

While this may seem fine, it turns out there's a lot of hidden assumptions above. One of the goals of this book is to emphasize the importance of stating problems clearly and fully. This is very different from a calculus or linear algebra class. In those courses it's pretty straightforward: find this derivative, integrate that function, solve this system of equations. As worded above, this question isn't specific enough. I'm married to an identical twin. Thus, at gatherings for her side of the family, there are always two people with the same birthday!* To correct for this trivial solution, we want to talk about a *generic* group of people. We need some information about how the birthdays of our people are distributed among the days of the year. More specifically, we'll assume that birthdays are **independent**, which means that knowledge of one person's birthday gives no information about another person's birthday. Independence is one of the most central concepts in probability, and as a result, we'll explore it in great detail in Chapter 4.

This leads us to our second formulation.

Birthday Problem (second formulation): Assume each day of the year is as likely to be someone's birthday as any other day. How many people do we need to have in a room before there's at least a 50% chance that two share a birthday?

Although this formulation is better, the problem is *still* too vague for us to study. In order to attack the problem we still need more information on the distribution of

*This isn't the only familial issue. Often siblings are almost exactly n years apart, for reasons ranging from life situation to fertile periods. My children (Cam and Kayla) were both born in March, two years apart. Their oldest first cousins (Eli and Matthew) are both September, also two years apart. Think about the people in your family. Do you expect the days of birthdays to be uncorrelated in your family?

birthdays throughout the year. You should be a bit puzzled right now, for haven't we completely specified how birthdays are distributed? We've just said each day is equally likely to be someone's birthday. So, assuming no one is ever born on February 29, that means roughly 1 out of 365 people are born on January 1, another 1 out 365 on January 2, and so on. What more information could be needed?

It's subtle, but we are *still* assuming something. What's the error? We're assuming that we have a random group of people at our event! Maybe the nature of the event causes some days to be more likely for birthdays than others. This seems absurd. After all, surely being born on certain days of the year has nothing to do with being good enough to be on the chess team or football team. Right?

Wrong! Consider the example raised by Malcolm Gladwell in his popular book, *Outliers* [Gl]. In the first chapter, the author investigates the claim that date of birth is strongly linked to success in some sports. In Canadian youth hockey leagues, for instance, "the eligibility cutoff for age-class hockey programs is January 1st." From a young age, the best players are given special attention. But think about it: at the ages of six, seven, and eight, the best players (for the most part) are also the oldest. So, the players who just make the cutoff—those born in January and February—can compete against younger players in the same age division, distinguish themselves, and then enter into a self-fulfilling cycle of advantages. They get better training, stronger competition, even more state-of-the-art equipment. Consequently, these older players get better at a faster rate, leading to more and more success down the road.

On page 23, Gladwell substitutes the birthdays for the players' names: "It no longer sounds like the championship of Canadian junior hockey. It now sounds like a strange sporting ritual for teenage boys born under the astrological signs Capricorn, Aquarius, and Pisces. *March 11 starts around one side of the Tigers' net, leaving the puck for his teammate January 4, who passes it to January 22, who flips it back to March 12, who shoots point-blank at the Tigers' goalie, April 27. April 27 blocks the shot, but it's rebounded by Vancouver's March 6. He shoots! Medicine Hat defensemen February 9 and February 14 dive to block the puck while January 10 looks on helplessly. March 6 scores!*" So, if we attend a party for professional hockey players from Canada, we shouldn't assume that everyone is equally likely to be born on any day of the year.

To simplify our analysis, let's assume that everyone actually *is* equally likely to be born on any day of the year, even though we understand that this might not always be a valid assumption; there's a nice article by Hurley [Hu] that studies what happens when all birthdays are not equally likely. We'll also assume that there are only 365 days in the year. (Unfortunately, if you were born on February 29, you won't be invited to the party.) In other words, we're assuming that the distribution of birthdays follows a **uniform distribution**. We'll discuss uniform distributions in particular and distributions more generally in Chapter 13. Thus, we reach our final version of the problem.

Birthday Problem (third formulation): Assuming that the birthdays *of our guests* are independent and equally likely to fall on any day of the year (except February 29), how many people do we need to have in the room before there's at least a 50% chance that two share a birthday?

1.1.2 Solving the Problem

We now have a well-defined problem; how should we approach it? Frequently, it's useful to look at extreme cases and try to get a sense of what the solution should be. The worst-case scenario for us is when everyone has a different birthday. Since we're assuming there are only 365 days in the year, we *must* have at least two people sharing a birthday once there are 366 people at the party (remember we're assuming no one was born on February 29). This is **Dirichlet's famous Pigeon-Hole Principle**, which we describe in Appendix A.11. On the other end of the spectrum, it's clear that if only one person attends the party, there can't be a shared birthday. Therefore, the answer lies somewhere between 2 and 365. But where? Thinking more deeply about the problem, we see that there should be *at least* a 50% chance when there are 184 people. The intuition is that if no one in the first 183 people shares a birthday with anyone else, then there's at least a 50% chance that they will share a birthday with someone in the room when the 184th person enters the party. More than half of the days of the year are taken! It's often helpful to spend a few minutes thinking about problems like this to get a feel for the answer. In just a few short steps, we've narrowed our set of solutions considerably. We know that the answer is somewhere between 2 and 184. This is still a pretty sizable range, but we think the answer should be a lot closer to 2 than to 184 (just imagine what happens when we have 170 people).

Let's compute the answer by brute force. This gives us our first recipe for finding probabilities. Let's say there are n people at our party, and each is as likely to have one day as their birthday as another. We can look at all possible lists of birthday assignments for n people and see how often at least two share a birthday. Unfortunately, this is a computational nightmare for large n. Let's try some small cases and build a feel for the problem.

With just two people, there are $365^2 = 133{,}225$ ways to assign two birthdays across the group of people. Why? There's 365 choices for the first person's birthday and 365 choices for the second person's birthday. Since the two events are independent (one of our previous assumptions), the number of possible combinations is just the product. The pairs range from (January 1, January 1), (January 1, January 2), and so on until we reach (December 31, December 31).

Of these 133,225 pairs, only 365 have two people sharing a birthday. To see this, note that once we've chosen the first person's birthday, there's only one possible choice for the second person's birthday if there's to be a match. Thus, with two people, the probability that there's a shared birthday is $365/365^2$ or about .27%. We computed this probability by looking at the number of successes (two people in our group of two sharing a birthday) divided by the number of possibilities (the number of possible pairs of birthdays).

If there are three people, there are $365^3 = 48{,}627{,}125$ ways to assign the birthdays. There are $365 \cdot 1 \cdot 364 = 132{,}860$ ways that the first two people share a birthday and the third has a different birthday (the first can have any birthday, the second must have the same birthday as the first, and then the final person must have a different birthday). Similarly, there are 132,860 ways that just the first and third share a birthday, and another 132,860 ways for only the second and third to share a birthday. We must be very careful, however, and ensure that we consider *all* the cases. A final possibility is that all three people could share a birthday. There are 365 ways that that could happen. Thus, the probability that at least two of three share a birthday is 398,945 / 48,627,125, or about .82%. Here 398,945 is $132{,}860 + 132{,}860 + 132{,}860 + 365$, the number of triples with at least two people sharing a birthday. One last note about the $n = 3$ case. It's always a good idea to check and see if an answer is reasonable. Do we expect there to be a greater chance of at least two people in a group of two sharing a birthday, or a group of two in a group of three? Clearly, the more people we have, the greater the chance of a shared birthday. Thus, our probability must be rising as we add more people, and we confirm that .82% is larger than .27%.

It's worth mentioning that we had to be *very careful* in our arguments above, as we didn't want to **double count** a triple. Double counting is one of the cardinal sins in probability, one which most of us have done a few times. For example, if all three people share a birthday this should only count as *one* success, not as three. Why might we mistakenly count it three times? Well, if the triple were (March 5, March 5, March 5) we could view it as the first two share a birthday, or the last two, or the first and last. We'll discuss double counting a lot when we do combinatorics and probability in Chapter 3.

For now, we'll leave it at the following (hopefully obvious) bit of advice: don't discriminate! Count each event once and only once! Of course, sometimes it's not clear what's being counted. One of my favorite scenes in *Superman II* is when Lex Luthor is at the White House, trying to ingratiate himself with the evil Kryptonians: General Zod, Ursa, and the slow-witted Non. He's trying to convince them that they can attack and destroy Superman. The dialogue below was taken from http://scifiscripts.com/scripts/superman_II_shoot.txt.

> General Zod: He has powers as we do.
> Lex Luthor: Certainly. But **-** Er. Oh Magnificent one, he's just one, but you are three (Non grunts disapprovingly), or four even, if you count him twice.

Here Non thought he wasn't being counted, that the "three" referred to General Zod, Ursa, and Lex Luthor. Be careful! Know what you're counting, and count carefully!

Okay. We shouldn't be surprised that the probability of a shared birthday increases as we increase the number of people, and we have to be careful in how we count. At this point, we could continue to attack this problem by brute force, computing how many ways at least two of four (and so on...) share a birthday. If you try doing four, you'll see we need a better way. Why? Here are the various possibilities we'd need to study. Not only could all four, exactly three of four, or exactly two of four share a birthday, but we could even have two pairs of distinct, shared birthdays (say the four birthdays are March 5, March 25, March 25, and March 5). This last case is a nice complement to our earlier concern. Before we worried about double counting an event; now we need to worry about forgetting to count an event! So, not only must we avoid double counting, we must be **exhaustive**, covering all possible cases.

Alright, the brute force approach isn't an efficient—or pleasant!—way to proceed. We need something better. In probability, it is often easier to calculate the **probability of the complementary event**—the probability that A doesn't happen—rather than determining the probability an event A happens. If we know that A doesn't happen with probability p, then A happens with probability $1 - p$. This is due to the fundamental relation that *something* must happen: A and not A are mutually exclusive events—either A happens or it doesn't. So, the sum of the probabilities must equal 1. These are intuitive notions on probabilities (probabilities are non-negative and sum to 1), which we'll deliberate when we formally define things in Chapter 2.

How does this help us? Let's calculate the probability that in a group of n people *no one* shares a birthday with anyone else. We imagine the people walking into the room one at a time. The first person can have any of the 365 days as her birthday since there's no one else in the room. Therefore, the probability that there are no shared birthdays when there's just one person in the room is 1. We'll rewrite this as $365/365$; we'll see in a moment why it's good to write it like this. When the second person enters, someone is already there. In order for the second person not to share a birthday, his birthday must fall on one of the 364 remaining days. Thus, the probability that we don't have a shared birthday is just $\frac{365}{365} \cdot \frac{364}{365}$. Here, we're using the fact that probabilities of independent events are multiplicative. This means that if A happens with probability p and B happens with probability q, then if A and B are independent—which means that knowledge of A happening gives us no information about whether or not B happens, and vice versa—the probability that both A and B happen is $p \cdot q$.

Similarly, when the third person enters, if we want to have no shared birthday we find that her birthday can be any of $365 - 2 = 363$ days. Thus, the probability that she doesn't share a birthday with either of the previous two people is $\frac{363}{365}$, and hence the probability of no shared birthday among three people is just $\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365}$. As a consistency check, this means the probability that there's a shared birthday among three people is $1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} = \frac{365^3 - 365 \cdot 364 \cdot 363}{365^3}$, which is 398,945 / 48,627,125. This agrees with what we found before.

> Note the relative simplicity of this calculation. By calculating the **complementary probability** (i.e., the probability that our desired event doesn't happen) we have eliminated the need to worry about double counting or leaving out ways in which an event can happen.

Arguing along these lines, we find that the probability of no shared birthday among n people is just

$$\frac{365}{365} \cdot \frac{364}{365} \cdots \frac{365 - (n-1)}{365}.$$

The tricky part in expressions like this is figuring out how far down to go. The first person has a numerator of 365, or $365 - 0$, the second has $364 = 365 - 1$. We see a pattern, and thus the n^{th} person will have a numerator of $365 - (n - 1)$ (as we subtract one less than the person's number). We may rewrite this using the **product notation**:

$$\prod_{k=0}^{n-1} \frac{365 - k}{365}.$$

This is a generalization of the **summation notation**; just as $\sum_{k=0}^{m} a_k$ is shorthand for $a_0 + a_1 + \cdots + a_{m-1} + a_m$, we use $\prod_{k=0}^{m} a_k$ as a compact way of writing $a_0 \cdot a_1 \cdots a_{m-1} a_m$. You might remember in calculus that empty sums are defined to be zero; it turns out that the "right" convention to take is to set an empty product to be 1.

If we introduce or recall another bit of notation, we can write our expression in a very nice way. The **factorial** of a positive integer is the product of all positive integers up to it. We denote the factorial by an exclamation point, so if m is a positive integer then $m! = m \cdot (m-1) \cdot (m-2) \cdots 3 \cdot 2 \cdot 1$. So $3! = 3 \cdot 2 \cdot 1 = 6$, $5! = 120$, and it turns out to be *very* useful to set $0! = 1$ (which is consistent with our convention that an empty product is 1). Using factorials, we find that the probability that no one in our group of n

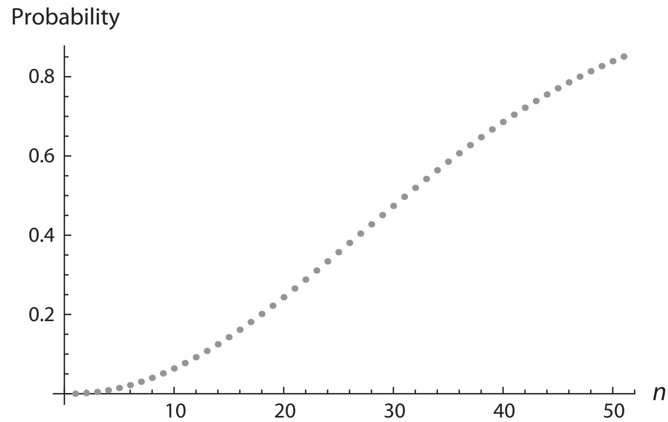

Figure 1.2. Probability that at least two of n people share a birthday (365 days in a year, all days equally likely to be a birthday, each birthday independent of the others).

shares a birthday is just

$$\prod_{k=0}^{n-1} \frac{365-k}{365} = \frac{365 \cdot 364 \cdots (365-(n-1))}{365^n}$$

$$= \frac{365 \cdot 364 \cdots (365-(n-1))}{365^n} \frac{(365-n)!}{(365-n)!} = \frac{365!}{365^n \cdot (365-n)!}. \qquad (1.1)$$

It's worth explaining why we multiplied by $(365-n)!/(365-n)!$. This is a very important technique in mathematics, **multiplying by one**. Clearly, if we multiply an expression by 1 we don't change its value; the reason this is often beneficial is it gives us an opportunity to regroup the algebra and highlight different relations. We'll see throughout the book advantages from **rewriting algebra** in different ways; sometimes these highlight different aspects of the problem, sometimes they simplify the computations. In this case, multiplying by 1 allows us to rewrite the numerator very simply as 365!.

To solve our problem, we must find the smallest value of n such that the product is less than 1/2, as this is the probability that no two persons out of n people share a birthday. Consequently, if that probability is less than 1/2, it means that there'll be at least a 50% chance that two people do in fact share a birthday (remember: complementary probabilities!). Unfortunately, this isn't an easy calculation to do. We have to multiply additional terms until the product first drops below 1/2. This isn't terribly enlightening, and it doesn't generalize. For example, what would happen if we moved to Mars, where the year is almost twice as long—what would the answer be then?

We could use trial and error to evaluate the formula on the right-hand side of (1.1) for various values of n. The difficulty with this is that if we are using a calculator or Microsoft Excel, 365! or 365^n will overflow the memory (though more advanced programs such as Mathematica and Matlab can handle numbers this large and larger). So, we seem forced to evaluate the product term-by-term. We do this and plot the results in Figure 1.2.

Doing the multiplication or looking at the plot, we see the answer to our question is 23. In particular, when there are 23 people in the room, there's approximately a 50.7% chance that at least two people share a birthday. The probability rises to about 70.6% when there are 30 people, about 89.1% when there are 40 people, and a whopping 97% when there are 50 people. Often, in large lecture courses, the professor will bet someone in the class $5 that at least two people share a birthday. The analysis above shows that the professor is very safe when there are 40 or more people (at least safe from losing the bet; their college or university may frown on betting with students).

As one of our objectives in this book is to highlight coding, we give a simple Mathematica program that generated Figure 1.2.

```
(* Mathematica code to compute birthday probabilities *)
(* initialize list of probabilities of sharing and not *)
(* as using recursion need to store previous value *)
noshare = {{1, 1}}; (* at start 100% chance don't share a bday *)
share = {{1, 0}};  (* at start 0% chance share a bday *)
currentnoshare = 1; (* current probability don't share *)
For[n = 2, n <= 50, n++, (* will calculate first 50 *)
 {
  newfactor = (365 - (n-1))/365; (*next term in product*)
  (* update probability don't share *)
  currentnoshare = currentnoshare * newfactor;
  noshare = AppendTo[noshare, {n, 1.0 currentnoshare}];
  (* update probability share *)
  share = AppendTo[share, {n, 1.0 - currentnoshare}];
  }];
(* print probability share *)
Print[ListPlot[share, AxesLabel -> {"n", "Probability"}]]
```

1.1.3 Generalizing the Problem and Solution: Efficiencies

Though we've solved the original Birthday Problem, our answer is somewhat unsatisfying from a computational point of view. If we change the number of days in the year, we have to redo the calculation. So while we know the answer on Earth, we don't immediately know what the answer would be on Mars, where there are about 687 days in a year. Interestingly, the answer is just 31 people!

While it's unlikely that we'll ever find ourselves at a party at Marsport with native Martians, this generalization is very important. We can interpret it as asking the following: given that there are D events which are equally likely to occur, how long do we have to wait before we have a 50% chance of seeing some event twice? Here are two possible applications. Imagine we have cereal boxes and each is equally likely to contain one of n different toys. How many toys do we expect to get before we have our first repeat? For another, imagine something is knocking out connections (maybe it's acid rain eating away at a building, or lightening frying cables), and it takes two hits to completely destroy something. If at each moment all places are equally likely to be struck, this problem becomes finding out how long we have until a systems failure.

This is a common theme in modern mathematics: it's not enough to have an algorithm to compute a quantity. We want more. We want the algorithm to be *efficient* and easy to use, and preferably, we want a nice closed form answer so that we can see how the solution varies as we change the parameters. Our solution above fails miserably in this regard.

The rest of this section assumes some familiarity and comfort with calculus; we need some basic facts about the Taylor series of $\log x$, and we need the formula for the sum of the first m positive integers (which we can and do quickly derive). *Remember that* $\log x$ *means the logarithm of* x *base* e*; mathematicians don't use* $\ln x$ *as the derivatives of* $\log x$ *and* e^x *are "nice," while the derivatives of* $\log_b x$ *and* b^x *are "messy" (forcing us to remember where to put the natural logarithm of* b*).* If you haven't seen calculus, just skim the arguments below to get a flavor of how that subject can be useful. If you haven't seen Taylor series, we can get a similar approximation for $\log x$ by using the tangent line approximation.

We're going to show how some simple algebra yields the following remarkable formula: If everyone at our party is equally likely to be born on any of D days, then we need about $\sqrt{D \cdot 2 \log 2}$ people to have a 50% probability that two people share a birthday.

Here are the needed calculus facts.

- The Taylor series expansion of $\log(1-x)$ is $-\sum_{\ell=1}^{\infty} x^{\ell}/\ell$ when $|x| < 1$. For x small, $\log(1-x) \approx -x$ plus a very small error since x^2 is much smaller than x. Alternatively, the tangent line to the curve $y = f(x)$ at $x = a$ is $y - f(a) = f'(a)(x-a)$; this is because we want a line going through the point $(a, f(a))$ with slope $f'(a)$ (remember the interpretation of the derivative of f at a is the slope of the tangent to the curve at $x = a$). Thus, if x is close to a, then $f(a) + f'(a)(x-a)$ should be a good approximation to $f(x)$. For us, $f(x) = \log(1-x)$ and $a = 0$. Thus $f(0) = \log 1 = 0$, $f'(x) = \frac{-1}{1-x}$ which implies $f'(0) = -1$, and therefore the tangent line is $y = 0 - 1 \cdot x$, or, in other words, $\log(1-x)$ is approximately $-x$ when x is small. *We'll encounter this expansion later in the book when we turn to the proof of the Central Limit Theorem.*
- $\sum_{\ell=0}^{m} \ell = m(m+1)/2$. This formula is typically proved by induction (see Appendix A.2.1), but it's possible to give a neat, direct proof. Write the original sequence, and then underneath it write the original sequence in reverse order. Now add column by column; the first column is $0 + m$, the next $1 + (m-1)$, and so on until the last, which is $m + 0$. Note each pair sums to m and we have $m + 1$ terms. Thus the sum of *twice* our sequence is $m(m+1)$, so our sum is $m(m+1)/2$.

We use these facts to analyze the product on the left-hand side of (1.1). Though we do the computation for 365 days in the year, it's easy to generalize these calculations to an arbitrarily long year—or arbitrarily many events.

We first rewrite $\frac{365-k}{365}$ as $1 - \frac{k}{365}$ and find that p_n—the probability that no two people share a birthday—is

$$p_n = \prod_{k=0}^{n-1} \left(1 - \frac{k}{365}\right),$$

where n is the number of people in our group. A very common technique is to take the logarithm of a product. From now on, whenever you see a product you should have a **Pavlovian response and take its logarithm**. If you've taken calculus, you've seen sums. We have a big theory that converts many sums to integrals and vice versa. You may remember terms such as Riemann sum and Riemann integral. Note that we do not have similar terms for products. We're just trained to look at sums; they should be comfortable and familiar. We don't have as much experience with products, but as we'll see in a moment, the logarithm can be used to convert from products to sums to move us into a familiar landscape. If you've never seen why logarithms are useful, that's about to change. You weren't just taught the log laws because they're on standardized tests; they're actually a great way to attack many problems.

Again, the reason why taking logarithms is so powerful is that we have a considerable amount of experience with sums, but very little experience with products. Since $\log(xy) = \log x + \log y$, we see that taking a logarithm converts our product to a sum:

$$\log p_n = \sum_{k=0}^{n-1} \log\left(1 - \frac{k}{365}\right).$$

We now Taylor expand the logarithm, setting $u = k/365$. Because we expect n to be much smaller than 365, we drop all error terms and find

$$\log p_n \approx \sum_{k=0}^{n-1} -\frac{k}{365}.$$

Using our second fact, we can evaluate this sum and find

$$\log p_n \approx -\frac{(n-1)n}{365 \cdot 2}.$$

As we're looking for the probability to be 50%, we set p_n equal to 1/2 and find

$$\log(1/2) \approx -\frac{(n-1)n}{365 \cdot 2},$$

or

$$(n-1)n \approx 365 \cdot 2 \log 2$$

(as $\log(1/2) = -\log 2$). As $(n-1)n \approx n^2$, we find

$$n \approx \sqrt{365 \cdot 2 \log 2}.$$

This leads to $n \approx 22.49$. And since n has to be an integer, this formula predicts that n should be about 22 or 23, which is exactly what we saw from our exact calculation above. Arguing along the lines above, we would find that if there are D days in the year then the answer would be $\sqrt{D \cdot 2 \log 2}$.

Instead of using $n(n-1) \approx n^2$, in the Birthday Problem we could use the better estimate $n(n-1) \approx (n-1/2)^2$ and show that this leads to the prediction that we need $\frac{1}{2} + \sqrt{365 \cdot 2 \log 2}$. This turns out to be 22.9944, which is stupendously close to 23. It's amazing: with a few simple approximations, we can get pretty close to 23; with just a little more work, we're only .0056 away! We completely avoid having to do big products.

In Figure 1.3, we compare our prediction to the actual answer for years varying in length from 10 days to a million days and note the spectacular agreement—visually, we can't see a difference! It shouldn't be surprising that our predicted answer is so close for long years—the longer the year, the greater n is and hence the smaller the Taylor expansion error.

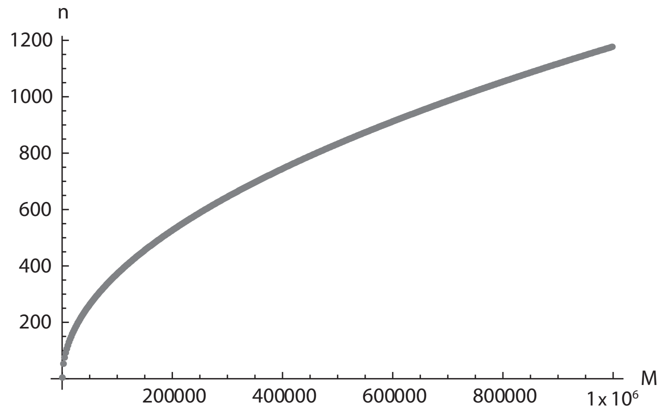

Figure 1.3. The first n so that there is a 50% probability that at least two people share a birthday when there are D days in a year (all days equally likely, all birthdays independent of each other). We plot the actual answer (black dots) versus the prediction $\sqrt{D \cdot 2 \log 2}$ (red line). Note the phenomenal fit: we can't see the difference between our approximation and the true answer.

1.1.4 Numerical Test

After doing a theoretical calculation, it's good to run numerical simulations to check and see if your answer is reasonable. Below is some Mathematica code to calculate the probability that there is at least one shared birthday among n people.

```
birthdaycdf[num_, days_] := Module[{},
  (* num is the number of times we do it *)
  (* days is the number of days in the year *)
  For[d = 1, d <= days, d++, numpeople[d] = 0];
  (* initializes to having d people be where the share happens
  to zero *)

  For[n = 1, n <= num, n++,
  { (* begin n loop *)
   share = 0;
   bdaylist = {}; (* will store bdays of people in room here *)
   k = 0; (* initialize to zero people *)
   While[share == 0,
   {
     (* randomly choose a new birthday *)
    x = RandomInteger[{1, days}];
     (* see if new birthday in the set observed *)
     (* if no add, if yes won and done *)
    If[MemberQ[bdaylist, x] == False,
    bdaylist = AppendTo[bdaylist, x],
    share = 1];
    k = k + 1; (* increase number people by 1 *)
     (* if just shared a birthday add one from that person
    onward *)
    If[share == 1, For[d = k , d <= days, d++,
     numpeople[d] = numpeople[d] + 1];
    ]; (* records when had match *)
```

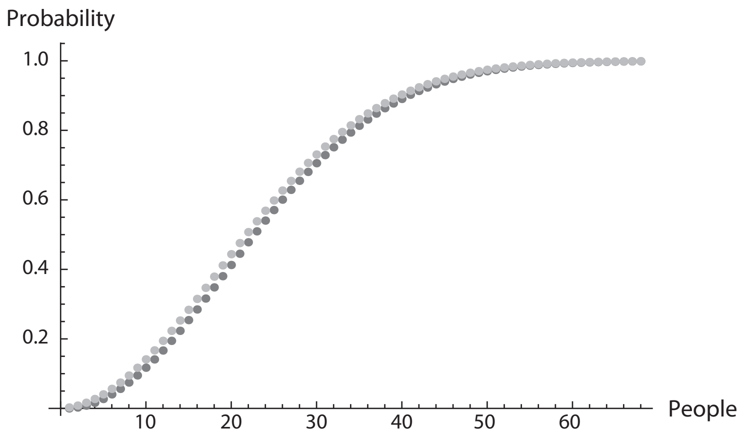

Figure 1.4. Comparison between experiment and theory: 100,000 trials with 365 days in a year.

```
    (* as doing cdf do from that point onward *)
   }]; (* end while loop *)
  }]; (* end n loop *);

bdaylistplot = {};
max = 3 * (.5 + Sqrt[days Log[4]]);
For[d = 1, d <= max, d++,
bdaylistplot =
   AppendTo[bdaylistplot, {d, numpeople[d] 1.0/num}]
  ]; (* end of d loop *)
 (* prints obs prob of shared birthday as a function of people*)
 Print[ListPlot[bdaylistplot, AxesLabel -> {People, Prob}]];
 Print[
  "Observed probability of success with 1/2 + Sqrt[D log(4)] people
  is ", numpeople[Floor[.5 + Sqrt[days Log[4.]]]]*100.0/num, "%."];
 (* this is our theoretical prediction *)
 f[x_] := 1 - Product[1 - k/days, {k, 0, Floor[x]}];
 (* this prints our obseerved data and our predicted at
 the same time using show *)
 Print[
  Show[Plot[f[x], {x, 1, max}],
   ListPlot[bdaylistplot, AxesLabel -> {People, Prob}]
   ]];
 theorybdaylistplot = {};
 For[d = 1, d <= max, d++,
  theorybdaylistplot = AppendTo[theorybdaylistplot, {d, f[d]}]];
 Print[
  ListPlot[{bdaylistplot, theorybdaylistplot},
   AxesLabel -> {People, Prob}]];
 ];
```

The above code looks at `num` groups, with `days` in the year, with various display options. It also computes our observed success rate at $1/2 + \sqrt{D \log 4}$. We record the results of one such simulation in Figure 1.4, where we took 100,000 trials in a 365-day year. Using the estimated point of $1/2 + \sqrt{D \log 4}$ led to a success rate of 47.8%, which is pretty good considering all the approximations we did. Further, a comparison of the cumulative probabilities of success between our experiment and our prediction is quite striking, which is highly suggestive of our not having made a mistake!

Figure 1.5. Larry Bird and Magic Johnson, game two of the 1985 NBA Finals (May 30) at the Boston Garden. Photo from Steve Lipofsky, Basketballphoto.com.

1.2 From Shooting Hoops to the Geometric Series

The purpose of this section is to introduce you to some important results in mathematics in general and probability in particular. While we'll motivate the material by considering a special basketball game, the results can be applied in many fields. It's thus good to have this material on your radar screen as you continue through the book. After discussing some generalizations we'll conclude with another interesting problem. Its solution is a bit involved and there's a nice paper with the solution, so we won't go through all the details here. Instead we'll concentrate on how to *attack* problems like this, which is a very important skill. It's easy to be frustrated upon encountering a difficult problem, and frequently it's unclear how to begin. We'll discuss some general problem solving techniques, which if you master you can then fruitfully apply to great effect again and again.

1.2.1 The Problem and Its Solution

The Great Shootout: Imagine that Larry Bird and Magic Johnson decide that instead of a rough game (see Figure 1.5), they'll just have a one-on-one shoot-out, winner takes all. (When I was growing up, these were two of the biggest superstars. If it would be easier to visualize, you may replace Larry Bird with Paul Pierce and Magic Johnson with Kobe Bryant, and muse on where I grew up and what year this section was written.) The two superstars take turns shooting, always releasing the ball from the same place. Suppose that Bird makes a basket with probability p (and

thus misses with probability $1 - p$), while Magic makes a basket with probability q (and thus misses with probability $1 - q$). If Bird shoots first, what is the probability that he wins the shootout?

Is this problem clear and concise? For the most part it is, but as we saw with the Birthday Problem it's worthwhile to take some time and think carefully about the problem, and make sure we're not making any hidden assumptions. There's one point worth highlighting: this is a mathematics problem, and not a real-world problem. We assume that Bird *always* makes a basket with probability p. He never tires, the crowd never gets to him (positively or negatively), and the same is true for Magic Johnson. Of course, in real life this would be absurd; if nothing else, after a year of doing nothing but shooting we'd expect our players to be tired, and thus shoot less effectively. However, we're in a math class, not a basketball arena, so we won't worry about endowing our players with superhuman stamina, and leave the generalization to "human" players to the reader.

While we chose to phrase this as a Basketball Problem, many games follow this general pattern. A common problem in probability involves finding the distribution of waiting times for the first successful iteration of some process. For example, imagine flipping a coin with probability p of heads and probability $1 - p$ of tails. Two (or more!) people take turns, and the first one to get something wins. There are many ways we can complicate the problem. We could have more people. We could also have the probabilities vary. We'll leave these generalizations for later, and stay with our simple game of hoops, for once we learn how to do this we'll be well-prepared for these other problems.

The standard way to attack this problem is to write down a number of probabilities and then evaluate their sum by using the Geometric Series Formula:

Geometric Series Formula: Let r be a real number less than 1 in absolute value. Then

$$\sum_{n=0}^{\infty} r^n \;=\; 1 + r + r^2 + r^3 + \cdots \;=\; \frac{1}{1-r}.$$

I'll review the proof of this useful formula at the end of this section. After first solving this problem by applying the geometric series formula, we'll discuss another approach that leads to a *proof* of the geometric series formula! We're going to use a powerful technique, which we'll call the **Bring It Over Method**. I'm indebted to Alex Cameron for coining this phrase in a Differential Equations class at Williams College. This strategy is important not just in probability, but also throughout much of mathematics, as we'll see shortly in some examples. It is precisely because this method is so important and useful that we've moved it to the beginning of the book. You should see from the beginning "good" math, which means math that's not only beautiful, but powerful and useful. There's a lot happening below, but you'll be in great shape if you can take the time and digest it.

First, we'll discuss the standard approach to solving this problem. For each positive integer n, we calculate the probability that Bird wins on his n^{th} shot. To get a sense of the answer, let's do some small n first. If $n = 1$, this means Bird wins on his first shot. In other words, he makes his first shot, which happens with probability p. If $n = 2$, then Bird wins on his second shot. In order for Bird to get a second shot, he and Magic must both miss their first shots. Since Bird misses his first shot with probability $1-p$ and Magic misses his first shot with probability $1-q$, we know that the probability that Bird misses his first, Magic misses his first, and Bird then makes his second shot is just $(1-p)(1-q)p = rp$, where we've set $r = (1-p)(1-q)$. Similarly, we see that if $n = 3$ then Bird must miss his first two shots, Magic must miss his first two shots, and Bird must make his third shot. The probability of this happening is $(1-p)(1-q)(1-p)(1-q)p = r^2p$. In general, the probability Bird wins on his n^{th} shot is $r^{n-1}p$. Note the exponent of r is $n-1$, as to win on his n^{th} shot he must miss his first $n-1$ shots and then make his n^{th} shot.

We've thus broken the probability of Bird winning into summing (infinitely many!) simpler probabilities. We haven't counted anything twice, and we've taken care of all the different ways for Bird to win. If Bird wins, then he must make the first basket of the shoot-out at some n. In other words, his probability of winning is

$$\text{Prob(Bird wins)} = p + rp + r^2p + r^3p + \cdots = \sum_{n=0}^{\infty} r^n p = p \sum_{n=0}^{\infty} r^n,$$

where as before, $r = (1-p)(1-q)$. *Using the geometric series formula to evaluate the probability*, we see that

$$\text{Prob(Bird wins)} = \frac{p}{1-r},$$

with $r = (1-p)(1-q)$.

Now we'll derive this probability *without* knowing the geometric series formula. In fact, we can use our probabilistic reasoning to derive an alternative proof of that formula. Let's denote the probability that Bird wins by x. We'll compute x in a different way than before. If Bird makes his first basket (which happens with probability p), then he wins. By definition, this happens with probability p. If Bird misses his first basket (which will happen with probability $1-p$), then the only way he can win is if Magic misses his first shot, which happens with probability $1-q$. But Magic missing isn't enough to ensure that Bird wins, though if Magic doesn't miss then Bird cannot win.

We've now reached a very interesting configuration. Both Bird and Magic have missed their first shots, and Bird is about to shoot his second shot. A little reflection reveals that if x is the probability that Bird wins the game *with Bird getting the first shot*, then x is also the probability that Bird wins after he and Magic miss their first shots. The reason for this is that it doesn't matter how we reach a point in this shootout. As long as Bird is shooting, his probability of winning in our model is the same regardless of how many times he and Magic have missed. This is an example of a **memoryless process**. The only thing that matters is what state we're in, not how we got there.

Amazingly, we can now find x, the probability that Bird wins! Recalling $r = (1-p)(1-q)$, we see that this probability is $p + (1-p)(1-q)x$, or

$$\begin{aligned} x &= p + (1-p)(1-q)x \\ x - rx &= p \\ x &= \frac{p}{1-r}. \end{aligned}$$

We've now computed the probability Bird wins two different ways, the first using the geometric series and the second noting that we have a memoryless process. Our two expressions must be equal, so if we set these answers equal to one another we see that we've also proved the geometric series formula:

$$\text{Since } p\sum_{n=0}^{\infty} r^n \;=\; \frac{p}{1-r} \quad \text{we have} \quad \sum_{n=0}^{\infty} r^n \;=\; \frac{1}{1-r}$$

provided $p \neq 0$! In mathematical arguments you must *always* be careful about dividing by zero; for example, if $p = 0$ then $4p = 9p$, but this doesn't mean $4 = 9$. Of course, if $p = 0$ then Bird has no chance of winning and we shouldn't even be considering this calculation. By choosing appropriate values of p, q (see Exercise 1.5.30) we can prove the geometric series for all r with $0 \leq r < 1$.

This turns out to be one of the most important methods in probability, and in fact is one of the reasons this problem made it into the introduction. Frequently we'll have a very difficult calculation, but if we're clever we'll see it equals something that's easier to find. It's of course very hard to "see" the simpler approach, but it does get easier the more problems you do. We call this the **Proof by Comparison** or **Proof by Story** method, and give some more examples and explanation in Appendix A.6.

Our second approach to finding the probability of Bird winning worked because we have something of the form

$$\text{unknown} = \text{good} + c \cdot \text{unknown},$$

where we just need $c \neq 1$. We must avoid $c = 1$; otherwise, we'd have the unknown on both sides of the equation occurring equally, meaning we wouldn't be able to isolate it. If, however, $c \neq 1$, then we find unknown $=$ good$/(1-c)$.

Example 1.2.1 (Bring It Over for Integrals): *The **Bring It Over Method** might be familiar from calculus, where it's used to evaluate certain integrals. The basic idea is to manipulate the equation to get the unknown integral on both sides and then solve for it from there. For example, consider*

$$I = \int_0^{\pi} e^{cx} \cos x\, dx.$$

We integrate by parts twice. Let $u = e^{cx}$ *and* $dv = \cos x dx$, *so* $du = ce^{cx} dx$ *and* $v = \sin x dx$. *Since* $\int_0^{\pi} u dv = uv|_0^{\pi} - \int_0^{\pi} v du$, *we have*

$$I \;=\; e^{cx} \sin x\Big|_0^{\pi} - \int_0^{\pi} ce^{cx} \sin x dx \;=\; -c\int_0^{\pi} e^{cx} \sin x dx.$$

We integrate by parts a second time. Then, we again take $u = e^{cx}$ *and set* $dv = \sin x$, *so* $du = ce^{cx} dx$ *and* $v = -\cos x$. *Thus,*

$$\begin{aligned} I \;&=\; -c\int_0^{\pi} e^{cx} \sin x dx \\ &=\; -c\left[e^{cx}(-\cos x)\Big|_0^{\pi} - \int_0^{\pi} ce^{cx}(-\cos x)\, dx\right] \end{aligned}$$

$$= -c\left[e^{\pi c} + 1 + c\int_0^{\pi} e^{cx}\cos x dx\right]$$
$$= -ce^{\pi c} - c - c^2\int_0^{\pi} e^{cx}\cos x dx \;=\; -ce^{\pi x} - c - c^2 I,$$

because the last integral is just what we're calling I. Rearranging yields

$$I + c^2 I = -ce^{\pi c} - c, \tag{1.2}$$

or

$$I \;=\; \int_0^{\pi} e^{cx}\cos x dx \;=\; -\frac{ce^{\pi c} + c}{c^2 + 1}.$$

This is a truly powerful method—we're able to evaluate the integral not by computing it directly, but by showing it equals something known minus a multiple of itself.

Remark 1.2.2: *Whenever we have a complicated expression such as* (1.2)*, it's worth checking the special cases of the parameter. This is a great way to see if we've made a mistake. Is it surprising, for example, that the final answer is negative for $c > 0$? Well, the cosine function is positive for $x \leq \pi/2$ and negative from $\pi/2$ to π, and the function e^{cx} is growing. Thus, the larger values of the exponential are hit with a negative term, and the resulting expression should be negative. (To be honest, I originally dropped a minus sign when writing this problem, and I noticed the error by doing this very test!) Another good check is to set $c = 0$. In this case we have $\int_0^{\pi}\cos x dx$, which is just 0. This is what we get in* (1.2) *upon setting $c = 0$.*

Remark 1.2.3 (Proof of the geometric series formula): *For completeness, let's do the standard proof of the geometric series formula. Consider $S_n = 1 + r + r^2 + \cdots + r^n$. Note $rS_n = r + r^2 + r^3 + \cdots + r^{n+1}$; thus $S_n - rS_N = 1 - r^{n+1}$, or*

$$S_n \;=\; \frac{1 - r^{n+1}}{1 - r}.$$

If $|r| < 1$, we can let $n \to \infty$, and find that

$$\lim_{n\to\infty} S_n \;=\; \sum_{n=0}^{\infty} r^n \;=\; \frac{1}{1 - r}.$$

The reason we multiplied through by r above is that it allowed us to have almost the same terms in our two expressions, and thus when we did the subtraction almost everything canceled. With practice, it becomes easier to see what algebra to do to lead to great simplifications, but this is one of the hardest parts of the subject.

Remark 1.2.4: *Technically, the probability proof we gave for the geometric series isn't quite as good as the standard proof. The reason is that for us, $r = (1 - p)(1 - q)$, which forces us to take $r \geq 0$. On the other hand, the standard proof allows us to take any r of absolute value at most 1. With some additional work, we can generalize our* *argument to handle negative r as well. Let $r = -s$ with $s \geq 0$. Then*

$$\sum_{n=0}^{\infty}(-s)^n = \sum_{n=0}^{\infty} s^{2n} - \sum_{n=0}^{\infty} s^{2n+1} = (1-s)\sum_{n=0}^{\infty} s^{2n}.$$

We now apply the geometric series formula to the sum of $s^{2n} = (s^2)^n$ and find that

$$\sum_{n=0}^{\infty}(-s)^n = (1-s)\cdot\frac{1}{1-s^2} = \frac{1-s}{(1-s)(1+s)} = \frac{1}{1+s} = \frac{1}{1-(-s)},$$

just as we claimed above. It may seem like all we've done is some clever algebra, but a lot of mathematics is learning how to ***rewrite algebra*** *to remove the clutter and see what's really going on. This example teaches us that we can often prove our result for a simpler case, and then with a little work get the more general case as well.*

Remark 1.2.5: *As the math you do becomes more and more involved, you'll appreciate the power of* ***good notation***. *Typically in probability we use q to denote $1 - p$, the complementary probability. In this problem, however, we use p and the next letter in the alphabet, q, for the two probabilities we care about most: the chance Bird has of making a basket, and the chance Magic has. We could use p_B for Bird's probability of getting a basket and p_M for Magic's; while the notation is now a bit more involved it has the advantage of being more descriptive: when we glance down, it's clear what item it describes. Along these lines, instead of writing x for the probability Bird wins we could write x_B. For this simple problem it wasn't worth it, but going forward this is something to consider.*

1.2.2 Related Problems

The techniques we developed for the Basketball Problem can be applied in many other cases; we give two nice examples below. The first is a great introduction to **generating functions**, which we explore in great detail in Chapter 19.

Example: Another fun example of the Bring It Over Method is the following problem: let F_n denote the n^{th} Fibonacci number. Compute $\sum_{n=0}^{\infty} F_n/3^n$.

Recall that the **Fibonacci numbers** are defined by the recurrence relation $F_{n+2} = F_{n+1} + F_n$, with initial conditions $F_0 = 0$ and $F_1 = 1$. Once the first two terms in the sequence are specified, the rest of the terms are uniquely determined by the recurrence relation. We'll see recurrence relations again when we study betting strategies in roulette in §23.

We now apply our method to solve this problem. Let $x = \sum_{n=0}^{\infty} F_n/3^n$. In the argument below we'll re-index the summation in order to use the Fibonacci recurrence; it shouldn't be surprising that we use this relation, as it is *the* defining property of the Fibonacci numbers. We have

$$\begin{aligned} x &= \sum_{n=0}^{\infty} \frac{F_n}{3^n} \\ &= \frac{F_0}{1} + \frac{F_1}{3} + \sum_{n=2}^{\infty} \frac{F_n}{3^n} \end{aligned}$$

$$= \frac{0}{1} + \frac{1}{3} + \sum_{m=0}^{\infty} \frac{F_{m+2}}{3^{m+2}}$$

$$= \frac{1}{3} + \sum_{m=0}^{\infty} \frac{F_{m+1} + F_m}{3^{m+2}}$$

$$= \frac{1}{3} + \sum_{m=0}^{\infty} \frac{F_{m+1}}{3^{m+1} \cdot 3} + \sum_{m=0}^{\infty} \frac{F_m}{3^m \cdot 9}$$

$$= \frac{1}{3} + \frac{1}{3} \sum_{n=1}^{\infty} \frac{F_n}{3^n} + \frac{1}{9} \sum_{n=0}^{\infty} \frac{F_n}{3^n}.$$

As $F_0 = 0$, we may extend the first sum in the last line over all n and find

$$x = \frac{1}{3} + \frac{x}{3} + \frac{x}{9},$$

which implies that $x = 3/5$.

It's annoying, but frequently in problems like the above you have to change the index of summation, moving it a bit. If you continue and take a course on differential equations, you'll do this non-stop when you reach the sections on series solutions. For another example along these lines, see the proof of the Binomial Theorem in Appendix A.2.3.

Example: We'll give one more example. Alice, Bob, and Charlie (whom you'll meet again if you take a cryptography course) are playing a game of cards. The first one to draw a diamond wins. They take turns drawing—Alice then Bob then Charlie then Alice and so on—until someone draws a diamond. After each person draws, if the card isn't a diamond it's put back in the deck and the deck is then thoroughly shuffled before the next person picks. What is the probability that each person wins?

WARNING: I hope the argument below seems plausible. I thought so at first, but it led to the wrong answer! After outlining it, we'll analyze what went wrong. As you read it below, see if you can find the mistake.

Let x denote the probability that Alice wins, y the probability that Bob wins, and z the probability that Charlie wins. Because there are 52 cards in a deck and 13 of these cards are diamonds, whomever is picking always has a $13/52 = 1/4$ chance of winning. The probability Alice wins is just

$$x = \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} x,$$

or $x = \frac{1}{4} + \frac{27}{64}x$, which implies that $\frac{37}{64}x = \frac{1}{4}$ or $x = \frac{16}{37}$. Why is this the answer? Either Alice wins on her first pick, which happens with probability 1/4, or to win she, Bob, and Charlie all miss on their first pick, which happens with probability $(3/4)^3$. At this point, it's as if we just started the game. You should see the similarity to the Basketball Problem now.

Similarly, we find the probability that Bob wins is

$$y = \frac{3}{4} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} y.$$

That is to say, either Bob wins on his first pick or they all miss once, Alice misses, and then Bob gets to pick again. After cleaning up the algebra, we get $y = \frac{48}{175}$. If we argue analogously for Charlie, we find that $z = \frac{9}{37}$.

As always, it's extremely valuable to check our answer. We must have $x + y + z = 1$, since exactly one of them must win. While we could have computed z directly from our knowledge of x and y, we prefer this method because it gives us an opportunity to talk about testing answers. Whenever possible, you should try to find an answer two different ways as a check against algebra (or other more serious) errors. In our case, we have

$$x + y + z = \frac{16}{37} + \frac{48}{175} + \frac{9}{37} = \frac{6151}{6475} \neq 1.$$

So, what went wrong? These probabilities should sum to 1, but they don't; we're off by a little bit. The problem is that we didn't compute the probabilities correctly. We defined y to be the probability that Bob wins when *Alice* draws first. Thus, the equation for y isn't $y = \frac{3}{4} \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^4 y$, but instead

$$y = \frac{3}{4} \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^3 y.$$

Remember, y is the probability that Bob wins when *Alice* picks first. So, when we start the game over, it must be Alice picking, not Bob. More explicitly, let's look at the two terms above. The $\frac{3}{4} \cdot \frac{1}{4}$ comes from Alice picking and not getting a diamond, followed by Bob immediately picking a diamond. Since y is the probability Bob wins when Alice is picking, we need to get back to Alice picking. Thus, in the second term the factor $\left(\frac{3}{4}\right)^3$ represents Alice, then Bob, and finally Charlie picking non-diamonds. At this point, it is again Alice's turn to take a card, and thus from *here* the probability Bob wins is y.

Thus $y = \frac{3}{4} \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^3 y$, as claimed. We can easily solve this for y, and find $y = \frac{12}{37}$. A similar argument gives $z = \frac{9}{37}$. Note that $x + y + z = \frac{16}{37} + \frac{12}{37} + \frac{9}{37} = 1$.

Alternatively, once we know x, we can immediately determine y by noting that $y = \frac{3}{4}x$. The intuition is simple: if we're calculating the probability that Bob wins, Alice must obviously not win on her first pick. After Alice fails on her first pick, it's Bob's turn. From this point forward, however, the probability that Bob wins is identical to the probability that Alice wins when Alice picks first, namely x. Therefore, $y = \frac{3}{4}x = \frac{12}{37}$. Similarly, we find that $z = \frac{3}{4} \cdot \frac{3}{4}x$, or $z = \frac{9}{37}$. It takes awhile to become comfortable looking at problems this way, but it is worth the effort. If you can correctly identify the memoryless components, you can frequently bypass infinite sums; it is far better to have a finite number of things on your "to-do" (or perhaps I should say "to-sum") list than infinitely many items!

We end this section with an appeal to you to learn how to write simple computer code. It is an incredibly useful, valuable skill to be able to numerically explore these problems as well as check your math. Let's revisit our incorrect logic, and let's write a simple program to see if our answer is reasonable. I often program in Mathematica because (1) it is freely available to me, (2) it has a lot of functions predefined that I like, (3) it's a fairly friendly environment with good display options, and (4) it's what I used when I was in college.

```
diamonddraw[num_] := Module[{},
   awin = 0; bwin = 0; cwin = 0; (* initialize win counts to 0 *)
   For[n = 1, n <= num, n++,
    { (* start of n loop *)
     diamond = 0;
     While[diamond == 0,
      { (* start of diamond loop, keep doing till get diamond *)
       (* randomly choose a card for each of three players,
       with replacement*)
       (* we'll order the deck so first 13 cards are the diamonds *)
       c1 = RandomInteger[{1, 52}];
       c2 = RandomInteger[{1, 52}];
       c3 = RandomInteger[{1, 52}];
       (* if one is a diamond we win and will stop *)
       If[c1 <= 13 || c2 <= 13 || c3 <= 13, diamond = 1];
       (* give credit to winner *)
       If[diamond == 1,
        If[c1 <= 13, awin = awin + 1,
         If[c2 <= 13, bwin = bwin + 1,
          If[c3 <= 13, cwin = cwin + 1]]]
        ]; (* end of if loop on diamond = 1 *)
       }]; (* end of while diamond loop *)
     }]; (* end of n loop *)
   Print["Here are the observed probabilities from ", num, " games."];
   Print["Percent Alice won (approx): ",  100.0 awin / num, "%."];
   Print["Percent Bob won (approx): ",  100.0 bwin / num, "%."];
   Print["Percent Charlie won (approx): ",  100.0 cwin / num, "%."];
   Print["Predictions (from our bad logic) were approx ", 1600.0/37,
    " ", 4800.0/175, " ", 900.0/37];
   ];
```

Playing one million games yielded:

- Percent Alice won (approx): 43.2202%.
- Percent Bob won (approx): 32.4069%.
- Percent Charlie won (approx): 24.3729%.
- Predictions (from our bad logic) were approx 43.2432%, 27.4286%, 24.3243%.

Thus while we're fairly confident about the probability for Alice, something looks fishy with our answer for Bob. One would hope with a million runs we would be close to the true answer; we'll return to figuring out how close we should be after we learn the Central Limit Theorem.

Remark 1.2.6: *Remember how we said that $y = \frac{3}{4}x$ and $z = (\frac{3}{4})^2 x$? We can use this to solve for x. Since someone wins, the sum of the probabilities is 1:*

$$1 = x + y + z = x + \frac{3}{4}x + \frac{9}{16}x = \frac{37}{16}x,$$

and thus $x = 16/37$! The reason we're able to so easily find x here is that there is a great deal of ***symmetry****; all players have the same chance of winning when they pick. This would be true in the Basketball Problem only if $p = q$.*

1.2.3 General Problem Solving Tips

We end this section by discussing another Basketball Problem. I heard about this from a beautiful article by Yigal Gerchak and Mordechai Henig, "The basketball shootout: strategy and winning probabilities" (see [GH]). Our goal is not to go through all the mathematics to solve the problem; if you want the solution you can go to their paper. Instead, our purpose is to explain good ways to attack problems like this. The ability to analyze something new is a very valuable skill, but a hard one to master. The more problems you do, the more experience you gain and the more connections you can make. You'll start to see that a new problem has some features in common with something you've done before, which can give you a clue on how to start your analysis. Of course, the more problems you master, the better chance you have of seeing connections. Our goal below is to highlight some good strategies for investigating new problems outside your comfort zone. Here's the problem.

Problem: N people are in a basketball shootout. Each gets one shot, and they're told if they're shooting first, second, third, and so on. Whomever makes a basket from the furthest distance wins. If you are the k^{th} person shooting, you know the outcome of the first $k - 1$ shots, and you know how many people will shoot after you. Where should you shoot from?

As with so many problems in this chapter, our first step is to make sure we understand the problem. We'll make several assumptions to simplify the problem. If after reading this section and their paper you're up for a challenge, try removing some of these assumptions and figuring out the new solutions.

- Let's assume all basketball players shoot from somewhere on the line connecting the two baskets. You might think this is an automatic assumption, since the players are shooting without any defenders pressuring them and thus all shots only depend on the distance. There is a flaw in that argument, however; the ball could bounce off the backboard, and thus perhaps the *angle* of the shot matters. If that's the case, we might need detailed information about how well people make different shots depending on both the distance and angle to the basket. Thus, let's make our lives simpler and assume everyone shoots from the same line.
- Next, we'll assume all players have the same ability. Of course this isn't true, but remember the great advice: *Walk before you run!* Always try to do simpler cases first. If we can't do the case when all players are the same, we have no chance of handling the general case.
- The description is vague as to what happens if two people make a basket from the same distance. We could say whoever made the shot first wins, in which case the other person would never shoot from the same place; however, they might shoot 10^{-10} centimeters further. To avoid such small ridiculous motions, let's just say if two people make a basket from the distance, whomever shot last gets the win. This avoids having to do a limiting argument, and really won't fundamentally change the solution.

- The probability of making a basket cannot increase as you move further away from the basket. While this should seem reasonable, it's important to realize we're making this assumption. Consider the following: extend your right hand to the sky. Try to touch your right shoulder with the thumb on your right hand. Now try to touch your right elbow with the same thumb. As the elbow is closer to the thumb when the arm is extended, it might seem reasonable to suppose it will be easier to reach, but this is clearly not the case.
- Related to the above, we'll assume the players can move so close to the basket that they can make a shot 100% of the time. This is a very useful assumption to include. Why? If the first $N-1$ players miss, the last player automatically wins by moving really close to the basket. If this couldn't happen, then it would be possible for there to be no winners in the game.

Okay, it's now time to try to solve the problem. As our players are all identical, instead of measuring their distance to the basket in feet or meters, we can record where they shoot by the *probability* they make a shot from there. Thus if we're close to the basket our p should be close to 1, and it should be non-increasing as we move further back.

Before we can solve the problem, however, it's worthwhile to spend some time and think about notation. We need to encode the given information and our analysis in math equations. Notation is very important. We need a symbol to denote the probability of person 1 winning given that there are N people playing and that they shot at *p and all subsequent people shoot from their optimum locations!* Let's denote this by $x_{1;N}(p)$. Why is this **good notation**? We often use x to represent unknown quantities. It should be a function of how far away we shoot, and thus writing it as a function of p is reasonable. What about the subscripts? The first subscript refers to person one, while the second tells us how many people there are. As the two numbers play different roles, we separate them by a colon. It's not as clear what the notation should be for the second person as where they shoot depends on where the first person shoots. We'll return to this later.

Armed with our notation, we now turn to determining $x_{1;N}(p)$. Whenever you have a hard problem, a great way to start is to look at simpler cases and try to detect a pattern. If there's just one player it's clear what happens: they win! They just shoot from where they have a 100% chance of making it, and thus $x_{1;1}(1) = 1$. Note that we would never have them shoot from anywhere else if they're the only shooter.

What about two players? If you think about it, everything is determined by where the first person shoots. If they miss then the second player automatically wins, as we've said they can move close enough to the basket to be assured of making their shot. If however the first player makes a basket, then the second player shoots from the same spot (as we've declared that if two people make a basket from the same place, then whoever shot second wins).

Before we convert the above analysis to mathematical notation, let's try and get a feel for the solution. This is a very valuable step. If you have a rough sense of what the answer should be, you're much more likely to catch an algebra error. The first question to ask is: do we think the first player has a better than 50% chance or worse than a 50% chance of winning? Another way of putting this is: would you rather shoot first or second? For me, I'd rather shoot second. If the first person misses then I automatically win, while if they make a shot all I have to do is make the same shot they did. Thus, it seems reasonable to expect that $x_{1;2}(p) \leq 1/2$.

Let's assume the first player shoots at position p (remember this means their probability of making the shot is p). There are two possibilities.

1. Person one can make the basket (which happens with probability p), in which case the second person shoots. If this happens then the second person makes a basket with probability p, so in this case person one wins with probability $1 - p$.
2. Person one can miss the basket (which happens with probability $1 - p$), in which case the second person wins with probability 1 and the first person wins with probability 0.

Combining the two cases, we find

$$x_{1;2}(p) \;=\; p \cdot (1-p) + (1-p) \cdot 0 \;=\; p(1-p).$$

We now want to find the value of p that maximizes the above expression; that will tell us where person one should shoot. If you know calculus you can take the derivative, set it equal to zero, and find that $p = 1/2$ gives the maximum. Alternatively, you can plot the function $x_{1;2}(p) = p(1 - p)$. This is a downward parabola with vertex at $p = 1/2$, and thus the maximum probability is 1/4 or 25%. Notice our answer is less than 50%, as expected.

We leave the rest of the analysis to the reader. I strongly encourage you to try the case of three shooters. For some problems the difficulty doesn't increase too much with increasing N, while for others new features emerge. Even figuring out good notation for 3 shooters is hard. For example, where the second person shoots will depend on whether or not the first person makes their shot. This observation does suggest one piece of good news: if the first person misses, the problem reduces to the two shooter case we just studied. Frequently we can make observations like this in our studies; you should always be on the lookout for simplifications, for reductions to earlier and simpler cases.

We end this section by explicitly culling out some useful observations on how to tackle new, hard problems.

General Problem Solving Strategies:

- Clearly define the problem. Be careful about hidden assumptions. Be explicit; if you need to assume something, do so but make note of the fact.
- Choose **good notation**. I've always been bothered by cosecant being the reciprocal of sine—shouldn't cosecant and cosine go together? In calculus we use F to denote the anti-derivative of f; by doing so, we make it easy to glance at the work and get a feel for what's happening.
- Do special cases first to build intuition. Walk before you run. Don't try to do the whole case at once; do some simpler cases first, and try to detect a pattern.

1.3 Gambling

No introduction to probability would be complete without at least a passing discussion of applications to gambling. This is both for historical reasons (a lot of the impetus for the development of the subject came from studying games of chance) and for current applications (consider how many billions of dollars are wagered, lost, and won in everything from football to poker to elections).

1.3.1 The 2008 Super Bowl Wager

I arrived at Williams in the summer of 2008. One of my favorite students relayed the story of a friend of his (let's call him Bob) who, in 2007, placed a \$500 wager with Las Vegas that the Patriots would go undefeated in the regular season and continue on and win the Super Bowl. He received 1000 to 1 odds, so if he wins he walks away with \$500,000, while if he loses he's down \$500.

As a Patriots fan, that season is still a little hard to talk about (though easier after the win over the Seahawks in 2015), but I'll try. The Patriots *did* go undefeated in the regular season, becoming the first team to do so in a 16-game season. They won their two AFC play-off games, and advanced to the Super Bowl and faced the New York Giants. The Patriots beat the Giants in the last game of the regular season, but it was a close game.

In the middle of the third quarter, with the Patriots enjoying a small lead, Vegas calls Bob and offers to buy the bet back at 300 to 1 odds; this means that they'll give him \$150,000 now to limit their exposure. Thus if Bob accepts, then Vegas immediately loses \$150,000 but protects themselves from losing the larger \$500,000; similarly it means Bob gets \$150,000 but loses the opportunity to get \$500,000.

Bob has faith in the Patriots and declines the offer, electing to go for the big payoff. I claim, and hope to convince you, that Bob made a *bad* choice; however, the reason Bob made a bad choice has nothing to do with the phenomenal catch by Giants wide receiver David Tyree on his helmet that kept the Giants' game-winning drive alive on their way to a huge upset win. Bob is living life on the edge: if the Patriots win he wins big, but if they lose he gets nothing. In the next subsection we'll look at a way for Bob to greatly minimize his risk; in fact, with a little bit of applied probability, Bob can ensure that he gets several hundred thousand dollars, *no matter who wins the game*!

1.3.2 Expected Returns

Right now Bob has bet \$500 on the Patriots; he stands to receive \$500,000 if the Patriots win but nothing if they lose. If the Patriots win with probability p, then p% of the time he makes \$500,000 and $(1 - p)$% of the time he makes \$0; also, no matter what, he loses the \$500 he bet.

The problem for Bob is that he's in a very risky position, and depending on the outcome of the game he can have huge fluctuations in his personal fortune. He can protect himself by placing a secondary bet on the Giants. *If* he were to make protective bets at the start of the season he'd be in trouble due to how the payoffs are calculated, but Bob is in a fortunate position (which sadly he didn't realize). We're not at the start of the season—the Patriots *have* made it to the Super Bowl, and we know their opponent. He can now protect himself by betting on just the Giants. As a Patriots fan, I can understand the reluctance to do so; as a mathematician, however, it's the only sensible decision!

Imagine that for every \$1 bet on the Giants you receive \$$x$ if they win, and \$0 if the Giants lose; as the Patriots were favored to win x must exceed 2. Why? Imagine the two teams were equal and each wins half the time. Then if $x = 2$ if we were to bet \$1 then half the time we would get \$2, half the time we would get \$0, and thus on average we expect to get \$1. Note this exactly equals the amount we wagered, so we should be indifferent to betting in this situation. As the Patriots were expected to win, however,

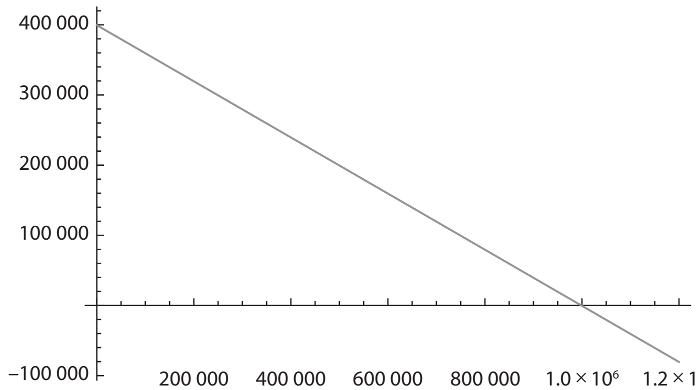

Figure 1.6. Plot of expected returns given an additional $\$B$ bet on the Giants, assuming that the Patriots have an 80% chance of winning and, if the Giants win, each dollar bet on them gives \$3.

Vegas needs to give people an incentive to place money on the Giants. As the odds of the Giants winning was believed to be less than 50%, there had to be a bigger payoff if the Giants won to make the wager more fair, and hence $x > 2$.

For definiteness, let's assume that the probability the Patriots win is $p = .8$, that $x = 3$, and that we now bet $\$B$ on the Giants winning. *Let's also assume that the Super Bowl will continue until one team wins and thus there is no tie; if you don't like this we can always phrase things as the Patriots win or the Patriots don't win, and note that not winning may be different than losing.* How do our returns look? If the Pats win, which happens with probability p, we make \$500,000; if the Giants win (which occurs with probability $1 - p$) we make $\$xB$; in both cases we have wagered $\$500 + \$B$.

Thus our expected return is

$$p \cdot \$500{,}000 + (1 - p)x \cdot \$B - \$500 - \$B;$$

we plot this in Figure 1.6.

Notice that the more we bet on the Giants, the lower our expected return is. This shouldn't be surprising, as we are assuming the Patriots win 80% of the time. In particular, if we bet a huge amount on the Giants we expect to lose a lot (the reason is that $(1 - p)x$ is less than 1).

At first, it appears that betting on the Giants is a bad idea—the more we bet on them, the lower our expected return. In the next subsection, however, we'll continue our analysis and show that this in fact *is* a good idea for most people.

1.3.3 The Value of Hedging

Figure 1.6 is misleading. Yes, the more we bet on the Giants the lower our expected return; however, this is not the right question to ask. Most people are risk averse. Which would you rather have: a guaranteed \$10,000 or a .001% chance of winning a million dollars and a 99.999% chance of getting nothing? Most people would take the sure \$10,000, especially when you calculated the expected return in the second situation: .001% of the time we get a million, while the rest we get nothing; thus we expect to make

$$.00001 \cdot \$1{,}000{,}000 + .99999 \cdot \$0 \ = \ \$100.$$

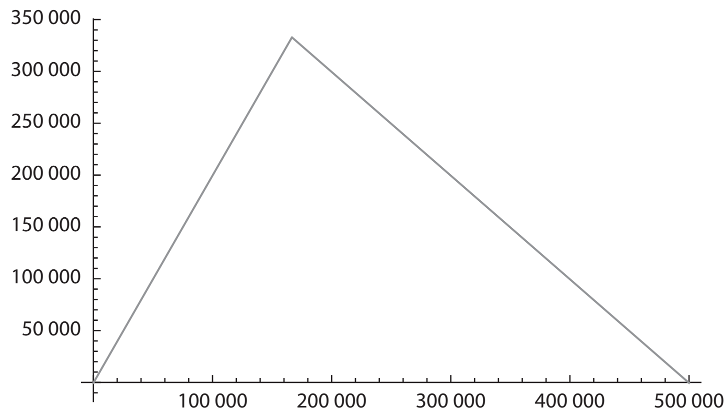

Figure 1.7. Plot of minimum guaranteed returns given an additional $\$B$ bet on the Giants, assuming that the Patriots have an 80% chance of winning and, if the Giants win, each dollar bet on them gives \$3.

While in the second situation when we win we win *big*, the chance is so low that the expected return is worse.

What if instead of a million dollars we now get a billion dollars in the second situation? In that case our expected return increases from \$100 to \$100,000. Now the situation isn't as clear. The expected value is greater in the second case, but most of the time we'll get nothing. Should we take the deal? The answer to that question is beyond the scope of this book, and falls to the realm of economics and psychology. It's worth briefly noting, though, what we are *not* being offered. We are *not* being offered the chance to play this game many times; we can only play once....

While the above problem is hard and involves personal choice, it's the wrong problem. What we'd rather do is have a situation where we can still win big, but no matter what we are still assured of getting something good. In general this is not possible; in the situation of Bob it fortunately is, and involves the beautiful concept of **hedging**. One of the hardest things to learn is to ask the right question. When we looked at the plot of the expected returns from a bet of $\$B$ on the Giants, that was the wrong object to study. What we should be looking at is how much money are we guaranteed to make from a bet of $\$B$ on the Giants.

Though the two questions sound similar, the answers are very different. If we have bets on both the Patriots and the Giants winning, then no matter what at least *one* of our bets must win. If the Patriots win we get \$500,000; if the Giants win we get $\$xB$ (note that regardless of the outcome we lose our initial wager of $\$500 + \$B$). Thus, no matter whether the Patriots win or the Giants win, we will get the minimum of \$500,000 and $\$xB$. We display our minimum guaranteed winnings in Figure 1.7.

This plot is very different than Figure 1.6: our minimum return increases at first as we increase our bet on the Giants, and then decreases! Our minimum return is

$$\min(500000, xB) - 500 - B;$$

assuming $x = 3$ and $p = .8$ we find the critical bet is when $500000 = 3B$, or when B is about \$166,667. At that special wager we're indifferent (from a financial point of view!) to whomever wins, and we are ensured of making approximately \$332,833.

It's worth pausing and letting this sink in. By placing a large bet on the Giants (\$166,667 is not small change for most of us!) we can make sure that we walk away with \$332,833 *no matter who wins in the game*! At this point we're no longer gambling as there is no longer an element of chance!

1.3.4 Consequences

There is a lot more that could be said about this problem, but this is enough to highlight some key points. Most of the time in life you cannot eliminate all risk, but sometimes it is possible! Why was that an option here? The reason was that we had the chance of placing a second bet late in the season (either right before, or even during, the Super Bowl!).

Why did Bob fail to do this? Sadly, Bob never took a course in probability (an advantage you have over him and others). Psychologically, however, Bob was focused on the big payoff, on winning the huge bet. He was so focused on maximizing his return that he completely forgot about minimizing his losses, or, in other words, maximizing his minimum return. It's very easy in life to look at the wrong item (magicians are wonderful at misdirection); one of the goals of this book is to help you learn how to ask the right questions and look at the right quantities. A great example of this is the Method of Least Squares versus the Method of Absolute Values (see Chapter 24); depending on what matters most to you there are different "best" choices to what curve "best" fits the data.

In this betting case, we could use basic probability to calculate our expected return, and we saw that with a large chance of a Patriots win it made no sense, from the point of view of maximizing our expected winnings, to bet on the Giants. For most of us, however, that's the wrong problem. Most of us are risk averse, and we'd rather have a guaranteed \$332,833 than a possible \$500,000 (the expected value is \$400,000, with 80% of the time us winning \$500,000 and 20% of the time was walking away with nothing). It's very interesting who would choose which option for various probabilities; if the Patriots really will win 80% of the time then the expected value is better when we don't bet on the Giants, but it is a lot riskier. For me, it's worth a little smaller expected return to have no risk at all on a good payout.

Interestingly, when we look at the minimum return it's no longer a probability problem. If we change x then the minimum return plot in Figure 1.7 changes; however, the plot does not change if we change the probability p of the Patriots winning! Why? The reason is that we're not looking at our expected return now, we're just looking at the minimum return and thus it doesn't matter who wins as we always assume the outcome is whatever is the worst for us.

When looking at math, be it an equation or a figure, you want to get a *feel* for the behavior. Try playing around with some of the parameters and intuiting the resulting change. For example, we talked about what should happen if we change p; how do you think the shape changes if we increase x?

1.4 Summary

I hope you've enjoyed these problems. The Birthday Problem makes an appearance in almost every first course in probability (a quick Google search turns up hundreds of millions of hits), and for good reason. It's ideally suited to introduce the course. It involves so many of the most important issues, including some obvious ones, such as the notion of independence, when probabilities multiply, the dangers of double counting, and the perils of missing cases, as well as a few less obvious ones, such as the need to state a problem clearly, the advantages of introducing new functions (like the factorial function) to simplify expressions, the power of taking logarithms and using log laws, and ways to approximate the answers to difficult calculations.

The Basketball Shootout is a less clear choice. I almost gave a great problem connecting the Fibonacci numbers and gambling strategies for playing roulette in Las Vegas; don't worry—we'll hit that in Chapter §23 (or go to https://www.youtube.com/watch?v=Esa2TYwDmwA). The point is that a probability instructor has a great deal of freedom in designing a course and choosing examples. It's impossible for this book to perfectly align with any class, nor should it. What we can do is talk in great detail about how to attack a problem, emphasizing the techniques, discussing how to check your answer, and highlighting the dangers and pitfalls. These can be transferred to almost anything you'll see in your class. Further, by choosing a few less standard examples you get to see some things you wouldn't have otherwise. The Basketball Problem quickly introduces us to the concept of a memoryless game, which is crucial in much of game theory (as well as advanced topics in probability, such as Markov processes).

If you've seen calculus before, there's the added advantage of revisiting what seemed like a one-time trick, namely the "Bring It Over" Method where we got our unknown integral on both sides of an equation. A technique is a trick that can be used successfully again and again, and this is a great one. We'll say more about this in a moment.

There are many possible gambling problems to choose; I chose the one above because (1) I'm a Patriots fan (while the 2008 Super Bowl was a painful loss, this section was written shortly after Butler's great interception and the Pats 2015 Super Bowl triumph), (2) it illustrates applications of probability and issues of applying it in the real world, and (3) it provides a terrific opportunity to talk about asking the right question. In many previous classes you have been given the problem to solve, which is frequently a trivial modification of worked out examples you've seen; in the real world often the hardest part is figuring out what the problem is or what the metric for success will be. Are we concerned with maximizing our expected returns, or maximizing our minimum return and eliminating as much risk as possible?

It's now time to explore the subject in earnest. We're forced to order the chapters and topics; while our choice is defensible, be aware that it's not the only one. Your instructor and your book may choose to do things in another order, so if you're using this book to supplement your course text, just be aware that you may be hopping around a bit. To assist you, I've tried to make the chapters as self-contained as possible. This means that if you read this book cover to cover, you'll notice passages suspiciously like earlier ones. This isn't accidental; it's to make the book as easy to use as possible. If you're having trouble in your class on the Gamma distribution, you can jump in at that chapter.

The next chapter is pretty standard for all courses. We'll cover the basic concepts in probability and discuss the definitions. While most courses use calculus, not all do. This isn't a problem here; we can cover the building blocks without calculus. Where is calculus most useful? It's really needed in expanding our domain of discourse; more examples are available with calculus, as well as more methods for finding probabilities (in fact, the Fundamental Theorem of Calculus allows us to interpret probabilities as areas under curves, which can be computed through integration).

There's one significant issue, though, in the next chapter. How rigorously do we want to define everything? This is a very important question, and no answer is right for all. Typically a first course doesn't assume familiarity with real analysis, and things are

(continued...)

INDEX