

---

## 14 Events and Probability Spaces

---

### 14.1 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.

#### 14.1.1 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. Other interpretations are at least as defensible, and some actually 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?”

---

## 14.2 The Four Step Method

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

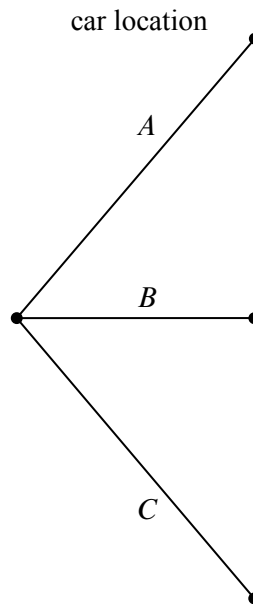
1. How do we model the situation mathematically?
2. How do we solve the resulting mathematical problem?

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, the structured thinking that this approach imposes 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.

### 14.2.1 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 14.1** The first level in a tree diagram for the Monty Hall Problem. The branches correspond to the door behind which the car is located.

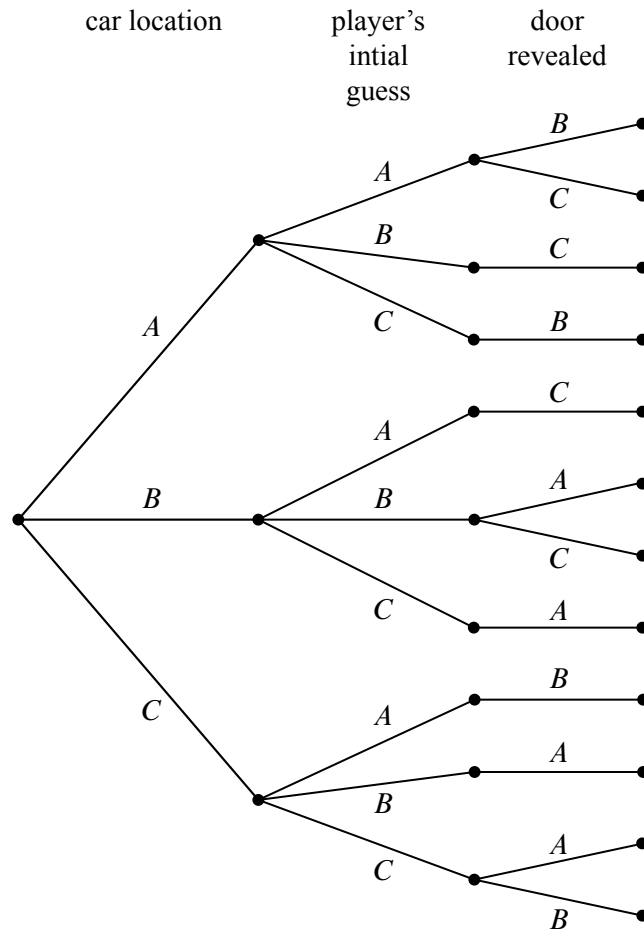
3. The door that the host opens to reveal a goat.

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 14.1. 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 14.2.

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 14.2** 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 14.3 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.

### 14.2.2 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”. Each of these phrases characterizes a set of outcomes. For example, the outcomes specified by “the prize is behind door C” is:

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

A set of outcomes is called an *event* and it is a subset of the sample space. So the event that the player initially picked the door concealing the prize is the set:

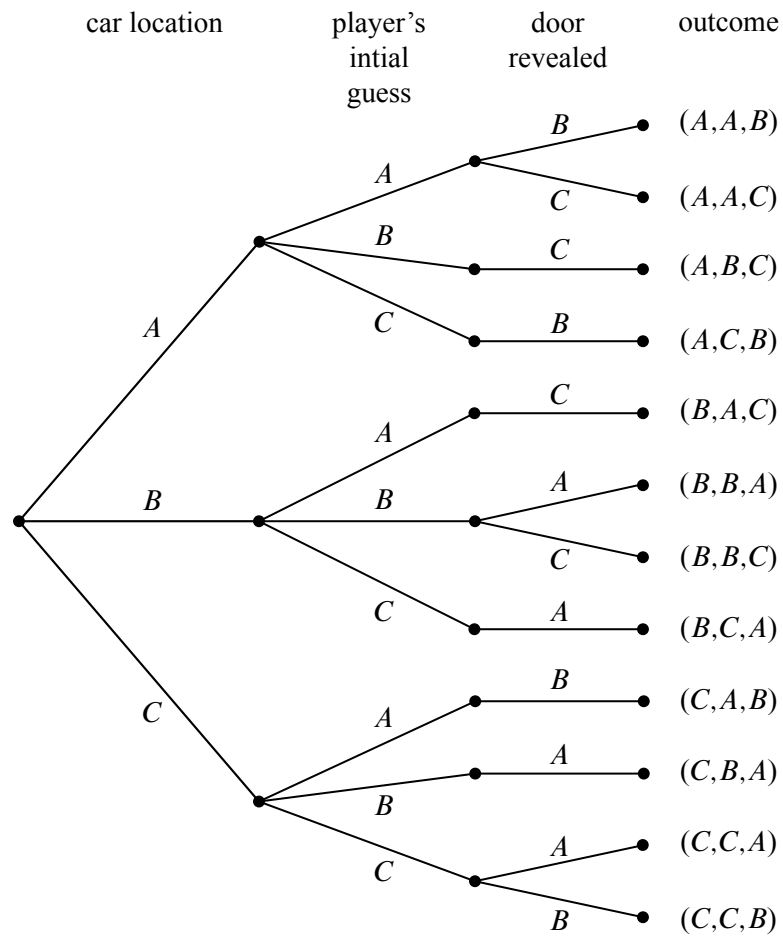
$$\{(A, A, B), (A, A, C), (B, B, A), (B, B, C), (C, C, A), (C, C, B)\}.$$

And what we’re really after, the event that the player wins by switching, is the set of outcomes:

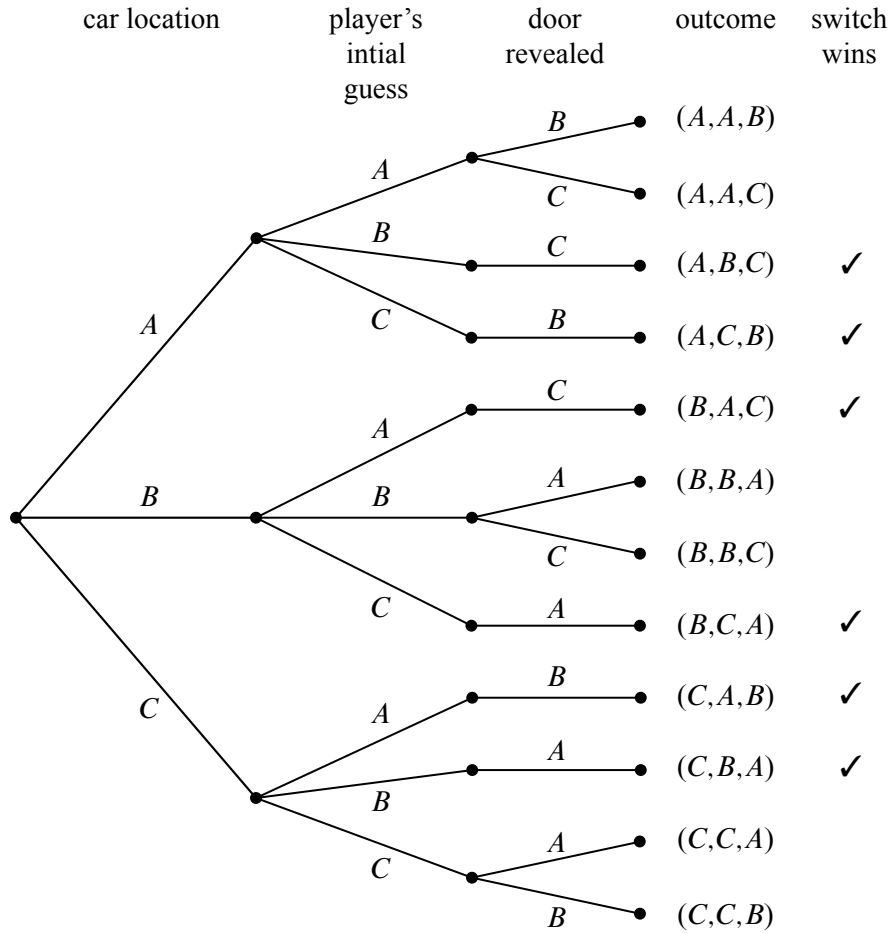
$$\{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\}.$$

These outcomes are denoted with a check mark in Figure 14.4.

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 14.3** The tree diagram for the Monty Hal 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 14.4** The tree diagram for the Monty Hall Problem where the outcomes in the event where the player wins by switching are denoted with a check mark.

### 14.2.3 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 outcome probabilities must be one, reflecting the fact that there always is 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 14.5.

#### Step 3b: Compute Outcome Probabilities

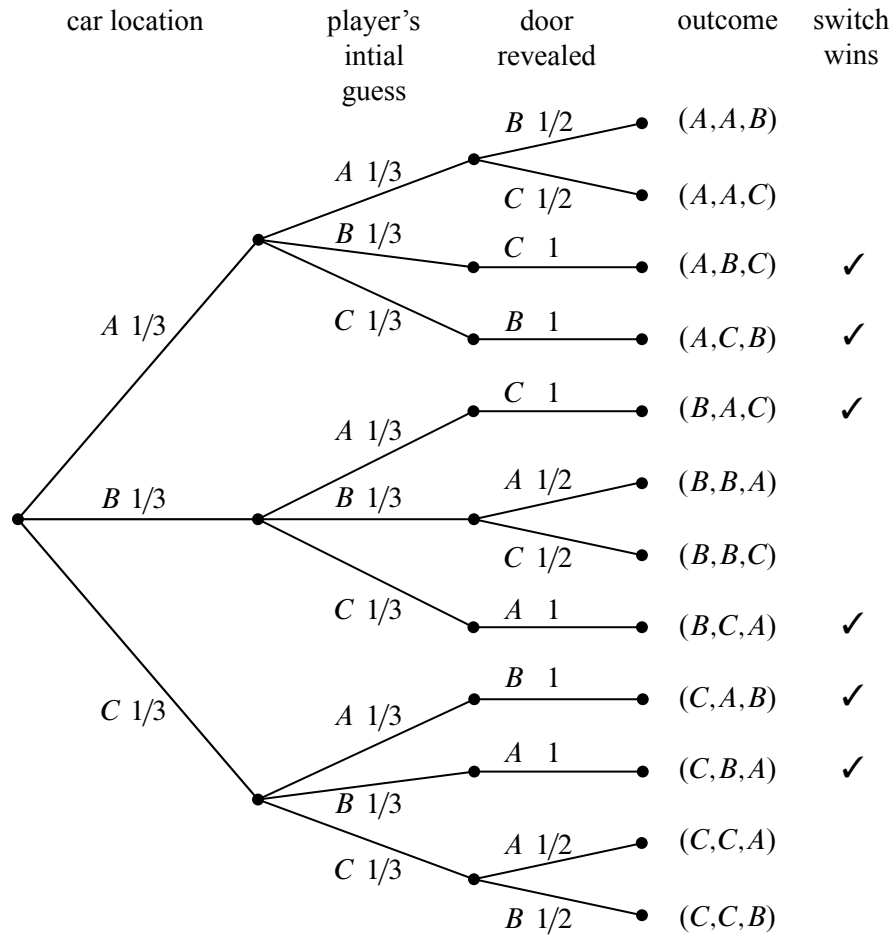
Our next job is to convert edge probabilities into outcome probabilities. This is a purely mechanical process: *the probability of an outcome is equal to the product of the edge-probabilities on the path from the root to that outcome*. For example, the probability of the topmost outcome in Figure 14.5,  $(A, A, B)$ , is

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

There’s an easy, intuitive justification for this rule. 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, it seems that 1 path in 18 should arrive at the  $(A, A, B)$  leaf, which is precisely the probability we assign it.





**Figure 14.5** 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.

We have illustrated all of the outcome probabilities in Figure 14.6.

Specifying the probability of each outcome amounts to defining a function that maps each outcome to a probability. This function is usually called **Pr**. In these terms, we’ve just determined that:

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

#### 14.2.4 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 event that the player wins by switching is:<sup>1</sup>

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

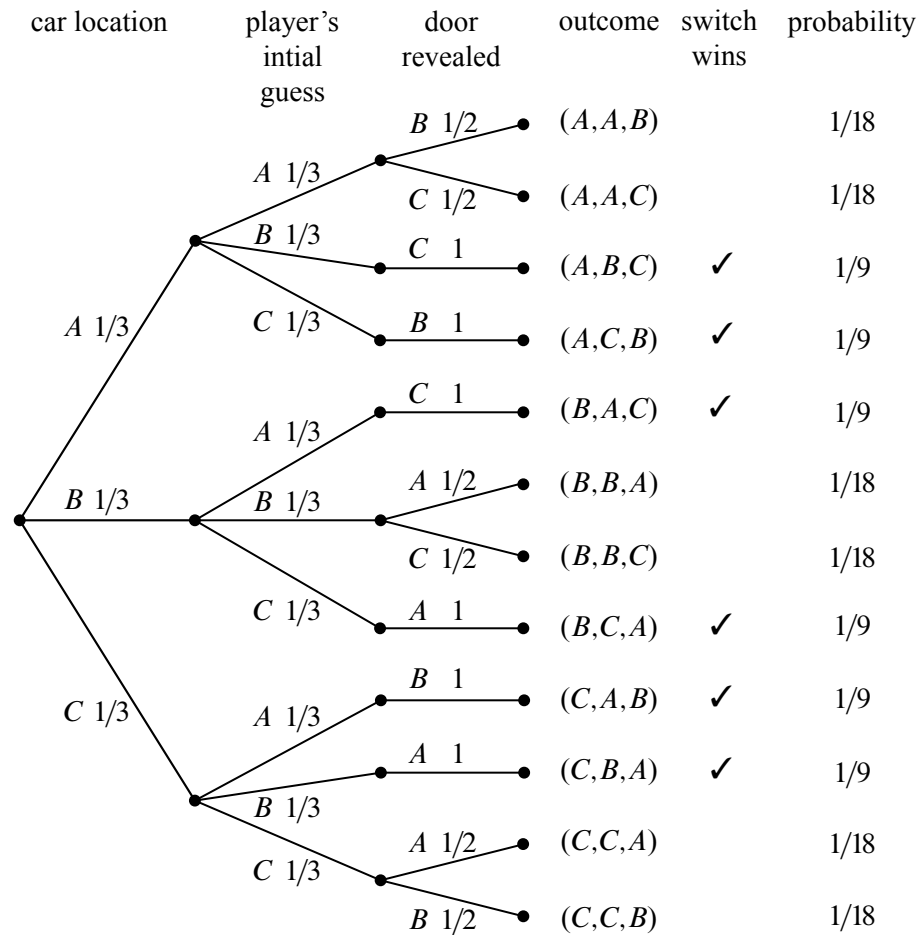
It seems Marilyn’s answer is correct! A player who switches doors wins the car with probability 2/3. In contrast, a player who stays with his or her original door wins with probability 1/3, since staying wins if and only if switching loses.

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.

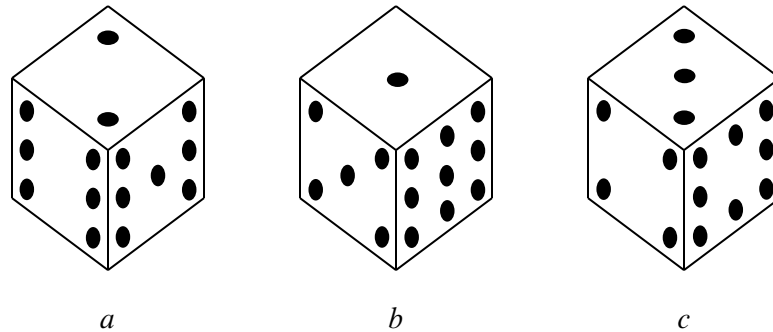
#### 14.2.5 An Alternative Interpretation of the Monty Hall Problem

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*

<sup>1</sup>“Switching wins” is shorthand for the set of outcomes where switching wins; namely,  $\{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\}$ . We will frequently use such shorthand to denote events.



**Figure 14.6** The rightmost column shows the outcome probabilities for the Monty Hall Problem. Each outcome probability is simply the product of the probabilities on the branches on the path from the root to the leaf for that outcome.



**Figure 14.7** 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  $1/3$ .

*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!

### 14.3 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  $\mathcal{P}(\mathcal{P}(\mathbb{R}))$ , biker dude slaps three strange-looking dice on the bar and challenges you to a \$100 wager.

The 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. A quick inspection reveals that these are not ordinary dice. They each have six sides, but the numbers on the dice are different, as shown in Figure 14.7.

Biker dude notices your hesitation and so he offers to let you pick a die first, and

then he will choose his die from the two that are left. That seals the deal since you figure that you now have an advantage.

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  $B$  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.<sup>2</sup> Not surprisingly, we will use the four-step method to compute this probability.

### 14.3.1 Die $A$ versus Die $B$

**Step 1: Find the sample space.**

The sample space for this experiment is worked out in the tree diagram shown in Figure 14.8.<sup>3</sup>

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

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

**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 14.8.

**Step 3: Determine outcome probabilities.**

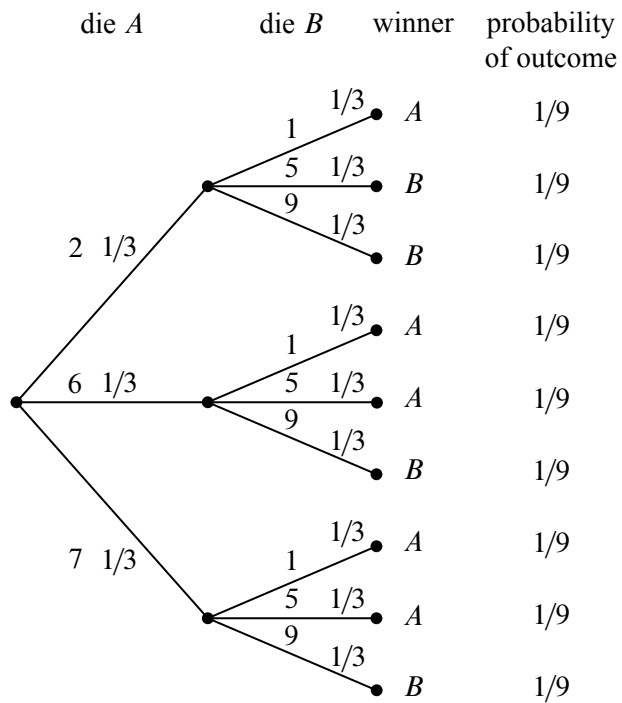
To find outcome probabilities, we first assign probabilities to edges in the tree diagram. Each number on each die comes up with probability  $1/3$ , regardless of the value of the other die. Therefore, we assign all edges probability  $1/3$ . The probability of an outcome is the product of the probabilities 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 14.8.

**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. In general, when the probability of every outcome is the same, we say that the sample space is *uniform*. Computing event probabilities for uniform sample spaces is particularly easy since

<sup>2</sup>Of course, you probably should have done this before picking die  $B$  in the first place.

<sup>3</sup>Actually, the whole probability space is worked out in this one picture. But pretend that each component sort of fades in—nyyrrroom!—as you read about the corresponding step below.



**Figure 14.8** The tree diagram for one roll of die *A* versus die *B*. Die *A* wins with probability  $\frac{5}{9}$ .

you just have to compute the number of outcomes in the event. In particular, for any event  $E$  in a uniform sample space  $\mathcal{S}$ ,

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|} \quad (14.1)$$

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.<sup>4</sup> 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.)

### 14.3.2 Die $A$ versus Die $C$

We can construct the three diagram and outcome probabilities as before. The result is shown in Figure 14.9 and there is bad news again. Die  $C$  will beat die  $A$  with probability  $5/9$ , and you lose once again.

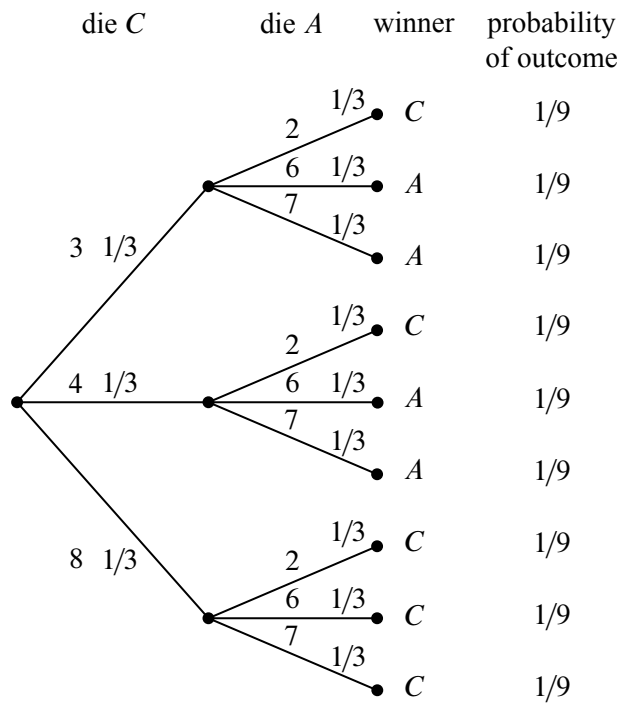
You now owe the biker dude \$200 and he asks for his money. You reply that you need to go to the bathroom.

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

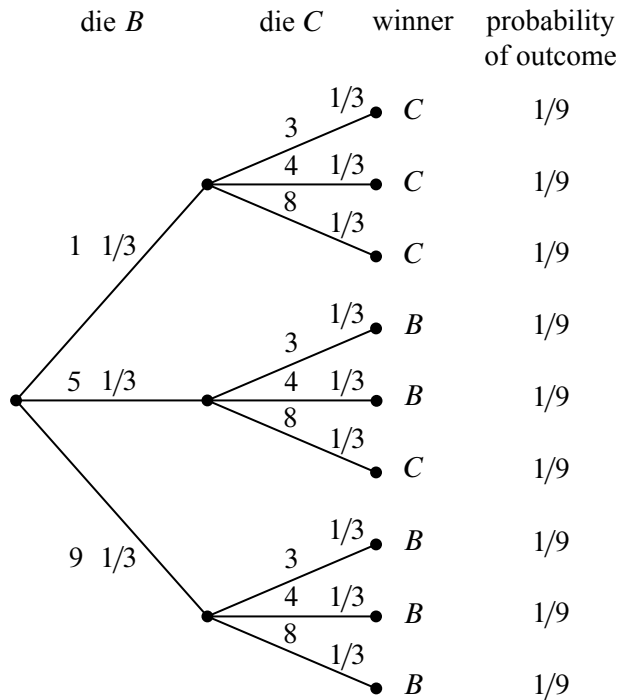
So you pick  $C$ , and then biker dude picks  $B$ . Let’s use the tree method to figure out the probability that you win.

<sup>4</sup>*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 now be even and you will owe him *nothing*.



**Figure 14.9** The tree diagram for one roll of die *C* versus die *A*. Die *C* wins with probability  $\frac{5}{9}$ .





**Figure 14.10** The tree diagram for one roll of die *B* versus die *C*. Die *B* wins with probability  $5/9$ .

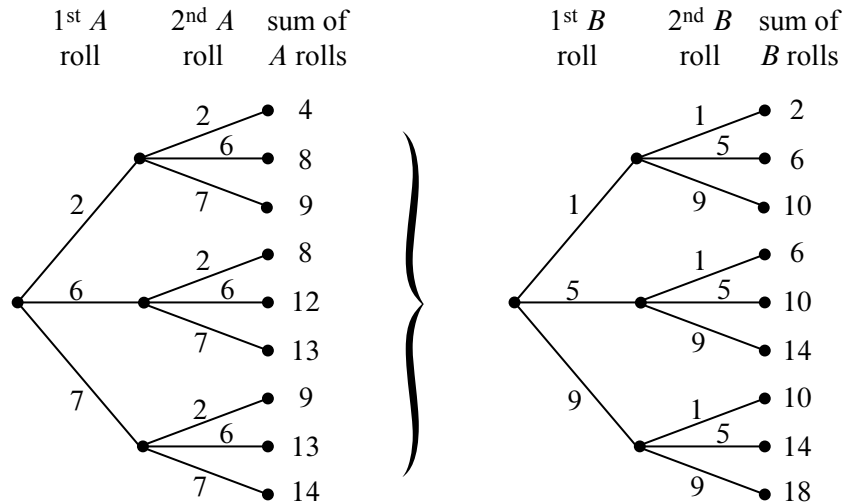
### 14.3.3 Die *B* versus Die *C*

The tree diagram and outcome probabilities for *B* versus *C* are shown in Figure 14.10. But surely there is a mistake! The data in Figure 14.10 shows that die *B* wins with probability  $5/9$ . How is it possible that

- C* beats *A* with probability  $5/9$ ,
- A* beats *B* with probability  $5/9$ , and
- B* beats *C* with probability  $5/9$ ?

The problem is not with the math, but with your intuition. It *seems* that the “likely-to-beat” relation should be transitive. But it is not, and whatever die you pick, biker dude can pick one of the others and be likely to win. So picking first is a big disadvantage and 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, you demand to choose second. Biker dude agrees,



**Figure 14.11** 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).

but with the condition that instead of rolling each die once, you each roll your die twice and your score is the sum of your rolls.

Believing that you finally have a winning wager, you agree.<sup>5</sup> Biker dude chooses die  $B$  and, of course, you grab die  $A$ . That’s because you know that die  $A$  will beat die  $B$  with probability  $5/9$  on one roll and so *surely* two rolls of die  $A$  are likely to beat two rolls of die  $B$ , right?

Wrong!

### 14.3.4 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. We can, however, easily write down the first two levels (as we have done in Figure 14.11(a)) and then notice that the remaining two levels consist of nine identical copies of the tree in Figure 14.11(b).

The probability of each outcome is  $(1/3)^4 = 1/81$  and so, once again, we have a uniform probability space. By Equation 14.1, 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 sum

<sup>5</sup>Did we mention that playing strange gambling games with strangers in a bar is a bad idea?

of the two rolls of die  $A$  is equally likely to be any element of the following multiset:

$$\mathcal{S}_A = \{4, 8, 8, 9, 9, 12, 13, 13, 14\}.$$

The sum of two rolls of die  $B$  is equally likely to be any element of the following multiset:

$$\mathcal{S}_B = \{2, 6, 6, 10, 10, 10, 14, 14, 18\}.$$

We can treat each outcome as a pair  $(x, y) \in \mathcal{S}_A \times \mathcal{S}_B$ , where  $A$  wins iff  $x > y$ . If  $x = 4$ , there is only one  $y$  (namely  $y = 2$ ) for which  $x > y$ . If  $x = 8$ , there are three values of  $y$  for which  $x > y$ . Continuing the count in this way, the number of pairs for which  $x > y$  is

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

A similar count shows that there are 42 pairs for which  $x > y$ , and there are two pairs  $((14, 14), (14, 14))$  which result in ties. 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 1 roll, but  $B$  is more likely to win with 2 rolls?!? Well, why not? The only reason we’d think otherwise is our (faulty) intuition. In fact, the die strength reverses no matter which two die we picked. So for 1 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. This is surprising even to us, but at least we don’t owe biker dude \$1400.

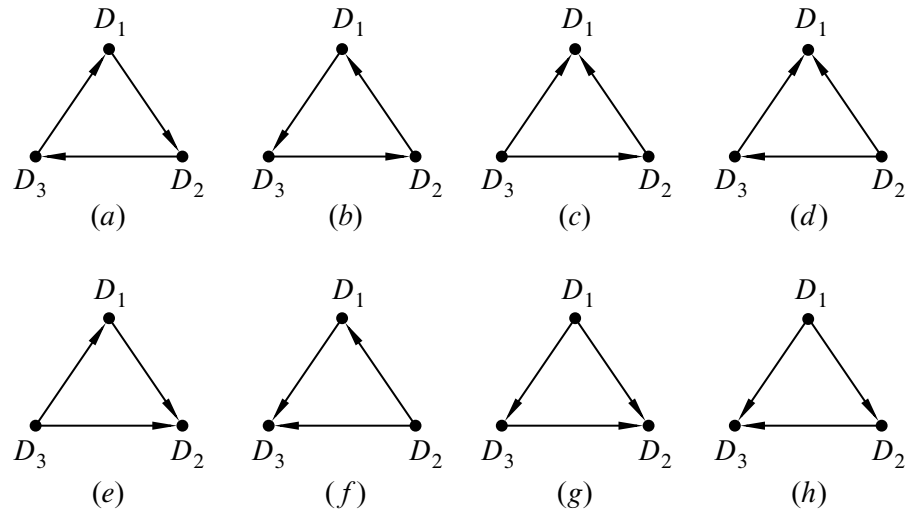
### 14.3.5 Even Stranger Dice

Now that we know that strange things can happen with strange dice, it is natural, at least for mathematicians, to ask how strange things can get. It turns out that things can get very strange. In fact, mathematicians<sup>6</sup> recently made the following discovery:

**Theorem 14.3.1.** *For any  $n \geq 2$ , there is a set of  $n$  dice  $D_1, D_2, \dots, D_n$  such that for any  $n$ -node tournament graph<sup>7</sup>  $G$ , there is a number of rolls  $k$  such that if each*

<sup>6</sup>Reference Ron Graham paper.

<sup>7</sup>Recall that a tournament graph is a directed graph for which there is precisely one directed edge between any two distinct nodes. In other words, for every pair of distinct nodes  $u$  and  $v$ , either  $u$  beats  $v$  or  $v$  beats  $u$ , but not both.



**Figure 14.12** All possible relative strengths for three dice  $D_1$ ,  $D_2$ , and  $D_3$ . The edge  $D_i \rightarrow D_j$  denotes that the sum of rolls for  $D_i$  is likely to be greater than the sum of rolls for  $D_j$ .

die is rolled  $k$  times, then for all  $i \neq j$ , the sum of the  $k$  rolls for  $D_i$  will exceed the sum for  $D_j$  with probability greater than  $1/2$  iff  $D_i \rightarrow D_j$  is in  $G$ .

It will probably take a few attempts at reading Theorem 14.3.1 to understand what it is saying. The idea is that for some sets of dice, by rolling them different numbers of times, the dice have varying strengths relative to each other. (This is what we observed for the dice in Figure 14.7.) Theorem 14.3.1 says that there is a set of (very) strange dice where every possible collection of relative strengths can be observed by varying the number of rolls. For example, the eight possible relative strengths for  $n = 3$  dice are shown in Figure 14.12.

Our analysis for the dice in Figure 14.7 showed that for 1 roll, we have the relative strengths shown in Figure 14.12(a), and for two rolls, we have the (reverse) relative strengths shown in Figure 14.12(b). Can you figure out what other relative strengths are possible for the dice in Figure 14.7 by using more rolls? This might be worth doing if you are prone to gambling with strangers in bars.

## 14.4 Set Theory and Probability

The study of probability is very closely tied to set theory. That is because any set can be a sample space and any subset can be an event. This means that most of the rules and identities that we have developed for sets extend very naturally to probability. We’ll cover several examples in this section, but first let’s review some definitions that should already be familiar.

### 14.4.1 Probability Spaces

**Definition 14.4.1.** A countable<sup>8</sup> *sample space*  $\mathcal{S}$  is a nonempty countable set. An element  $w \in \mathcal{S}$  is called an *outcome*. A subset of  $\mathcal{S}$  is called an *event*.

**Definition 14.4.2.** A *probability function* on a sample space  $\mathcal{S}$  is a total function  $\Pr : \mathcal{S} \rightarrow \mathbb{R}$  such that

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

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  $E$ :

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

### 14.4.2 Probability Rules from Set Theory

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, as follows.

**Rule 14.4.3 (Sum Rule).** If  $\{E_0, E_1, \dots\}$  is collection of disjoint events, then

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

<sup>8</sup>Yes, sample spaces can be infinite. We’ll see some examples shortly. If you did not read Chapter 13, don’t worry—*countable* means that you can list the elements of the sample space as  $w_1, w_2, w_3, \dots$

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 North America is 70%.

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

**Rule 14.4.4** (Complement Rule).

$$\Pr[\bar{A}] = 1 - \Pr[A].$$

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:

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

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  $n$  events in the same way as the corresponding Inclusion-Exclusion rule for  $n$  sets. Boole's inequality also generalizes to

$$\Pr[E_1 \cup \dots \cup E_n] \leq \Pr[E_1] + \dots + \Pr[E_n]. \quad \text{(Union Bound)}$$

This simple Union Bound is useful in many calculations. For example, suppose that  $E_i$  is the event that the  $i$ -th critical component in a spacecraft fails. Then  $E_1 \cup \dots \cup E_n$  is the event that *some* critical component fails. If  $\sum_{i=1}^n \Pr[E_i]$  is small, then the Union Bound can give an adequate upper bound on this vital probability.

### 14.4.3 Uniform Probability Spaces

**Definition 14.4.5.** A finite probability space  $\mathcal{S}$ ,  $\Pr$  is said to be *uniform* if  $\Pr[w]$  is the same for every outcome  $w \in \mathcal{S}$ .

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

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|} \tag{14.2}$$

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 11.7.2, we know that

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

and

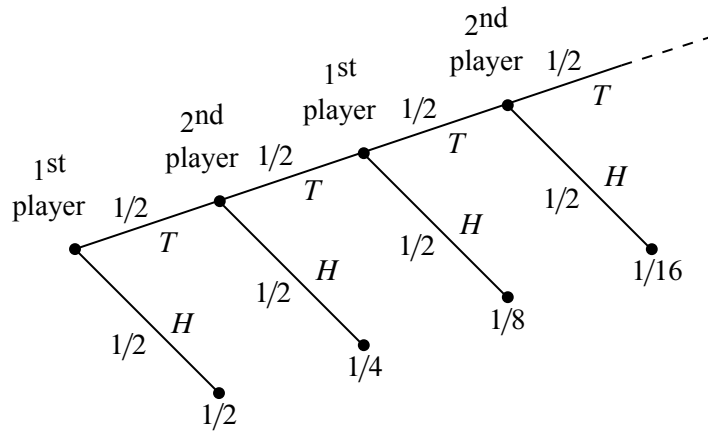
$$|E| = 13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$$

where  $E$  is the event that we have a full house. Since every five-card hand is equally likely, we can apply Equation 14.2 to find that

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

## 14.5 Infinite Probability Spaces

General probability theory deals with uncountable sets like  $\mathbb{R}$ , but in computer science, it is usually sufficient to restrict our attention to countable probability spaces.



**Figure 14.13** The tree diagram for the game where players take turns flipping a fair coin. The first player to flip heads wins.

It’s also a lot easier—infinite sample spaces are hard enough to work with without having to deal with uncountable 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 14.13.

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

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

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

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

In this case, the sample space is the infinite set

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



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

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

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. Nonnegativity is obvious, and applying the formula for the sum of a geometric series, we find that

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

Notice that this model does not have an outcome corresponding to the possibility that both players keep flipping tails forever.<sup>9</sup> That's because the probability of flipping forever would be

$$\lim_{n \rightarrow \infty} \frac{1}{2^{n+1}} = 0,$$

and outcomes with probability zero will have no impact on our calculations.

---

<sup>9</sup>In the diagram, flipping forever corresponds to following the infinite path in the tree without ever reaching a leaf or outcome. Some texts deal with this case by adding a special “infinite” sample point  $w_{\text{forever}}$  to the sample space, but we will follow the more traditional approach of excluding such sample points, as long as they collectively have probability 0.



MIT OpenCourseWare  
<http://ocw.mit.edu>

6.042J / 18.062J Mathematics for Computer Science  
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.