17 Events and Probability Spaces

17.1 Let's Make a Deal

In the September 9, 1990 issue of *Parade* magazine, columnist Marilyn vos Savant responded to this letter:

> *Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?*
>
> Craig. F. Whitaker
> Columbia, MD

The letter describes a situation like one faced by contestants in the 1970's game show *Let's Make a Deal*, hosted by Monty Hall and Carol Merrill. Marilyn replied that the contestant should indeed switch. She explained that if the car was behind either of the two unpicked doors—which is twice as likely as the the car being behind the picked door—the contestant wins by switching. But she soon received a torrent of letters, many from mathematicians, telling her that she was wrong. The problem became known as the *Monty Hall Problem* and it generated thousands of hours of heated debate.

This incident highlights a fact about probability: the subject uncovers lots of examples where ordinary intuition leads to completely wrong conclusions. So until you've studied probabilities enough to have refined your intuition, a way to avoid errors is to fall back on a rigorous, systematic approach such as the Four Step Method that we will describe shortly. First, let's make sure we really understand the setup for this problem. This is always a good thing to do when you are dealing with probability.

17.1.1 Clarifying the Problem

Craig's original letter to Marilyn vos Savant is a bit vague, so we must make some assumptions in order to have any hope of modeling the game formally. For example, we will assume that:

1. The car is equally likely to be hidden behind each of the three doors.
2. The player is equally likely to pick each of the three doors, regardless of the car's location.
3. After the player picks a door, the host *must* open a different door with a goat behind it and offer the player the choice of staying with the original door or switching.
4. If the host has a choice of which door to open, then he is equally likely to select each of them.

In making these assumptions, we're reading a lot into Craig Whitaker's letter. There are other plausible interpretations that lead to different answers. But let's accept these assumptions for now and address the question, "What is the probability that a player who switches wins the car?"

17.2 The Four Step Method

Every probability problem involves some sort of randomized experiment, process, or game. And each such problem involves two distinct challenges:

1. How do we model the situation mathematically?
2. How do we solve the resulting mathematical problem?

In this section, we introduce a four step approach to questions of the form, "What is the probability that...?" In this approach, we build a probabilistic model step by step, formalizing the original question in terms of that model. Remarkably, this structured approach provides simple solutions to many famously confusing problems. For example, as you'll see, the four step method cuts through the confusion surrounding the Monty Hall problem like a Ginsu knife.

17.2.1 Step 1: Find the Sample Space

Our first objective is to identify all the possible outcomes of the experiment. A typical experiment involves several randomly-determined quantities. For example, the Monty Hall game involves three such quantities:

1. The door concealing the car.
2. The door initially chosen by the player.

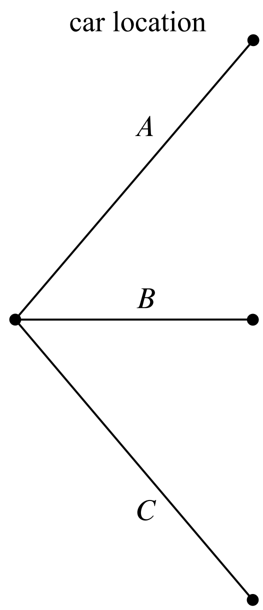

Figure 17.1 The first level in a tree diagram for the Monty Hall Problem. The branches correspond to the door behind which the car is located.

3. The door that the host opens to reveal a goat.

Every possible combination of these randomly-determined quantities is called an *outcome*. The set of all possible outcomes is called the *sample space* for the experiment.

A *tree diagram* is a graphical tool that can help us work through the four step approach when the number of outcomes is not too large or the problem is nicely structured. In particular, we can use a tree diagram to help understand the sample space of an experiment. The first randomly-determined quantity in our experiment is the door concealing the prize. We represent this as a tree with three branches, as shown in Figure 17.1. In this diagram, the doors are called A, B and C instead of 1, 2, and 3, because we'll be adding a lot of other numbers to the picture later.

For each possible location of the prize, the player could initially choose any of the three doors. We represent this in a second layer added to the tree. Then a third layer represents the possibilities of the final step when the host opens a door to reveal a goat, as shown in Figure 17.2.

Notice that the third layer reflects the fact that the host has either one choice or two, depending on the position of the car and the door initially selected by the player. For example, if the prize is behind door A and the player picks door B, then

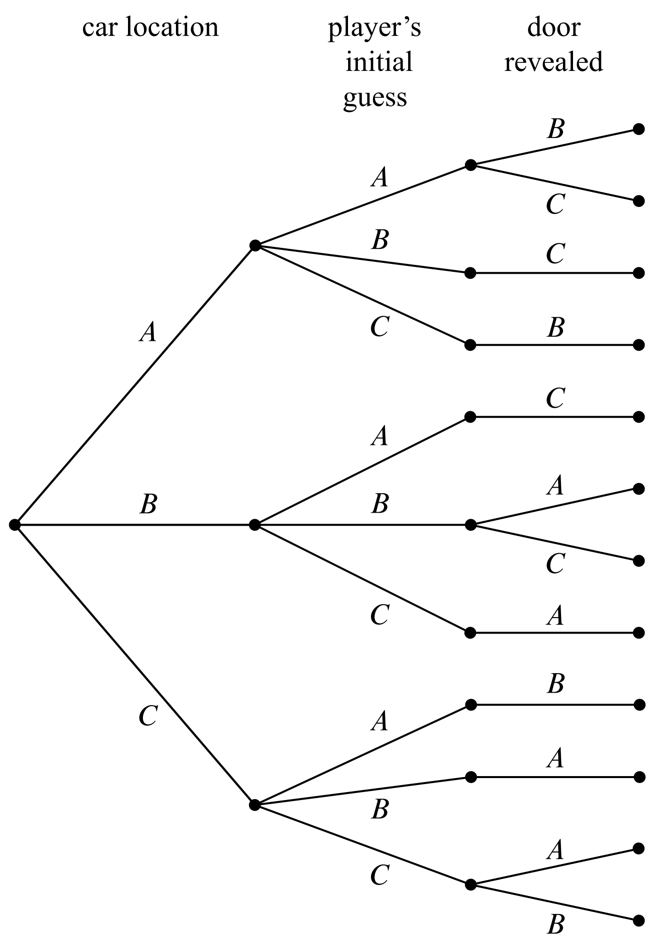

Figure 17.2 The full tree diagram for the Monty Hall Problem. The second level indicates the door initially chosen by the player. The third level indicates the door revealed by Monty Hall.

the host must open door C. However, if the prize is behind door A and the player picks door A, then the host could open either door B or door C.

Now let's relate this picture to the terms we introduced earlier: the leaves of the tree represent *outcomes* of the experiment, and the set of all leaves represents the *sample space*. Thus, for this experiment, the sample space consists of 12 outcomes. For reference, we've labeled each outcome in Figure 17.3 with a triple of doors indicating:

(door concealing prize, door initially chosen, door opened to reveal a goat).

In these terms, the sample space is the set

$$\mathcal{S} = \left\{ \begin{array}{l} (A, A, B),\ (A, A, C),\ (A, B, C),\ (A, C, B),\ (B, A, C),\ (B, B, A), \\ (B, B, C),\ (B, C, A),\ (C, A, B),\ (C, B, A),\ (C, C, A),\ (C, C, B) \end{array} \right\}$$

The tree diagram has a broader interpretation as well: we can regard the whole experiment as following a path from the root to a leaf, where the branch taken at each stage is "randomly" determined. Keep this interpretation in mind; we'll use it again later.

17.2.2 Step 2: Define Events of Interest

Our objective is to answer questions of the form "What is the probability that ... ?", where, for example, the missing phrase might be "the player wins by switching," "the player initially picked the door concealing the prize," or "the prize is behind door C."

A set of outcomes is called an *event*. Each of the preceding phrases characterizes an event. For example, the event [prize is behind door C] refers to the set:

$$\{(C, A, B), (C, B, A), (C, C, A), (C, C, B)\},$$

and the event [prize is behind the door first picked by the player] is:

$$\{(A, A, B), (A, A, C), (B, B, A), (B, B, C), (C, C, A), (C, C, B)\}.$$

Here we're using square brackets around a property of outcomes as a notation for the event whose outcomes are the ones that satisfy the property.

What we're really after is the event [player wins by switching]:

$$\{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\}. \tag{17.1}$$

The outcomes in this event are marked with checks in Figure 17.4.

Notice that exactly half of the outcomes are checked, meaning that the player wins by switching in half of all outcomes. You might be tempted to conclude that a player who switches wins with probability $1/2$. *This is wrong.* The reason is that these outcomes are not all equally likely, as we'll see shortly.

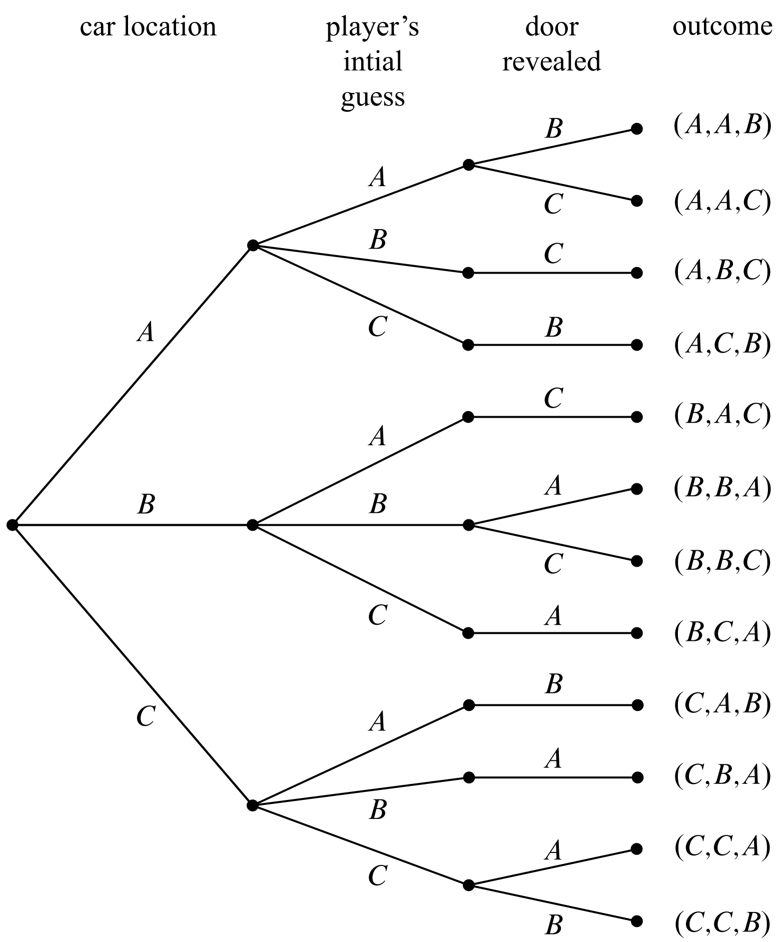

Figure 17.3 The tree diagram for the Monty Hall Problem with the outcomes labeled for each path from root to leaf. For example, outcome (A, A, B) corresponds to the car being behind door A, the player initially choosing door A, and Monty Hall revealing the goat behind door B.

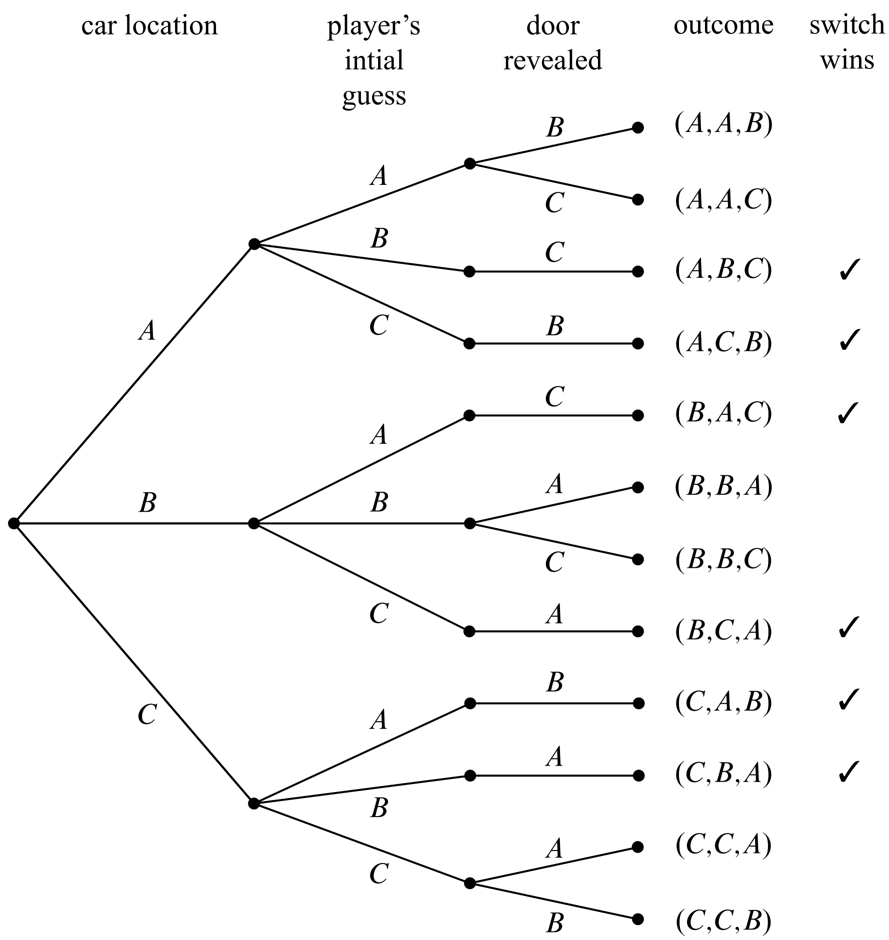

Figure 17.4 The tree diagram for the Monty Hall Problem, where the outcomes where the player wins by switching are denoted with a check mark.

17.2.3 Step 3: Determine Outcome Probabilities

So far we’ve enumerated all the possible outcomes of the experiment. Now we must start assessing the likelihood of those outcomes. In particular, the goal of this step is to assign each outcome a probability, indicating the fraction of the time this outcome is expected to occur. The sum of all the outcome probabilities must equal one, reflecting the fact that there always must be an outcome.

Ultimately, outcome probabilities are determined by the phenomenon we’re modeling and thus are not quantities that we can derive mathematically. However, mathematics can help us compute the probability of every outcome *based on fewer and more elementary modeling decisions.* In particular, we’ll break the task of determining outcome probabilities into two stages.

Step 3a: Assign Edge Probabilities

First, we record a probability on each *edge* of the tree diagram. These edge-probabilities are determined by the assumptions we made at the outset: that the prize is equally likely to be behind each door, that the player is equally likely to pick each door, and that the host is equally likely to reveal each goat, if he has a choice. Notice that when the host has no choice regarding which door to open, the single branch is assigned probability 1. For example, see Figure 17.5.

Step 3b: Compute Outcome Probabilities

Our next job is to convert edge probabilities into outcome probabilities. This is a purely mechanical process:

> calculate the probability of an outcome by multiplying the edge-probabilities on the path from the root to that outcome.

For example, the probability of the topmost outcome in Figure 17.5, (A, A, B), is

$$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18}. \tag{17.2}$$

We’ll examine the official justification for this rule in Section 18.4, but here’s an easy, intuitive justification: as the steps in an experiment progress randomly along a path from the root of the tree to a leaf, the probabilities on the edges indicate how likely the path is to proceed along each branch. For example, a path starting at the root in our example is equally likely to go down each of the three top-level branches.

How likely is such a path to arrive at the topmost outcome (A, A, B)? Well, there is a 1-in-3 chance that a path would follow the A-branch at the top level, a 1-in-3 chance it would continue along the A-branch at the second level, and 1-in-2 chance it would follow the B-branch at the third level. Thus, there is half of a one third of a one third chance, of arriving at the (A, A, B) leaf. That is, the chance is $1/3 \cdot 1/3 \cdot 1/2 = 1/18$—the same product (in reverse order) we arrived at in (17.2).

We have illustrated all of the outcome probabilities in Figure 17.5.

Specifying the probability of each outcome amounts to defining a function that maps each outcome to a probability. This function is usually called $\Pr[\cdot]$. In these terms, we've just determined that:

$$\begin{aligned}
\Pr[(A, A, B)] &= \frac{1}{18}, \\
\Pr[(A, A, C)] &= \frac{1}{18}, \\
\Pr[(A, B, C)] &= \frac{1}{9}, \\
&\text{etc.}
\end{aligned}$$

17.2.4 Step 4: Compute Event Probabilities

We now have a probability for each *outcome*, but we want to determine the probability of an *event*. The probability of an event E is denoted by $\Pr[E]$, and it is the sum of the probabilities of the outcomes in E. For example, the probability of the [switching wins] event (17.1) is

$$\begin{aligned}
&\Pr[\text{switching wins}] \\
&= \Pr[(A, B, C)] + \Pr[(A, C, B)] + \Pr[(B, A, C)] + \\
&\qquad \Pr[(B, C, A)] + \Pr[(C, A, B)] + \Pr[(C, B, A)] \\
&= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} \\
&= \frac{2}{3}.
\end{aligned}$$

It seems Marilyn's answer is correct! A player who switches doors wins the car with probability $2/3$. In contrast, a player who stays with his or her original door wins with probability $1/3$, since staying wins if and only if switching loses.

We're done with the problem! We didn't need any appeals to intuition or ingenious analogies. In fact, no mathematics more difficult than adding and multiplying fractions was required. The only hard part was resisting the temptation to leap to an “intuitively obvious” answer.

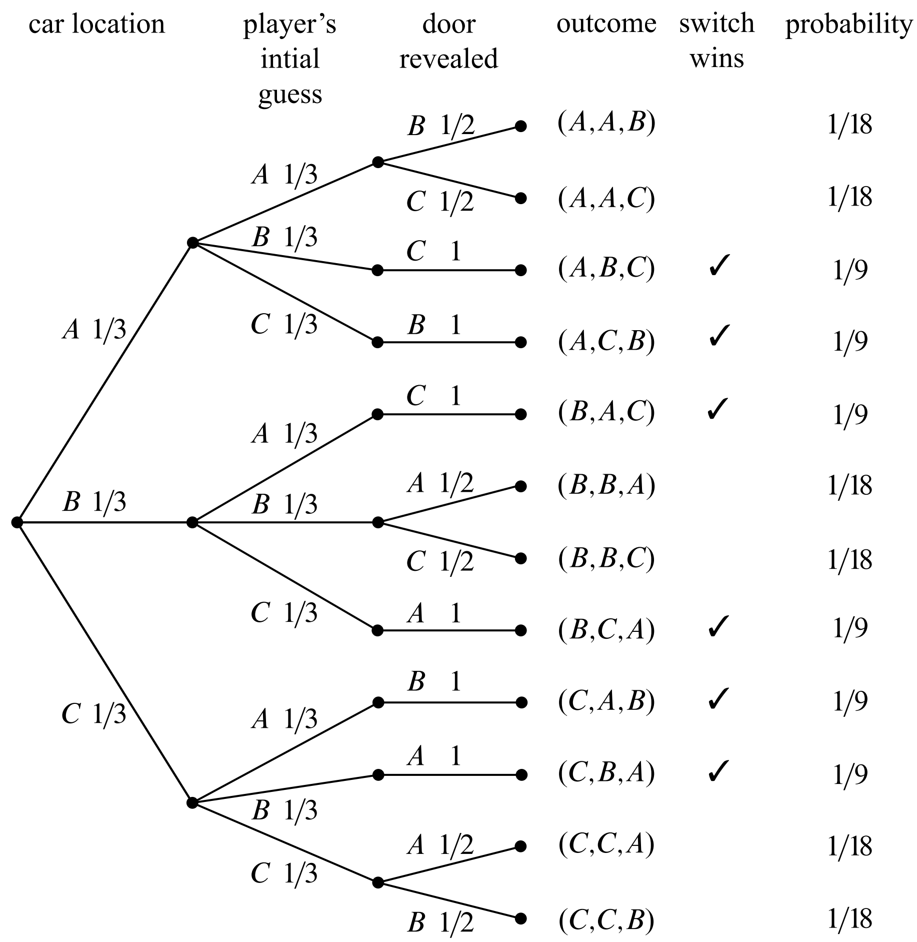

Figure 17.5 The tree diagram for the Monty Hall Problem where edge weights denote the probability of that branch being taken given that we are at the parent of that branch. For example, if the car is behind door A, then there is a 1/3 chance that the player's initial selection is door B. The rightmost column shows the outcome probabilities for the Monty Hall Problem. Each outcome probability is simply the product of the probabilities on the path from the root to the outcome leaf.

17.2.5 An Alternative Interpretation of the Monty Hall Problem

Was Marilyn really right? Our analysis indicates that she was. But a more accurate conclusion is that her answer is correct *provided we accept her interpretation of the question*. There is an equally plausible interpretation in which Marilyn's answer is wrong. Notice that Craig Whitaker's original letter does not say that the host is *required* to reveal a goat and offer the player the option to switch, merely that he *did* these things. In fact, on the *Let's Make a Deal* show, Monty Hall sometimes simply opened the door that the contestant picked initially. Therefore, if he wanted to, Monty could give the option of switching only to contestants who picked the correct door initially. In this case, switching never works!

17.3 Strange Dice

The four-step method is surprisingly powerful. Let's get some more practice with it. Imagine, if you will, the following scenario.

It's a typical Saturday night. You're at your favorite pub, contemplating the true meaning of infinite cardinalities, when a burly-looking biker plops down on the stool next to you. Just as you are about to get your mind around $\text{pow}(\text{pow}(\mathbb{R}))$, biker dude slaps three strange-looking dice on the bar and challenges you to a \$100 wager. His rules are simple. Each player selects one die and rolls it once. The player with the lower value pays the other player \$100.

Naturally, you are skeptical, especially after you see that these are not ordinary dice. Each die has the usual six sides, but opposite sides have the same number on them, and the numbers on the dice are different, as shown in Figure 17.6.

Biker dude notices your hesitation, so he sweetens his offer: he will pay you \$105 if you roll the higher number, but you only need pay him \$100 if he rolls higher, *and* he will let you pick a die first, after which he will pick one of the other two. The sweetened deal sounds persuasive since it gives you a chance to pick what you think is the best die, so you decide you will play. But which of the dice should you choose? Die B is appealing because it has a 9, which is a sure winner if it comes up. Then again, die A has two fairly large numbers, and die C has an 8 and no really small values.

In the end, you choose die B because it has a 9, and then biker dude selects die A. Let's see what the probability is that you will win. (Of course, you probably should have done this before picking die B in the first place.) Not surprisingly, we will use the four-step method to compute this probability.

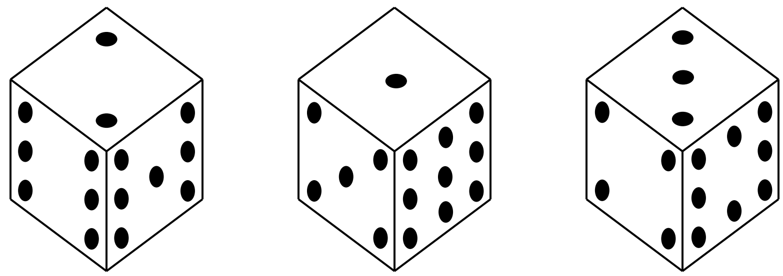

Figure 17.6 The strange dice. The number of pips on each concealed face is the same as the number on the opposite face. For example, when you roll die A, the probabilities of getting a 2, 6, or 7 are each $1/3$.

17.3.1 Die A versus Die B

Step 1: Find the sample space.
The tree diagram for this scenario is shown in Figure 17.7. In particular, the sample space for this experiment are the nine pairs of values that might be rolled with Die A and Die B:

For this experiment, the sample space is a set of nine outcomes:

$$\mathcal{S} = \{\,(2,1),\ (2,5),\ (2,9),\ (6,1),\ (6,5),\ (6,9),\ (7,1),\ (7,5),\ (7,9)\,\}.$$

Step 2: Define events of interest.
We are interested in the event that the number on die A is greater than the number on die B. This event is a set of five outcomes:

$$\{\,(2,1),\ (6,1),\ (6,5),\ (7,1),\ (7,5)\,\}.$$

These outcomes are marked A in the tree diagram in Figure 17.7.

Step 3: Determine outcome probabilities.
To find outcome probabilities, we first assign probabilities to edges in the tree diagram. Each number on each die comes up with probability $1/3$, regardless of the value of the other die. Therefore, we assign all edges probability $1/3$. The probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path, which means that every outcome has probability $1/9$. These probabilities are recorded on the right side of the tree diagram in Figure 17.7.

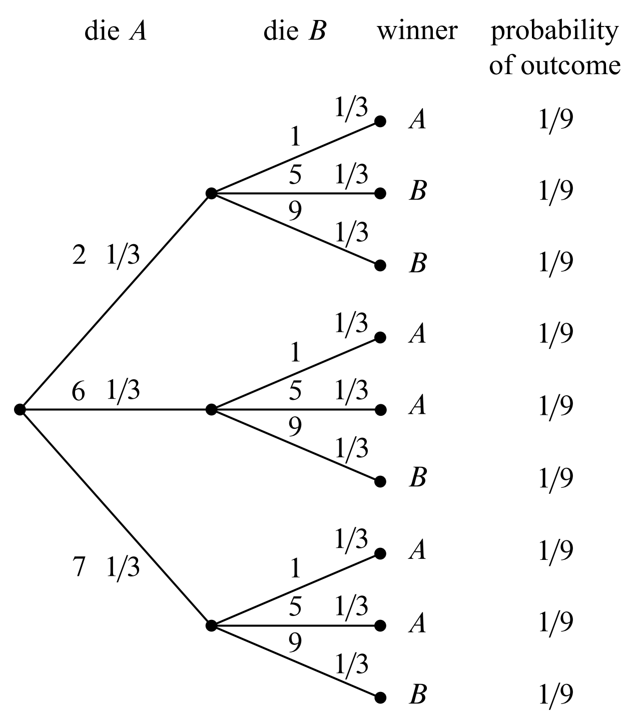

Figure 17.7 The tree diagram for one roll of die A versus die B. Die A wins with probability $5/9$.

Step 4: Compute event probabilities.

The probability of an event is the sum of the probabilities of the outcomes in that event. In this case, all the outcome probabilities are the same, so we say that the sample space is *uniform*. Computing event probabilities for uniform sample spaces is particularly easy since you just have to compute the number of outcomes in the event. In particular, for any event E in a uniform sample space $\mathcal{S}$,

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|}. \tag{17.3}$$

In this case, E is the event that die A beats die B, so $|E| = 5$, $|\mathcal{S}| = 9$, and

$$\Pr[E] = 5/9.$$

This is bad news for you. Die A beats die B more than half the time and, not surprisingly, you just lost \$100.

Biker dude consoles you on your "bad luck" and, given that he's a sensitive guy beneath all that leather, he offers to go double or nothing.[1] Given that your wallet only has \$25 in it, this sounds like a good plan. Plus, you figure that choosing die A will give *you* the advantage.

So you choose A, and then biker dude chooses C. Can you guess who is more likely to win? (Hint: it is generally not a good idea to gamble with someone you don't know in a bar, especially when you are gambling with strange dice.)

17.3.2 Die A versus Die C

We can construct the tree diagram and outcome probabilities as before. The result is shown in Figure 17.8, and there is bad news again. Die C will beat die A with probability $5/9$, and you lose once again.

You now owe the biker dude \$200 and he asks for his money. You reply that you need to go to the bathroom.

17.3.3 Die B versus Die C

Being a sensitive guy, biker dude nods understandingly and offers yet another wager. This time, he'll let you have die C. He'll even let you raise the wager to \$200 so you can win your money back.

This is too good a deal to pass up. You know that die C is likely to beat die A and that die A is likely to beat die B, and so die C is *surely* the best. Whether biker

[1] *Double or nothing* is slang for doing another wager after you have lost the first. If you lose again, you will owe biker dude *double* what you owed him before. If you win, you will owe him *nothing*; in fact, since he should pay you \$210 if he loses, you would come out \$10 ahead.

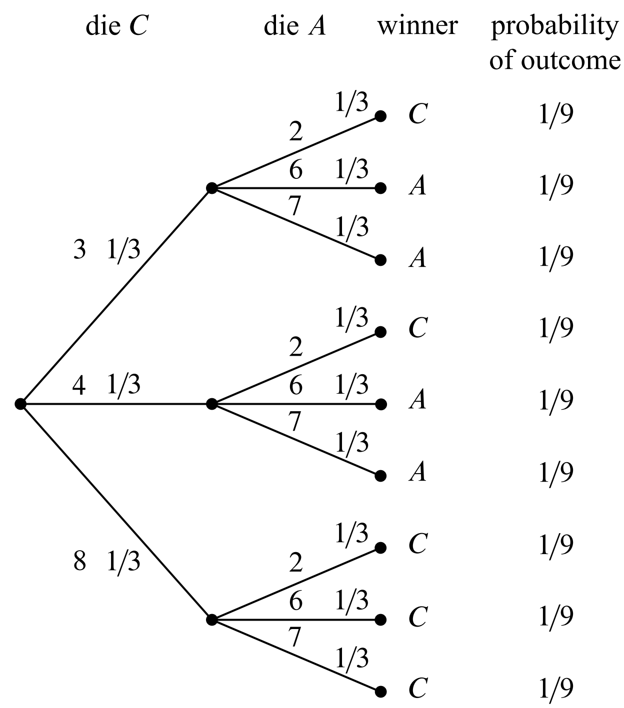

Figure 17.8 The tree diagram for one roll of die C versus die A. Die C wins with probability $5/9$.

dude picks A or B, the odds would be in your favor this time. Biker dude must really be a nice guy.

So you pick C, and then biker dude picks B. Wait—how come you haven't caught on yet and worked out the tree diagram before you took this bet? If you do it now, you'll see by the same reasoning as before that B beats C with probability $5/9$. But surely there is a mistake! How is it possible that

C beats A with probability $5/9$,

A beats B with probability $5/9$,

B beats C with probability $5/9$?

The problem is not with the math, but with your intuition. Since A will beat B more often than not, and B will beat C more often than not, it *seems* like A ought to beat C more often than not, that is, the "beats more often" relation ought to be *transitive*. But this intuitive idea is simply false: whatever die you pick, biker dude can pick one of the others and be likely to win. So picking first is actually a disadvantage, and as a result, you now owe biker dude \$400.

Just when you think matters can't get worse, biker dude offers you one final wager for \$1,000. This time, instead of rolling each die once, you will each roll your die twice, and your score is the sum of your rolls, and he will even let you pick your die second, that is, after he picks his. Biker dude chooses die B. Now you know that die A will beat die B with probability $5/9$ on one roll, so, jumping at this chance to get ahead, you agree to play, and you pick die A. After all, you figure that since a roll of die A beats a roll of die B more often that not, two rolls of die A are even more likely to beat two rolls of die B, right?

Wrong! (Did we mention that playing strange gambling games with strangers in a bar is a bad idea?)

17.3.4 Rolling Twice

If each player rolls twice, the tree diagram will have four levels and $3^4 = 81$ outcomes. This means that it will take a while to write down the entire tree diagram. But it's easy to write down the first two levels as in Figure 17.9(a) and then notice that the remaining two levels consist of nine identical copies of the tree in Figure 17.9(b).

The probability of each outcome is $(1/3)^4 = 1/81$ and so, once again, we have a uniform probability space. By equation (17.3), this means that the probability that A wins is the number of outcomes where A beats B divided by 81.

To compute the number of outcomes where A beats B, we observe that the two rolls of die A result in nine equally likely outcomes in a sample space $\mathcal{S}_A$ in which

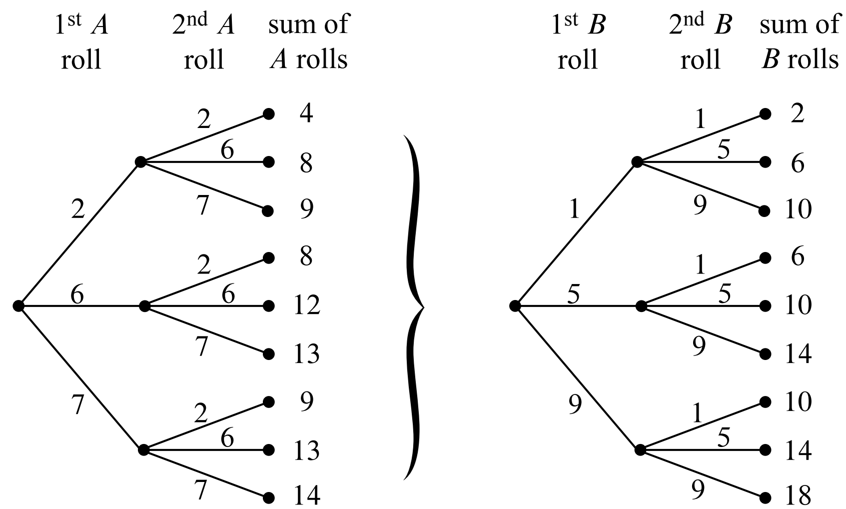

Figure 17.9 Parts of the tree diagram for die B versus die A where each die is rolled twice. The first two levels are shown in (a). The last two levels consist of nine copies of the tree in (b).

the two-roll sums take the values

$$(4, 8, 8, 9, 9, 12, 13, 13, 14).$$

Likewise, two rolls of die B result in nine equally likely outcomes in a sample space $\mathcal{S}_B$ in which the two-roll sums take the values

$$(2, 6, 6, 10, 10, 10, 14, 14, 18).$$

We can treat the outcome of rolling both dice twice as a pair $(x, y) \in \mathcal{S}_A \times \mathcal{S}_B$, where A wins iff the sum of the two A-rolls of outcome x is larger the sum of the two B-rolls of outcome y. If the A-sum is 4, there is only one y with a smaller B-sum, namely, when the B-sum is 2. If the A-sum is 8, there are three y's with a smaller B-sum, namely, when the B-sum is 2 or 6. Continuing the count in this way, the number of pairs (x, y) for which the A-sum is larger than the B-sum is

$$1 + 3 + 3 + 3 + 3 + 6 + 6 + 6 + 6 = 37.$$

A similar count shows that there are 42 pairs for which B-sum is larger than the A-sum, and there are two pairs where the sums are equal, namely, when they both equal 14. This means that A *loses* to B with probability $42/81 > 1/2$ and ties with probability $2/81$. Die A wins with probability only $37/81$.

How can it be that A is more likely than B to win with one roll, but B is more likely to win with two rolls? Well, why not? The only reason we'd think otherwise is our unreliable, untrained intuition. (Even the authors were surprised when they first learned about this, but at least they didn't lose $1400 to biker dude.) In fact, the die strength reverses no matter which two die we picked. So for one roll,

$$A \succ B \succ C \succ A,$$

but for two rolls,

$$A \prec B \prec C \prec A,$$

where we have used the symbols $\succ$ and $\prec$ to denote which die is more likely to result in the larger value.

The weird behavior of the three strange dice above generalizes in a remarkable way: there are arbitrarily large sets of dice which will beat each other in any desired pattern according to how many times the dice are rolled.[2]

17.4 The Birthday Principle

There are 95 students in a class. What is the probability that some birthday is shared by two people? Comparing 95 students to the 365 possible birthdays, you might guess the probability lies somewhere around $1/4$—but you'd be wrong: the probability that there will be two people in the class with matching birthdays is actually more than 0.9999.

To work this out, we'll assume that the probability that a randomly chosen student has a given birthday is $1/d$. We'll also assume that a class is composed of n randomly and independently selected students. Of course $d = 365$ and $n = 95$ in this case, but we're interested in working things out in general. These randomness assumptions are not really true, since more babies are born at certain times of year, and students' class selections are typically not independent of each other, but simplifying in this way gives us a start on analyzing the problem. More importantly, these assumptions are justifiable in important computer science applications of birthday matching. For example, birthday matching is a good model for collisions between items randomly inserted into a hash table. So we won't worry about things like spring procreation preferences that make January birthdays more common, or about twins' preferences to take classes together (or not).

[2] **TBA - Reference Ron Graham paper.**

17.4.1 Exact Formula for Match Probability

There are d^n sequences of n birthdays, and under our assumptions, these are equally likely. There are $d(d-1)(d-2)\cdots(d-(n-1))$ length n sequences of distinct birthdays. That means the probability that everyone has a different birthday is:

$$\begin{aligned}
&\frac{d(d-1)(d-2)\cdots(d-(n-1))}{d^n} \\
&= \frac{d}{d}\cdot\frac{d-1}{d}\cdot\frac{d-2}{d}\cdots\frac{d-(n-1)}{d} && (17.4)\\
&= \left(1-\frac{0}{d}\right)\left(1-\frac{1}{d}\right)\left(1-\frac{2}{d}\right)\cdots\left(1-\frac{n-1}{d}\right) && (17.5)
\end{aligned}$$

Now we simplify (17.5) using the fact that $1-x < e^{-x}$ for all $x > 0$. This follows by truncating the Taylor series $e^{-x} = 1 - x + x^2/2! - x^3/3! + \cdots$. The approximation $e^{-x} \approx 1 - x$ is pretty accurate when x is small.

$$\begin{aligned}
&\left(1-\frac{0}{d}\right)\left(1-\frac{1}{d}\right)\left(1-\frac{2}{d}\right)\cdots\left(1-\frac{n-1}{d}\right) \\
&< e^0 \cdot e^{-1/d} \cdot e^{-2/d} \cdots e^{-(n-1)/d} && (17.6)\\
&= e^{-\left(\sum_{i=1}^{n-1} i/d\right)} \\
&= e^{-(n(n-1)/(2d))}. && (17.7)
\end{aligned}$$

For $n = 95$ and $d = 365$, the value of (17.7) is less than $1/200{,}000$, which means the probability of having some pair of matching birthdays actually is more than $1 - 1/200{,}000 > 0.99999$. So it would be pretty astonishing if there were no pair of students in the class with matching birthdays.

For $d \le n^2/2$, the probability of no match turns out to be asymptotically equal to the upper bound (17.7). For $d = n^2/2$ in particular, the probability of no match is asymptotically equal to $1/e$. This leads to a rule of thumb which is useful in many contexts in computer science:

The Birthday Principle

If there are d days in a year and $\sqrt{2d}$ people in a room, then the probability that two share a birthday is about $1 - 1/e \approx 0.632$.

For example, the Birthday Principle says that if you have $\sqrt{2 \cdot 365} \approx 27$ people in a room, then the probability that two share a birthday is about 0.632. The actual probability is about 0.626, so the approximation is quite good.

Among other applications, it implies that to use a hash function that maps n items into a hash table of size d, you can expect many collisions if n^2 is more than a small fraction of d. The Birthday Principle also famously comes into play as the basis of “birthday attacks” that crack certain cryptographic systems.

17.5 Set Theory and Probability

Let’s abstract what we’ve just done into a general mathematical definition of sample spaces and probability.

17.5.1 Probability Spaces

Definition 17.5.1. A countable *sample space* $\mathcal{S}$ is a nonempty countable set.[3] An element $\omega \in \mathcal{S}$ is called an *outcome*. A subset of $\mathcal{S}$ is called an *event*.

Definition 17.5.2. A *probability function* on a sample space $\mathcal{S}$ is a total function $\Pr : \mathcal{S} \to \mathbb{R}$ such that

- $\Pr[\omega] \geq 0$ for all $\omega \in \mathcal{S}$, and
- $\sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1$.

A sample space together with a probability function is called a *probability space*. For any event $E \subseteq \mathcal{S}$, the *probability of E* is defined to be the sum of the probabilities of the outcomes in E:

$$\Pr[E] ::= \sum_{\omega \in E} \Pr[\omega].$$

In the previous examples there were only finitely many possible outcomes, but we’ll quickly come to examples that have a countably infinite number of outcomes.

The study of probability is closely tied to set theory because any set can be a sample space and any subset can be an event. General probability theory deals with uncountable sets like the set of real numbers, but we won’t need these, and sticking to countable sets lets us define the probability of events using sums instead of integrals. It also lets us avoid some distracting technical problems in set theory like the Banach-Tarski “paradox” mentioned in Chapter 8.

[3] Yes, sample spaces can be infinite. If you did not read Chapter 8, don’t worry—*countable* just means that you can list the elements of the sample space as $\omega_0, \omega_1, \omega_2, \ldots$.

17.5.2 Probability Rules from Set Theory

Most of the rules and identities that we have developed for finite sets extend very naturally to probability.

An immediate consequence of the definition of event probability is that for *disjoint* events E and F,

$$\Pr[E \cup F] = \Pr[E] + \Pr[F].$$

This generalizes to a countable number of events:

Rule 17.5.3 (Sum Rule)**.** *If $E_0, E_1, \ldots, E_n, \ldots$ are pairwise disjoint events, then*

$$\Pr\left[\bigcup_{n \in \mathbb{N}} E_n\right] = \sum_{n \in \mathbb{N}} \Pr[E_n].$$

The Sum Rule lets us analyze a complicated event by breaking it down into simpler cases. For example, if the probability that a randomly chosen MIT student is native to the United States is 60%, to Canada is 5%, and to Mexico is 5%, then the probability that a random MIT student is native to one of these three countries is 70%.

Another consequence of the Sum Rule is that $\Pr[A] + \Pr[\overline{A}] = 1$, which follows because $\Pr[\mathcal{S}] = 1$ and $\mathcal{S}$ is the union of the disjoint sets A and $\overline{A}$. This equation often comes up in the form:

$$\Pr[\overline{A}] = 1 - \Pr[A]. \qquad \text{(Complement Rule)}$$

Sometimes the easiest way to compute the probability of an event is to compute the probability of its complement and then apply this formula.

Some further basic facts about probability parallel facts about cardinalities of finite sets. In particular:

$$\begin{aligned}
&\Pr[B - A] = \Pr[B] - \Pr[A \cap B], && \text{(Difference Rule)}\\
&\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B], && \text{(Inclusion-Exclusion)}\\
&\Pr[A \cup B] \le \Pr[A] + \Pr[B], && \text{(Boole's Inequality)}\\
&\text{If } A \subseteq B, \text{ then } \Pr[A] \le \Pr[B]. && \text{(Monotonicity Rule)}
\end{aligned}$$

The Difference Rule follows from the Sum Rule because B is the union of the disjoint sets $B - A$ and $A \cap B$. Inclusion-Exclusion then follows from the Sum and Difference Rules, because $A \cup B$ is the union of the disjoint sets A and $B - A$. Boole's inequality is an immediate consequence of Inclusion-Exclusion since probabilities are nonnegative. Monotonicity follows from the definition of event probability and the fact that outcome probabilities are nonnegative.

The two-event Inclusion-Exclusion equation above generalizes to any finite set of events in the same way as the corresponding Inclusion-Exclusion rule for n sets. Boole's inequality also generalizes to both finite and countably infinite sets of events:

Rule 17.5.4 (Union Bound)**.**

$$\Pr[E_1 \cup \cdots \cup E_n \cup \cdots] \le \Pr[E_1] + \cdots + \Pr[E_n] + \cdots. \tag{17.8}$$

The Union Bound is useful in many calculations. For example, suppose that E_i is the event that the i-th critical component among n components in a spacecraft fails. Then $E_1 \cup \cdots \cup E_n$ is the event that *some* critical component fails. If $\sum_{i=1}^{n} \Pr[E_i]$ is small, then the Union Bound can provide a reassuringly small upper bound on this overall probability of critical failure.

17.5.3 Uniform Probability Spaces

Definition 17.5.5. A finite probability space $\mathcal{S}$ is said to be *uniform* if $\Pr[\omega]$ is the same for every outcome $\omega \in \mathcal{S}$.

As we saw in the strange dice problem, uniform sample spaces are particularly easy to work with. That's because for any event $E \subseteq \mathcal{S}$,

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|}. \tag{17.9}$$

This means that once we know the cardinality of E and $\mathcal{S}$, we can immediately obtain $\Pr[E]$. That's great news because we developed lots of tools for computing the cardinality of a set in Part III.

For example, suppose that you select five cards at random from a standard deck of 52 cards. What is the probability of having a full house? Normally, this question would take some effort to answer. But from the analysis in Section 15.7.2, we know that

$$|\mathcal{S}| = \binom{52}{5}$$

and

$$|E| = 13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$$

where E is the event that we have a full house. Since every five-card hand is equally

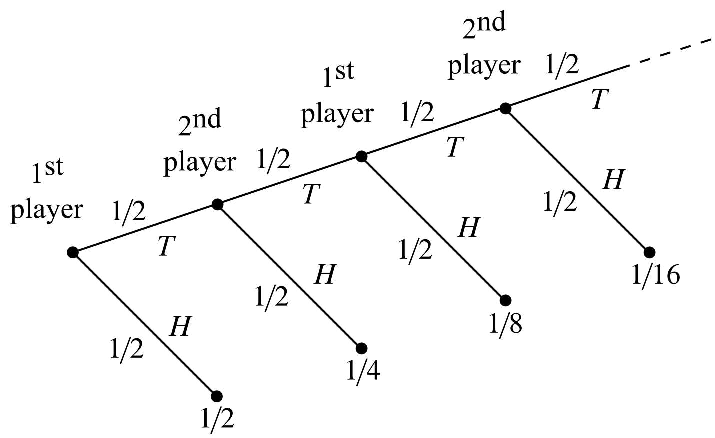

Figure 17.10 The tree diagram for the game where players take turns flipping a fair coin. The first player to flip heads wins.

likely, we can apply equation (17.9) to find that

$$\begin{aligned}
\Pr[E] &= \frac{13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}}{\binom{52}{5}} \\
&= \frac{13 \cdot 12 \cdot 4 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = \frac{18}{12495} \\
&\approx \frac{1}{694}.
\end{aligned}$$

17.5.4 Infinite Probability Spaces

Infinite probability spaces are fairly common. For example, two players take turns flipping a fair coin. Whoever flips heads first is declared the winner. What is the probability that the first player wins? A tree diagram for this problem is shown in Figure 17.10.

The event that the first player wins contains an infinite number of outcomes, but we can still sum their probabilities:

$$\begin{aligned}
\Pr[\text{first player wins}] &= \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \frac{1}{128} + \cdots \\
&= \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{1}{4}\right)^n \\
&= \frac{1}{2}\left(\frac{1}{1-1/4}\right) = \frac{2}{3}.
\end{aligned}$$

Similarly, we can compute the probability that the second player wins:

$$\Pr[\text{second player wins}] = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + \cdots = \frac{1}{3}.$$

In this case, the sample space is the infinite set

$$\mathcal{S} ::= \{\, \mathtt{T}^n\mathtt{H} \mid n \in \mathbb{N} \,\},$$

where $\mathtt{T}^n$ stands for a length n string of $\mathtt{T}$'s. The probability function is

$$\Pr[\mathtt{T}^n\mathtt{H}] ::= \frac{1}{2^{n+1}}.$$

To verify that this is a probability space, we just have to check that all the probabilities are nonnegative and that they sum to 1. The given probabilities are all nonnegative, and applying the formula for the sum of a geometric series, we find that

$$\sum_{n \in \mathbb{N}} \Pr[\mathtt{T}^n\mathtt{H}] = \sum_{n \in \mathbb{N}} \frac{1}{2^{n+1}} = 1.$$

Notice that this model does not have an outcome corresponding to the possibility that both players keep flipping tails forever. (In the diagram, flipping forever corresponds to following the infinite path in the tree without ever reaching a leaf/outcome.) If leaving this possibility out of the model bothers you, you're welcome to fix it by adding another outcome ω_{forever} to indicate that that's what happened. Of course since the probabililities of the other outcomes already sum to 1, you have to define the probability of ω_{forever} to be 0. Now outcomes with probability zero will have no impact on our calculations, so there's no harm in adding it in if it makes you happier. On the other hand, in countable probability spaces it isn't necessary to have outcomes with probability zero, and we will generally ignore them.

17.6 References

[19], [26], [30], [34], [38], [39] [43], [42], [51]

Problems for Section 17.2

Practice Problems

Problem 17.1.
Let B be the number of heads that come up on $2n$ independent tosses of a fair coin.

(a) $\Pr[B = n]$ is asymptotically equal to one of the expressions given below. Explain which one.

1. $\frac{1}{\sqrt{2\pi n}}$
2. $\frac{2}{\sqrt{\pi n}}$
3. $\frac{1}{\sqrt{\pi n}}$
4. $\sqrt{\frac{2}{\pi n}}$

Exam Problems

Problem 17.2. **(a)** What's the probability that 0 doesn't appear among k digits chosen independently and uniformly at random?

(b) A box contains 90 good and 10 defective screws. What's the probability that if we pick 10 screws from the box, none will be defective?

(c) First one digit is chosen uniformly at random from $\{1, 2, 3, 4, 5\}$ and is removed from the set; then a second digit is chosen uniformly at random from the remaining digits. What is the probability that an odd digit is picked the second time?

(d) Suppose that you *randomly* permute the digits $1, 2, \cdots, n$, that is, you select a permutation uniformly at random. What is the probability the digit k ends up in the ith position after the permutation?

(e) A fair coin is flipped n times. What's the probability that all the heads occur at the end of the sequence? (If no heads occur, then "all the heads are at the end of the sequence" is vacuously true.)

Class Problems

Problem 17.3.
The New York Yankees and the Boston Red Sox are playing a two-out-of-three

series. In other words, they play until one team has won two games. Then that team is declared the overall winner and the series ends. Assume that the Red Sox win each game with probability $3/5$, regardless of the outcomes of previous games.

Answer the questions below using the four step method. You can use the same tree diagram for all three problems.

(a) What is the probability that a total of 3 games are played?

(b) What is the probability that the winner of the series loses the first game?

(c) What is the probability that the *correct* team wins the series?

Problem 17.4.
To determine which of two people gets a prize, a coin is flipped twice. If the flips are a Head and then a Tail, the first player wins. If the flips are a Tail and then a Head, the second player wins. However, if both coins land the same way, the flips don't count and the whole process starts over.

Assume that on each flip, a Head comes up with probability p, regardless of what happened on other flips. Use the four step method to find a simple formula for the probability that the first player wins. What is the probability that neither player wins?

Hint: The tree diagram and sample space are infinite, so you're not going to finish drawing the tree. Try drawing only enough to see a pattern. Summing all the winning outcome probabilities directly is cumbersome. However, a neat trick solves this problem—and many others. Let s be the sum of all winning outcome probabilities in the whole tree. Notice that *you can write the sum of all the winning probabilities in certain subtrees as a function of* s. Use this observation to write an equation in s and then solve.

Homework Problems

Problem 17.5.
Let's see what happens when *Let's Make a Deal* is played with **four** doors. A prize is hidden behind one of the four doors. Then the contestant picks a door. Next, the host opens an unpicked door that has no prize behind it. The contestant is allowed to stick with their original door or to switch to one of the two unopened, unpicked doors. The contestant wins if their final choice is the door hiding the prize.

Let's make the same assumptions as in the original problem:

1. The prize is equally likely to be behind each door.

2. The contestant is equally likely to pick each door initially, regardless of the prize's location.

3. The host is equally likely to reveal each door that does not conceal the prize and was not selected by the player.

Use The Four Step Method to find the following probabilities. The tree diagram may become awkwardly large, in which case just draw enough of it to make its structure clear.

(a) Contestant Stu, a sanitation engineer from Trenton, New Jersey, stays with his original door. What is the probability that Stu wins the prize?

(b) Contestant Zelda, an alien abduction researcher from Helena, Montana, switches to one of the remaining two doors with equal probability. What is the probability that Zelda wins the prize?

Now let's revise our assumptions about how contestants choose doors. Say the doors are labeled A, B, C, and D. Suppose that Carol always opens the *earliest* door possible (the door whose label is earliest in the alphabet) with the restriction that she can neither reveal the prize nor open the door that the player picked.

This gives contestant Mergatroid—an engineering student from Cambridge, MA—just a little more information about the location of the prize. Suppose that Mergatroid always switches to the earliest door, excluding his initial pick and the one Carol opened.

(c) What is the probability that Mergatroid wins the prize?

Problem 17.6.

There were n Immortal Warriors born into our world, but in the end there can be *only one*. The Immortals' original plan was to stalk the world for centuries, dueling one another with ancient swords in dramatic landscapes until only one survivor remained. However, after a thought-provoking discussion probability, they opt to give the following protocol a try:

(i) The Immortals forge a coin that comes up heads with probability p.

(ii) Each Immortal flips the coin once.

(iii) If *exactly one* Immortal flips heads, then they are declared The One. Otherwise, the protocol is declared a failure, and they all go back to hacking each other up with swords.

One of the Immortals (Kurgan from the Russian steppe) argues that as n grows large, the probability that this protocol succeeds must tend to zero. Another (McLeod from the Scottish highlands) argues that this need not be the case, provided p is chosen carefully.

(a) A natural sample space to use to model this problem is $\{H, T\}^n$ of length-n sequences of H and T's, where the successive H's and T's in an outcome correspond to the Head or Tail flipped on each one of the n successive flips. Explain how a tree diagram approach leads to assigning a probability to each outcome that depends only on p, n and the number h of H's in the outcome.

(b) What is the probability that the experiment succeeds as a function of p and n?

(c) How should p, the bias of the coin, be chosen in order to maximize the probability that the experiment succeeds?

(d) What is the probability of success if p is chosen in this way? What quantity does this approach when n, the number of Immortal Warriors, grows large?

Problem 17.7.
We play a game with a deck of 52 regular playing cards, of which 26 are red and 26 are black. I randomly shuffle the cards and place the deck face down on a table. You have the option of “taking” or “skipping” the top card. If you skip the top card, then that card is revealed and we continue playing with the remaining deck. If you take the top card, then the game ends; you win if the card you took was revealed to be black, and you lose if it was red. If we get to a point where there is only one card left in the deck, you must take it. Prove that you have no better strategy than to take the top card—which means your probability of winning is 1/2.

Hint: Prove by induction the more general claim that for a randomly shuffled deck of n cards that are red or black—not necessarily with the same number of red cards and black cards—there is no better strategy than taking the top card.

Problems for Section 17.5

Class Problems

Problem 17.8.
Suppose there is a system with n components, and we know from past experience that any particular component will fail in a given year with probability p. That is,

letting F_i be the event that the ith component fails within one year, we have

$$\Pr[F_i] = p$$

for $1 \le i \le n$. The *system* will fail if *any one* of its components fails. What can we say about the probability that the system will fail within one year?

Let F be the event that the system fails within one year. Without any additional assumptions, we can't get an exact answer for $\Pr[F]$. However, we can give useful upper and lower bounds, namely,

$$p \le \Pr[F] \le np. \tag{17.10}$$

We may as well assume $p < 1/n$, since the upper bound is trivial otherwise. For example, if $n = 100$ and $p = 10^{-5}$, we conclude that there is at most one chance in 1000 of system failure within a year and at least one chance in 100,000.

Let's model this situation with the sample space $\mathcal{S} ::= \text{pow}([1..n])$ whose outcomes are subsets of positive integers $\le n$, where $s \in \mathcal{S}$ corresponds to the indices of exactly those components that fail within one year. For example, $\{2, 5\}$ is the outcome that the second and fifth components failed within a year and none of the other components failed. So the outcome that the system did not fail corresponds to the empty set $\emptyset$.

(a) Show that the probability that the system fails could be as small as p by describing appropriate probabilities for the outcomes. Make sure to verify that the sum of your outcome probabilities is 1.

(b) Show that the probability that the system fails could actually be as large as np by describing appropriate probabilities for the outcomes. Make sure to verify that the sum of your outcome probabilities is 1.

(c) Prove inequality (17.10).

Problem 17.9.

Here are some handy rules for reasoning about probabilities that all follow directly from the Disjoint Sum Rule. Prove them.

$$\Pr[A - B] = \Pr[A] - \Pr[A \cap B] \qquad \text{(Difference Rule)}$$

$$\Pr[\overline{A}] = 1 - \Pr[A] \qquad \text{(Complement Rule)}$$

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] \qquad \text{(Inclusion-Exclusion)}$$

$$\Pr[A \cup B] \le \Pr[A] + \Pr[B] \qquad \text{(2-event Union Bound)}$$

$$A \subseteq B \text{ IMPLIES } \Pr[A] \le \Pr[B] \qquad \text{(Monotonicity)}$$

Homework Problems

Problem 17.10.
Prove the following probabilistic inequality, referred to as the *Union Bound*.

Let $A_1, A_2, \ldots, A_n, \ldots$ be events. Then

$$\Pr\left[\bigcup_{n \in \mathbb{N}} A_n\right] \le \sum_{n \in \mathbb{N}} \Pr[A_n].$$

Hint: Replace the A_n's by pairwise disjoint events and use the Sum Rule.

Problem 17.11.
The results of a round robin tournament in which every two people play each other and one of them wins can be modelled a *tournament digraph*—a digraph with exactly one edge between each pair of distinct vertices, but we'll continue to use the language of players beating each other.

An n-player tournament is k*-neutral* for some $k \in [0, n)$, when, for every set of k players, there is another player who beats them all. For example, being 1-neutral is the same as not having a "best" player who beats everyone else.

This problem shows that for any fixed k, if n is large enough, there will be a k-neutral tournament of n players. We will do this by reformulating the question in terms of probabilities. In particular, for any fixed n, we assign probabilities to each n-vertex tournament digraph by choosing a direction for the edge between any two vertices, independently and with equal probability for each edge.

(a) For any set S of k players, let B_S be the event that no contestant beats everyone in S. Express $\Pr[B_S]$ in terms of n and k.

(b) Let Q_k be the event that the tournament digraph is *not* k-neutral. Prove that

$$\Pr[Q_k] \le \binom{n}{k} \alpha^{n-k},$$

where $\alpha ::= 1 - (1/2)^k$.

Hint: Let S range over the size-k subsets of players, so

$$Q_k = \bigcup_S B_S\,.$$

Use Boole's inequality.

(c) Conclude that if n is large enough (relative to k), then $\Pr[Q_k] < 1$.

(d) Explain why the previous result implies that for every integer k, there is an n-player k-neutral tournament (for a large enough $n \in \mathbb{N}$).

Homework Problems

Problem 17.12.
Suppose you repeatedly flip a fair coin until three consecutive flips match the pattern `HHT` or the pattern `TTH` occurs. What is the probability you will see `HHT` first? Define a suitable probability space that models the coin flipping and use it to explain your answer.

Hint: Symmetry between Heads and Tails.