IV Probability

Last section

Also I should do good at since 6.041

So ce study cest of stuff

Introduction

Probability is one of the most important disciplines in all of the sciences. It is also one of the least well understood.

Probability is especially important in computer science—it arises in virtually every branch of the field. In algorithm design and game theory, for example, *randomized* algorithms and strategies (those that use a random number generator as a key input for decision making) frequently outperform deterministic algorithms and strategies. In information theory and signal processing, an understanding of randomness is critical for filtering out noise and compressing data. In cryptography and digital rights management, probability is crucial for achieving security. The list of examples is long.

Given the impact that probability has on computer science, it seems strange that probability should be so misunderstood by so many. Perhaps the trouble is that basic human intuition is wrong as often as it is right when it comes to problems involving random events. As a consequence, many students develop a fear of probability. Indeed, we have witnessed many graduate oral exams where a student will solve the most horrendous calculation, only to then be tripped up by the simplest probability question. Indeed, even some faculty will start squirming if you ask them a question that starts "What is the probability that. . . ?"

last chap basically prob I think

Our goal in the remaining chapters is to equip you with the tools that will enable you to solve basic problems involving probability easily and confidently.

Chapter 16 introduces the basic definitions and an elementary 4-step process that can be used to determine the probability that a specified event occurs. We illustrate the method on two famous problems where your intuition will probably fail you. The key concepts of Conditional probability and independence are introduced, along with examples of their use, and regrettable misuse, in practice: the probability you have a disease given that a diagnostic test says you do, and the probability

6.01

that a suspect is guilty given that his blood type matches the blood found at the scene of the crime.

We study random variables and their probability distributions in the following Chapter. Random variables provide a more quantitative way to measure random events. For example, instead of determining the probability that it will rain, we may want to determine *how much* or *how long* it is likely to rain. The fundamental concept of the *expected value* of a random variable is introduced and some of its key properties are developed.

After that, we examine the probability that a random variable deviates significantly from its expected value. This is especially important in practice, where things are generally fine if they are going according to expectation, and you would like to be assured that the probability of deviating from the expectation is very low.

We conclude with final chapter that applies the previous results to solve problems involving more complex random processes. We will see why you will probably never get very far ahead at the casino, and how two Stanford graduate students became gazillionaires by combining graph theory and probability theory to design a better search engine for the web.

16 Events and Probability Spaces

16.1 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

Monty Hall

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.

16.1.1 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.

goats are wrong door

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.

gets extra info

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?"

16.2 The Four Step Method

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

Convert math

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.

16.2.1 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.

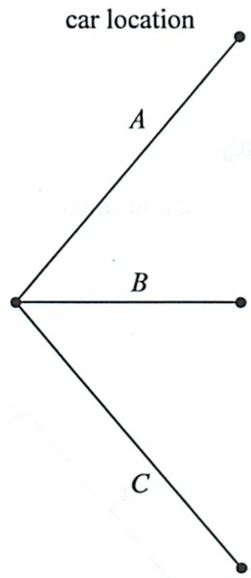


Figure 16.1 The first level in a tree diagram for the Monty Hall Problem. The branches correspond to the door behind which the car is located.

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.

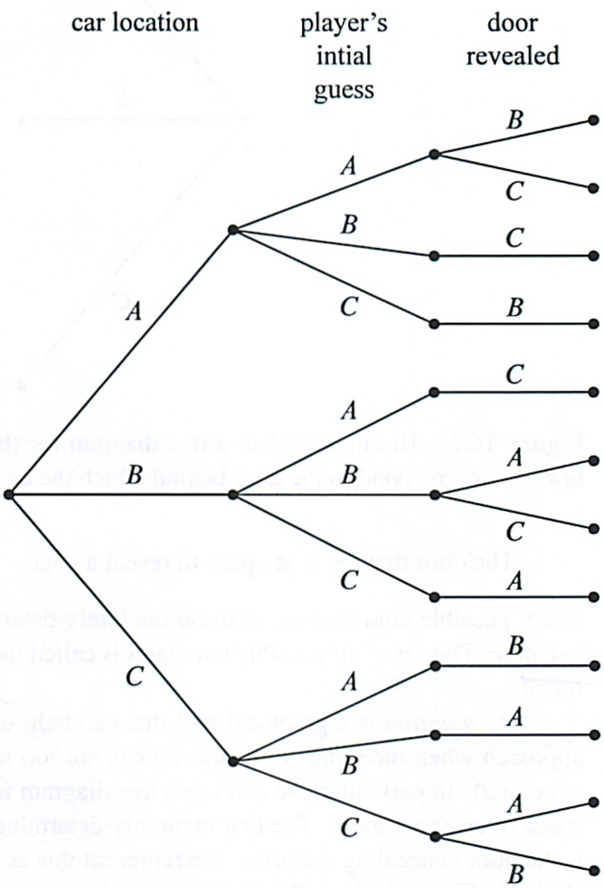
Ω

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 16.1. 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.

I spent a lot of time on this last semester

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 16.2.

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 figuring out which things to branch was very difficult



then add up branches

Figure 16.2 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 16.3 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.

16.2.2 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:

$$[\text{switching-wins}] ::= \{(A,B,C),(A,C,B),(B,A,C),(B,C,A),(C,A,B),(C,B,A)\}. \quad (16.1)$$

These outcomes have check marks in Figure 16.4.

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.

partial prob

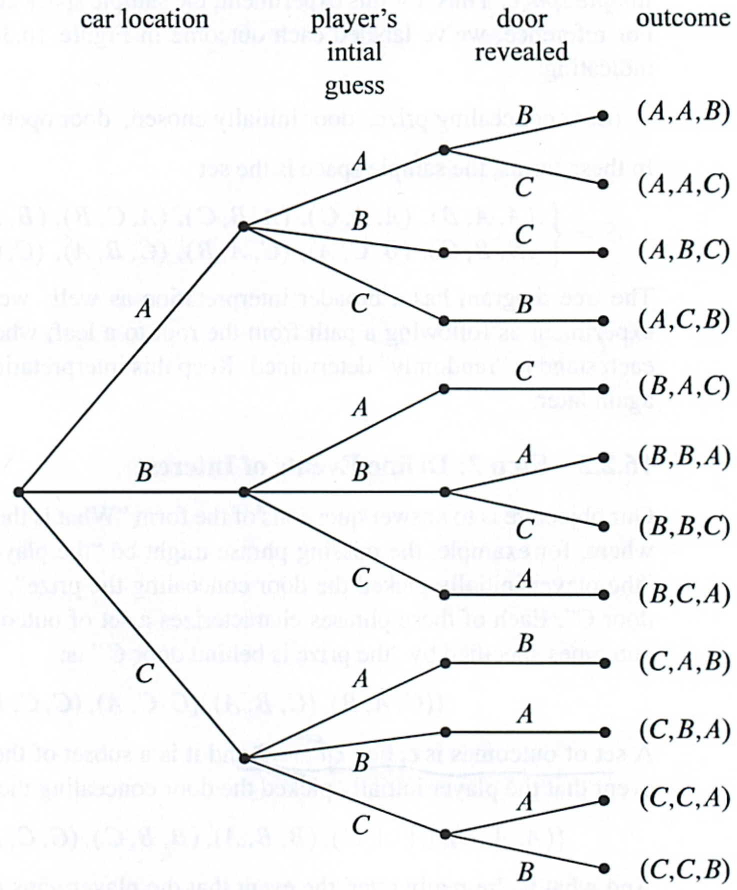


Figure 16.3 The tree diagram for the Monty Hal Problem with the outcomes labeled for each path from root to leaf. For example, outcome (A, A, B) corresponds to the car being behind door A, the player initially choosing door A, and Monty Hall revealing the goat behind door B.

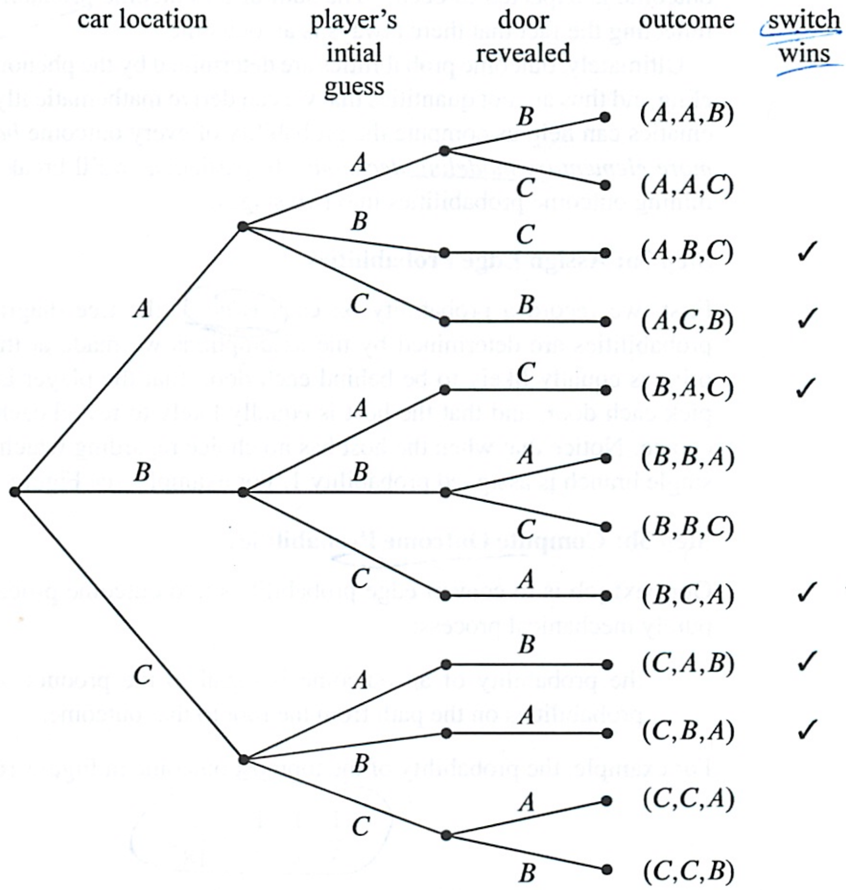


Figure 16.4 The tree diagram for the Monty Hall Problem where the outcomes in the event where the player wins by switching are denoted with a check mark.

16.2.3 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 16.5.

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 16.5, (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.

We have illustrated all of the outcome probabilities in Figure 16.5.

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:

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

16.2.4 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 (16.1) is

add it up

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

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.

16.2.5 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*

→ p 523

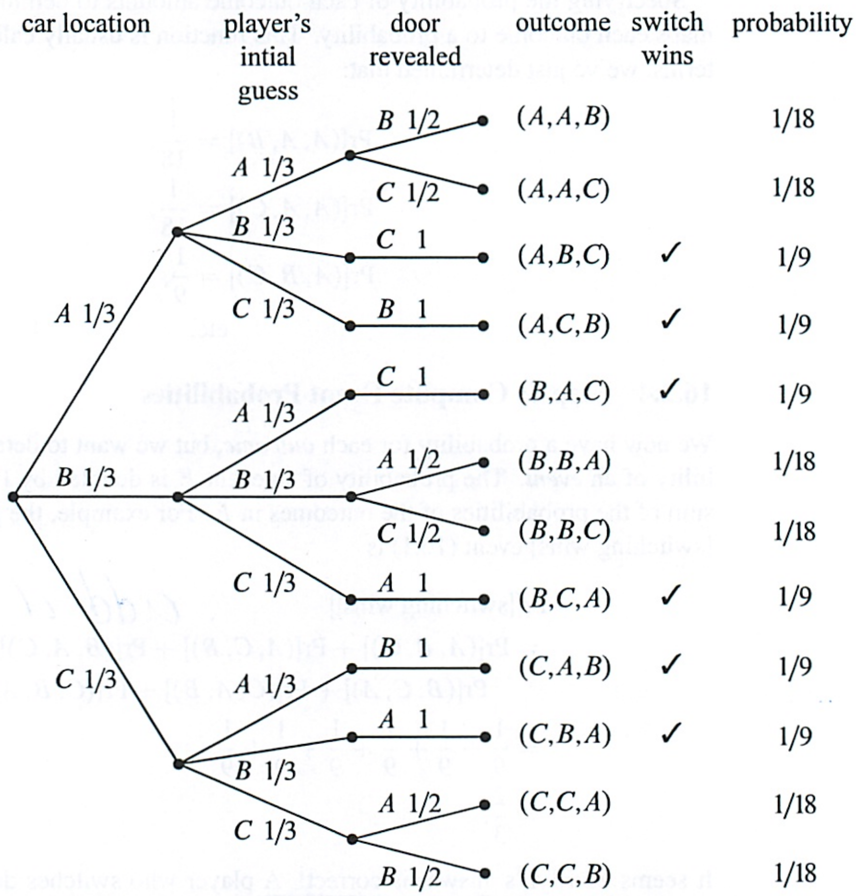
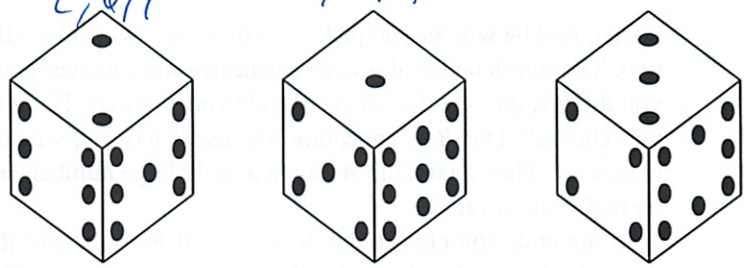


Figure 16.5 The tree diagram for the Monty Hall Problem where edge weights denote the probability of that branch being taken given that we are at the parent of that branch. For example, if the car is behind door A, then there is a 1/3 chance that the player’s initial selection is door B. The rightmost 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.

2,6,7 1,5,9 3,4,8



a *b* *c*

Figure 16.6 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 $1/3$.

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!

Wait

it opened initial door picked · would see its empty and always want to switch

16.3 Strange Dice ~~Oh I see - if you were rig~~

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. 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 16.6.

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.

16.3.1 Die A versus Die B

Step 1: Find the sample space.

The tree diagram for this scenario is shown in Figure 16.7. In particular, the sample space for this experiment are the nine pairs of values that might be rolled with Die A and Die B:

all possible outcomes

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: where A wins

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

These outcomes are marked A in the tree diagram in Figure 16.7.

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 16.7. tree prob

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

add up all tree probs where A wins

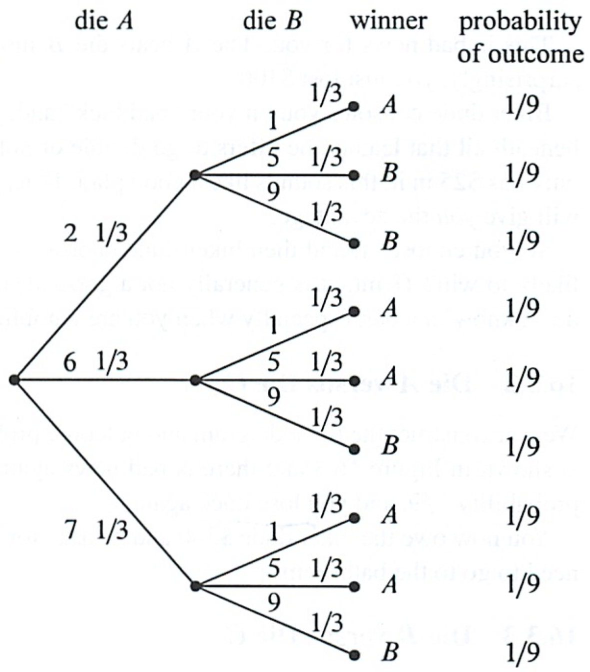


Figure 16.7 The tree diagram for one roll of die A versus die B. Die A wins with probability $5/9$.

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 $\mathcal{S}$,

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|}. \tag{16.2}$$

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.)

16.3.2 Die A versus Die C

We can construct the three diagram and outcome probabilities as before. The result is shown in Figure 16.8 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.

16.3.3 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 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

[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.

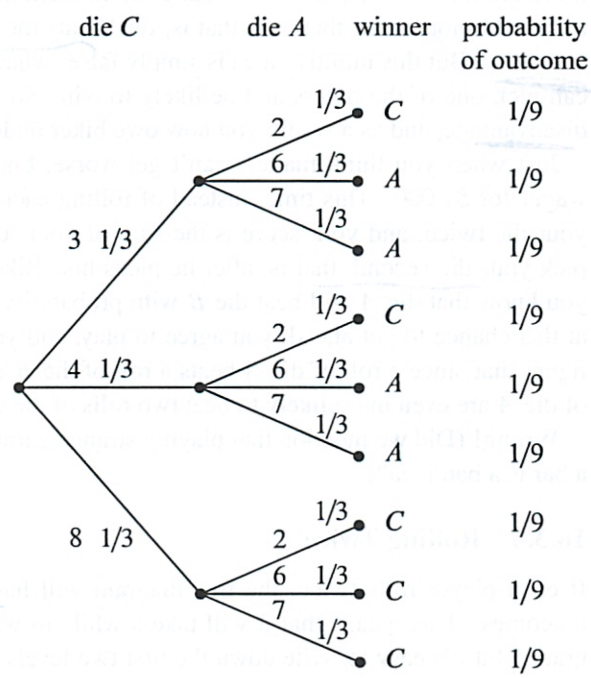


Figure 16.8 The tree diagram for one roll of die C versus die A. Die C wins with probability $5/9$.

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 big disadvantage, and as a result, you now owe biker dude \$400.

I was just about to think of this word

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?)

16.3.4 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 16.9(a) and then notice that the remaining two levels consist of nine identical copies of the tree in Figure 16.9(b).

The probability of each outcome is $(1/3)^4 = 1/81$ and so, once again, we have a uniform probability space. By Equation 16.2, 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 of the two rolls of die A is equally likely to be any element of the following multiset:

$$S_A = \{4, 8, 8, 9, 9, 12, 13, 13, 14\}.$$

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

$$S_B = \{2, 6, 6, 10, 10, 10, 14, 14, 18\}.$$

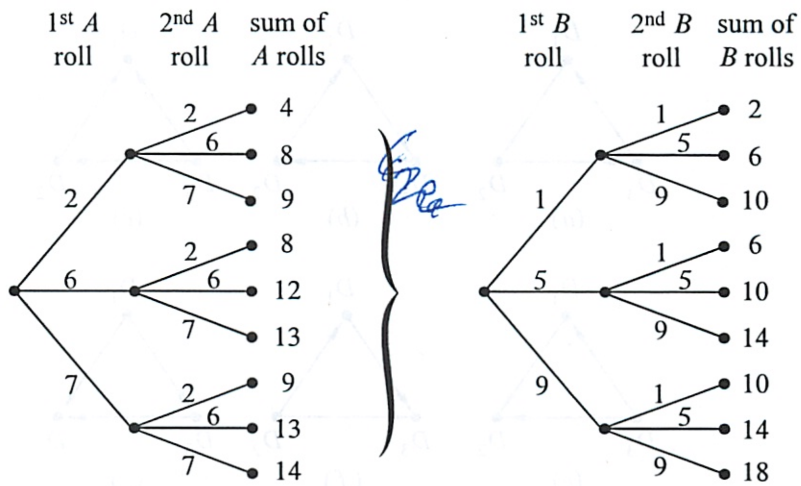


Figure 16.9 Parts of the tree diagram for die B versus die A where each die is rolled twice. The first two levels are shown in (a). The last two levels consist of nine copies of the tree in (b).

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 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 we didn't lose \$1400 to biker dude.) 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.

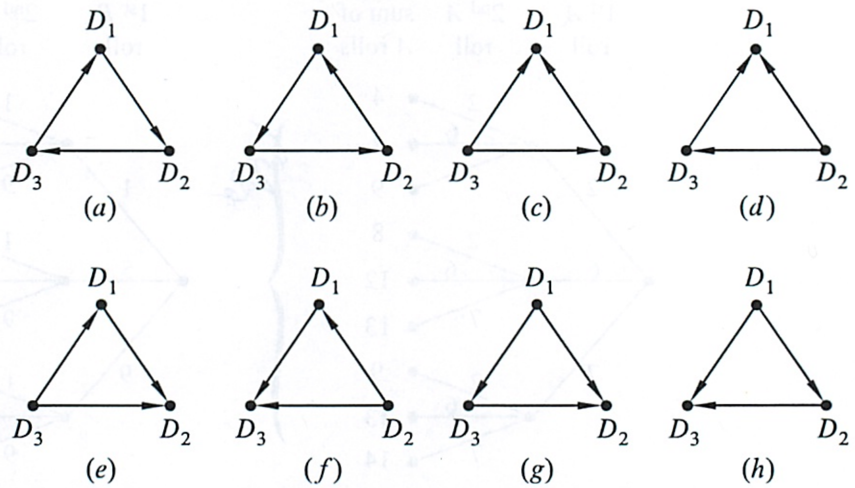


Figure 16.10 All possible relative strengths for three dice D_1, D_2, and D_3. The edge $\langle D_i \to D_j \rangle$ denotes that the sum of rolls for D_i is likely to be greater than the sum of rolls for D_j.

Even Stranger Dice

The weird behavior of the three strange dice above generalizes in a remarkable way.[2] The idea is that you can find arbitrarily large sets of dice which will beat each other in any desired pattern according to how many times the dice are rolled. prob-wise The precise statement of this result involves several alternations of universal and existential quantifiers, so it may take a few readings to understand what it is saying:

Theorem 16.3.1. *For any $n \geq 2$, there is a set of n dice with the following property: for any n-node digraph with exactly one directed edge between every two distinct nodes,[3] there is a number of rolls k such that the sum of k rolls of the ith die is bigger than the sum for the jth die with probability greater than $1/2$ iff there is an edge from the ith to the jth node in the graph.*

For example, the eight possible relative strengths for $n = 3$ dice are shown in Figure 16.10.

Our analysis for the dice in Figure 16.6 showed that for 1 roll, we have the relative strengths shown in Figure 16.10(a), and for two rolls, we have the (reverse) relative strengths shown in Figure 16.10(b). If you are prone to gambling with

[2] Reference Ron Graham paper.

[3] In other words, for every pair of nodes $u \neq v$, either $\langle u \to v \rangle$ or $\langle v \to u \rangle$, but not both, are edges of the graph. Such graphs are called *tournament graphs*, see Problem 9.4.

strangers in bars, it would be a good idea to try figuring out what other relative strengths are possible for the dice in Figure 16.6 when using more rolls.

16.4 Set Theory and Probability

Let's abstract what we've just done with the Monty Hall and strange dice examples into a general mathematical definition of sample spaces and probability.

16.4.1 Probability Spaces

4/29

Definition 16.4.1. A countable *sample space* $\mathcal{S}$ is a nonempty countable set.[4] An element $w \in \mathcal{S}$ is called an *outcome*. A subset of $\mathcal{S}$ is called an *event*. one item? No a collection of outcomes ? 1 item

Definition 16.4.2. A *probability function* on a sample space $\mathcal{S}$ is a total function $\Pr : \mathcal{S} \to \mathbb{R}$ such that

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

the 2 rules

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:

add each outcome

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

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 5.

Continuous

16.4.2 Probability Rules from Set Theory

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

[4] Yes, sample spaces can be infinite. If you did not read Chapter 5, don't worry —*countable* just means that you can list the elements of the sample space as $w_0, w_1, w_2, \ldots$.

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

$$\Pr[E \cup F] = \Pr[E] + \Pr[F].$$ Must be disjoint!

This generalizes to a countable number of events, as follows.

Rule 16.4.3 (Sum Rule). *If $\{E_0, E_1, \ldots\}$ is collection of disjoint events, then*

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

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 $\Pr[A] + \Pr[\overline{A}] = 1$, 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:

Rule 16.4.4 (Complement Rule).

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

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:

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

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

$$\Pr[E_1 \cup \cdots \cup E_n] \le \Pr[E_1] + \cdots + \Pr[E_n]. \qquad \text{(Union Bound)}$$

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 $E_1 \cup \cdots \cup E_n$ is the event that *some* critical component fails. If $\sum_{i=1}^{n} \Pr[E_i]$ is small, then the Union Bound can give an adequate upper bound on this vital probability.

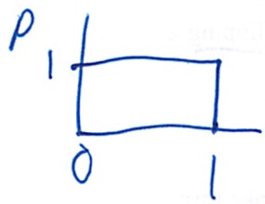
≥ 1

~~If each is different~~
oh if interrelated (not disjoint)

16.4.3 Uniform Probability Spaces

Definition 16.4.5. 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 $E \subseteq \mathcal{S}$,



$$\Pr[E] = \frac{|E|}{|\mathcal{S}|}. \tag{16.3}$$

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.

just fraction of area
? well in all cases

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 15.9.2, we know that

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

and

$$|E| = 13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$$ ← from last chap

where E is the event that we have a full house. Since every five-card hand is equally likely, we can apply Equation 16.3 to find that

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

← just do the math

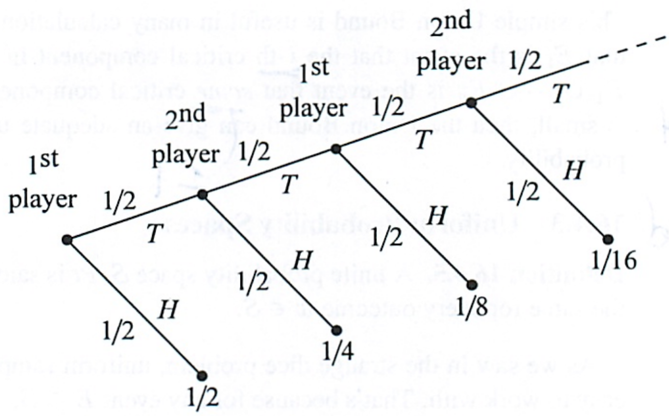


Figure 16.11 The tree diagram for the game where players take turns flipping a fair coin. The first player to flip heads wins.

16.4.4 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 16.11.

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

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

last chap ∞ sums

← note how sum is done

↑ definite form

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

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

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

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

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

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

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, w_{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 w_{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, since it has no impact, we will exclude such sample points whose probability is 0, which is the usual thing to do.

in the limit

16.5 Conditional Probability

Suppose that we pick a random person in the world. Everyone has an equal chance of being selected. Let A be the event that the person is an MIT student, and let B be the event that the person lives in Cambridge. What are the probabilities of these events? Intuitively, we're picking a random point in the big ellipse shown in Figure 16.12 and asking how likely that point is to fall into region A or B.

The vast majority of people in the world neither live in Cambridge nor are MIT students, so events A and B both have low probability. But what about the probability that a person is an MIT student, *given* that the person lives in Cambridge? This should be much greater —but what is it exactly?

What we're asking for is called a *conditional probability*; that is, the probability that one event happens, given that some other event definitely happens. Questions about conditional probabilities come up all the time:

- What is the probability that it will rain this afternoon, given that it is cloudy this morning?

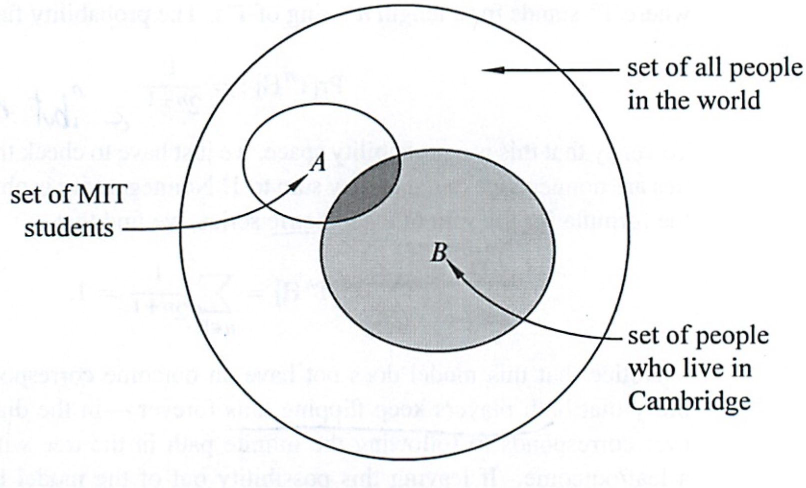


Figure 16.12 Selecting a random person. A is the event that the person is an MIT student. B is the even that the person lives in Cambridge.

- What is the probability that two rolled dice sum to 10, given that both are odd?
- What is the probability that I'll get four-of-a-kind in Texas No Limit Hold 'Em Poker, given that I'm initially dealt two queens?

There is a special notation for conditional probabilities. In general, $\Pr[A \mid B]$ denotes the probability of event A, given that event B happens. So, in our example, $\Pr[A \mid B]$ is the probability that a random person is an MIT student, given that he or she is a Cambridge resident.

How do we compute $\Pr[A \mid B]$? Since we are *given* that the person lives in Cambridge, we can forget about everyone in the world who does not. Thus, all outcomes outside event B are irrelevant. So, intuitively, $\Pr[A \mid B]$ should be the fraction of Cambridge residents that are also MIT students; that is, the answer should be the probability that the person is in set $A \cap B$ (the darkly shaded region in Figure 16.12) divided by the probability that the person is in set B (the lightly shaded region). This motivates the definition of conditional probability:

remember Critical formula ✓ intuitive explanation

Definition 16.5.1.

$$\Pr[A \mid B] ::= \frac{\Pr[A \cap B]}{\Pr[B]}$$

If $\Pr[B] = 0$, then the conditional probability $\Pr[A \mid B]$ is undefined.

Pure probability is often counterintuitive, but conditional probability is even worse! Conditioning can subtly alter probabilities and produce unexpected results in randomized algorithms and computer systems as well as in betting games. Yet, the mathematical definition of conditional probability given above is very simple and should give you no trouble —provided that you rely on mathematical reasoning and not intuition. The four-step method will also be very helpful as we will see in the next examples.

16.5.1 Using the Four-Step Method to Determine Conditional Probability

16.5.2 The "Halting Problem"

Hockey Game

The *Halting Problem* was the first example of a property that could not be tested by any program. It was introduced by Alan Turing in his seminal 1936 paper. The problem is to determine whether a Turing machine halts on a given ... yadda yadda yadda ... more importantly, it was the name of the MIT EECS department's famed C-league hockey team.

wtf?

In a best-of-three tournament, the Halting Problem wins the first game with probability $1/2$. In subsequent games, their probability of winning is determined by the outcome of the previous game. If the Halting Problem won the previous game, then they are invigorated by victory and win the current game with probability $2/3$. If they lost the previous game, then they are demoralized by defeat and win the current game with probability only $1/3$. What is the probability that the Halting Problem wins the tournament, given that they win the first game?

Make a chart! -otherwise really hard to think about it

This is a question about a conditional probability. Let A be the event that the Halting Problem wins the tournament, and let B be the event that they win the first game. Our goal is then to determine the conditional probability $\Pr[A \mid B]$.

We can tackle conditional probability questions just like ordinary probability problems: using a tree diagram and the four step method. A complete tree diagram is shown in Figure 16.13.

Step 1: Find the Sample Space

Each internal vertex in the tree diagram has two children, one corresponding to a win for the Halting Problem (labeled W) and one corresponding to a loss (labeled L). The complete sample space is:

$$\mathcal{S} = \{WW,\ WLW,\ WLL,\ LWW,\ LWL,\ LL\}.$$

Step 2: Define Events of Interest

The event that the Halting Problem wins the whole tournament is:

$$T = \{WW,\ WLW,\ LWW\}.$$

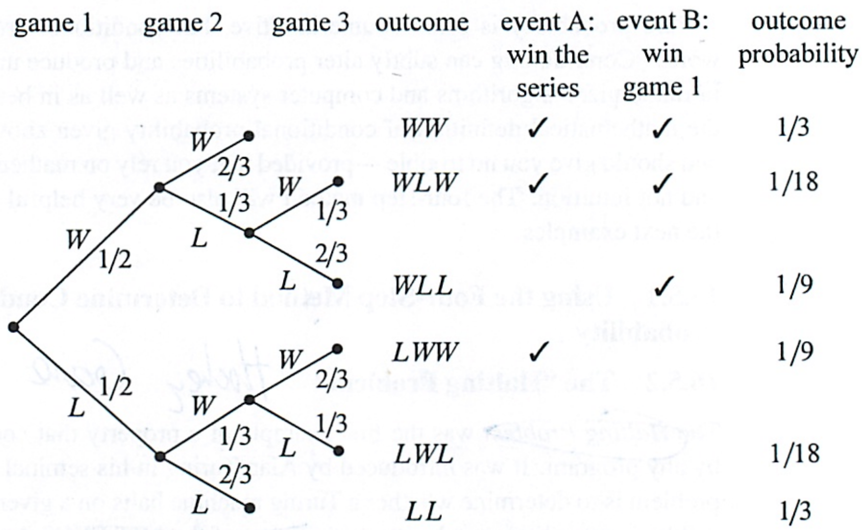


Figure 16.13 The tree diagram for computing the probability that the "Halting Problem" wins two out of three games given that they won the first game.

And the event that the Halting Problem wins the first game is:

$$F = \{WW, WLW, WLL\}.$$

The outcomes in these events are indicated with check marks in the tree diagram in Figure 16.13.

Step 3: Determine Outcome Probabilities

Next, we must assign a probability to each outcome. We begin by labeling edges as specified in the problem statement. Specifically, The Halting Problem has a $1/2$ chance of winning the first game, so the two edges leaving the root are each assigned probability $1/2$. Other edges are labeled $1/3$ or $2/3$ based on the outcome of the preceding game. We then find the probability of each outcome by multiplying all probabilities along the corresponding root-to-leaf path. For example, the probability of outcome WLL is:

$$\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{9}.$$

along each branch

Step 4: Compute Event Probabilities

We can now compute the probability that The Halting Problem wins the tournament, given that they win the first game:

$$\begin{aligned}
\Pr[A \mid B] &= \frac{\Pr[A \cap B]}{\Pr[B]} \\
&= \frac{\Pr[\{WW, WLW\}]}{\Pr[\{WW, WLW, WLL\}]} \\
&= \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} \\
&= \frac{7}{9}.
\end{aligned}$$

We're done! If the Halting Problem wins the first game, then they win the whole tournament with probability $7/9$.

16.5.3 Why Tree Diagrams Work

We've now settled into a routine of solving probability problems using tree diagrams. But we've left a big question unaddressed: what is the mathematical justification behind those funny little pictures? Why do they work?

The answer involves conditional probabilities. In fact, the probabilities that we've been recording on the edges of tree diagrams *are* conditional probabilities. For example, consider the uppermost path in the tree diagram for the Halting Problem, which corresponds to the outcome WW. The first edge is labeled $1/2$, which is the probability that the Halting Problem wins the first game. The second edge is labeled $2/3$, which is the probability that the Halting Problem wins the second game, *given* that they won the first—that's a conditional probability! More generally, on each edge of a tree diagram, we record the probability that the experiment proceeds along that path, given that it reaches the parent vertex.

So we've been using conditional probabilities all along. But why can we multiply edge probabilities to get outcome probabilities? For example, we concluded that:

$$\Pr[WW] = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Why is this correct?

now break it down + prove

The answer goes back to Definition 16.5.1 of conditional probability which could be written in a form called the *Product Rule* for probabilities:

Rule (Product Rule: 2 Events). *If* $\Pr[E_1] \neq 0$, *then:*

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2 \mid E_1].$$

Multiplying edge probabilities in a tree diagram amounts to evaluating the right side of this equation. For example:

$$\begin{aligned}
&\Pr[\text{win first game} \cap \text{win second game}] \\
&\qquad = \Pr[\text{win first game}] \cdot \Pr\left[\text{win second game} \mid \text{win first game}\right] \\
&\qquad = \frac{1}{2} \cdot \frac{2}{3}.
\end{aligned}$$

probabilities to get outcome probabilities! Of course to justify multiplying edge probabilities along longer paths, we need a Product Rule for n events.

Rule (Product Rule: n Events).

$$\Pr[E_1 \cap E_2 \cap \ldots \cap E_n] = \Pr[E_1] \cdot \Pr\left[E_2 \mid E_1\right] \cdot \Pr\left[E_3 \mid E_1 \cap E_2\right] \cdots \cdot \Pr\left[E_n \mid E_1 \cap E_2 \cap \ldots \cap E_{n-1}\right]$$

provided that

$$\Pr[E_1 \cap E_2 \cap \cdots \cap E_{n-1}] \neq 0.$$

This rule follows by routine induction from the definition of conditional probability.

6.01 covered this too

16.5.4 Medical Testing

There is an unpleasant condition called *BO* suffered by 10% of the population. There are no prior symptoms; victims just suddenly start to stink. Fortunately, there is a test for latent *BO* before things start to smell. The test is not perfect, however:

- If you have the condition, there is a 10% chance that the test will say you do not have it. These are called "false negatives."
- If you do not have the condition, there is a 30% chance that the test will say you do. These are "false positives."

Suppose a random person is tested for latent *BO*. If the test is positive, then what is the probability that the person has the condition?

Step 1: Find the Sample Space

The sample space is found with the tree diagram in Figure 16.14.

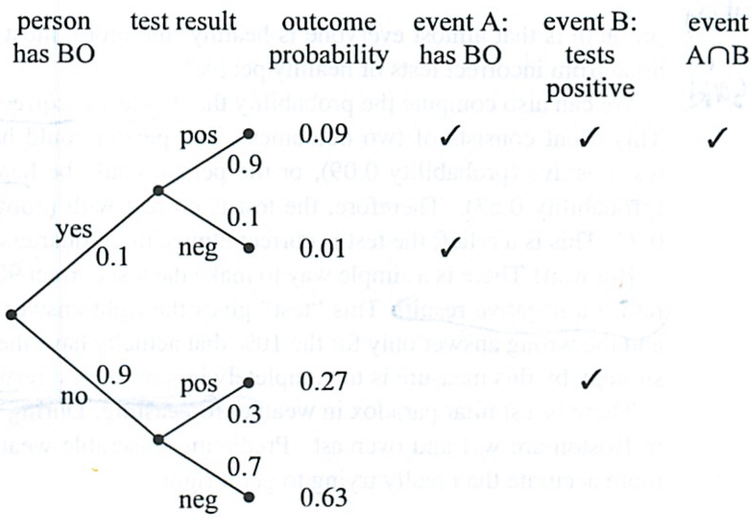


Figure 16.14 The tree diagram for the BO problem.

Step 2: Define Events of Interest

Let A be the event that the person has *BO*. Let B be the event that the test was positive. The outcomes in each event are marked in the tree diagram. We want to find $\Pr[A \mid B]$, the probability that a person has *BO*, given that the test was positive.

Step 3: Find Outcome Probabilities

First, we assign probabilities to edges. These probabilities are drawn directly from the problem statement. By the Product Rule, the probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path. All probabilities are shown in Figure 16.14.

Step 4: Compute Event Probabilities

From Definition 16.5.1, we have

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{0.09}{0.09 + 0.27} = \frac{1}{4}.$$

So, if you test positive, then there is only a 25% chance that you have the condition! This answer is initially surprising, but makes sense on reflection. There are two ways you could test positive. First, it could be that you have the condition and the test is correct. Second, it could be that you are healthy and the test is incorrect. The

Pay close attention to what is said

-false (+)

-false (−)

problem is that almost everyone is healthy; therefore, most of the positive results arise from incorrect tests of healthy people!

We can also compute the probability that the test is correct for a random person. This event consists of two outcomes. The person could have the condition and test positive (probability 0.09), or the person could be healthy and test negative (probability 0.63). Therefore, the test is correct with probability $0.09 + 0.63 = 0.72$. This is a relief; the test is correct almost three-quarters of the time.

But wait! There is a simple way to make the test correct 90% of the time: always return a negative result! This "test" gives the right answer for all healthy people and the wrong answer only for the 10% that actually have the condition. So a better strategy by this measure is to completely ignore the test result! ← since no result at all more accurate

There is a similar paradox in weather forecasting. During winter, almost all days in Boston are wet and overcast. Predicting miserable weather every day may be more accurate than really trying to get it right!

16.5.5 *A Posteriori* Probabilities

If you think about it too much, the medical testing problem we just considered could start to trouble you. The concern would be that by the time you take the test, you either have the BO condition or you don't—you just don't know which it is. So you may wonder if a statement like "If you tested positive, then you have the condition with probability 25%" makes sense.

In fact, such a statement does make sense. It means that 25% of the people who test positive actually have the condition. It is true that any particular person has it or they don't, but a *randomly selected* person among those who test positive will have the condition with probability 25%.

Anyway, if the medical testing example bothers you, you will definitely be worried by the following examples, which go even further down this path.

16.5.6 The "Halting Problem," in Reverse

Hockey Game

Suppose that we turn the hockey question around: what is the probability that the Halting Problem won their first game, given that they won the series?

This seems like an absurd question! After all, if the Halting Problem won the series, then the winner of the first game has already been determined. Therefore, who won the first game is a question of fact, not a question of probability. However, our mathematical theory of probability contains no notion of one event preceding another—there is no notion of time at all. Therefore, from a mathematical perspective, this is a perfectly valid question. And this is also a meaningful question from a practical perspective. Suppose that you're told that the Halting Problem won the series, but not told the results of individual games. Then, from your perspective, it

or during the transmission
an error burst
corrupted the bit ---

makes perfect sense to wonder how likely it is that The Halting Problem won the first game.

Thought it meant something else ...

A conditional probability $\Pr[B \mid A]$ is called *a posteriori* if event B precedes event A in time. Here are some other examples of a posteriori probabilities:

- The probability it was cloudy this morning, given that it rained in the afternoon.
- The probability that I was initially dealt two queens in Texas No Limit Hold 'Em poker, given that I eventually got four-of-a-kind.

Mathematically, a posteriori probabilities are *no different* from ordinary probabilities; the distinction is only at a higher, philosophical level. Our only reason for drawing attention to them is to say, "Don't let them rattle you."

Let's return to the original problem. The probability that the Halting Problem won their first game, given that they won the series is $\Pr[B \mid A]$. We can compute this using the definition of conditional probability and the tree diagram in Figure 16.13:

$$\Pr[B \mid A] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} = \frac{7}{9}.$$

This answer is suspicious! In the preceding section, we showed that $\Pr[A \mid B]$ was also $7/9$. Could it be true that $\Pr[A \mid B] = \Pr[B \mid A]$ in general? Some reflection suggests this is unlikely. For example, the probability that I feel uneasy, given that I was abducted by aliens, is pretty large. But the probability that I was abducted by aliens, given that I feel uneasy, is rather small.

NO!

Let's work out the general conditions under which $\Pr[A \mid B] = \Pr[B \mid A]$. By the definition of conditional probability, this equation holds if an only if:

$$\frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\Pr[A \cap B]}{\Pr[A]}$$

This equation, in turn, holds only if the denominators are equal or the numerator is 0; namely if

$$\Pr[B] = \Pr[A] \qquad \text{or} \qquad \Pr[A \cap B] = 0.$$

The former condition holds in the hockey example; the probability that the Halting Problem wins the series (event A) is equal to the probability that it wins the first game (event B) since both probabilities are $1/2$.

In general, such pairs of probabilities are related by Bayes' Rule:

Oh leads to:

All review from before

Theorem 16.5.2 (Bayes' Rule). *If* $\Pr[A]$ *and* $\Pr[B]$ *are nonzero, then:*

$$\Pr\left[B \mid A\right] = \frac{\Pr\left[A \mid B\right] \cdot \Pr[B]}{\Pr[A]} \tag{16.4}$$

Proof. When $\Pr[A]$ and $\Pr[B]$ are nonzero, we have

$$\Pr\left[A \mid B\right] \cdot \Pr[B] = \Pr[A \cap B] = \Pr\left[B \mid A\right] \cdot \Pr[A]$$

by definition of conditional probability. Dividing by $\Pr[A]$ gives (16.4). ■

16.5.7 The Law of Total Probability

Breaking a probability calculation into cases simplifies many problems. The idea is to calculate the probability of an event A by splitting into two cases based on whether or not another event E occurs. That is, calculate the probability of $A \cap E$ and $A \cap \overline{E}$. By the Sum Rule, the sum of these probabilities equals $\Pr[A]$. Expressing the intersection probabilities as conditional probabilities yields:

Rule 16.5.3 (Law of Total Probability, single event). *If* $\Pr[E]$ *and* $\Pr[\overline{E}]$ *are nonzero, then*

$$\Pr[A] = \Pr\left[A \mid E\right] \cdot \Pr[E] + \Pr\left[A \mid \overline{E}\right] \cdot \Pr[\overline{E}].$$

For example, suppose we conduct the following experiment. First, we flip a fair coin. If heads comes up, then we roll one die and take the result. If tails comes up, then we roll two dice and take the sum of the two results. What is the probability that this process yields a 2? Let E be the event that the coin comes up heads, and let A be the event that we get a 2 overall. Assuming that the coin is fair, $\Pr[E] = \Pr[\overline{E}] = 1/2$. There are now two cases. If we flip heads, then we roll a 2 on a single die with probability $\Pr\left[A \mid E\right] = 1/6$. On the other hand, if we flip tails, then we get a sum of 2 on two dice with probability $\Pr\left[A \mid \overline{E}\right] = 1/36$. Therefore, the probability that the whole process yields a 2 is

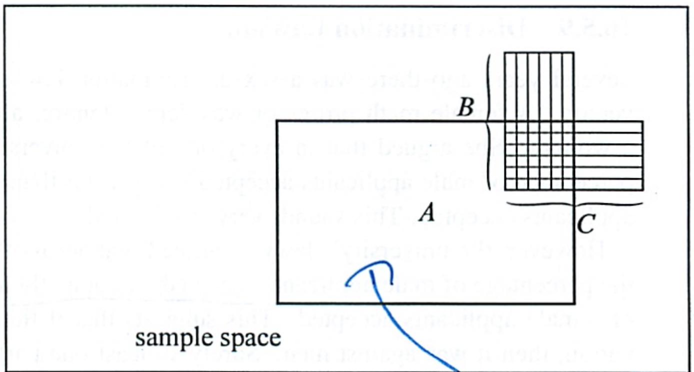
$$\Pr[A] = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{36} = \frac{7}{72}.$$

There is also a form of the rule to handle more than two cases.

Rule 16.5.4 (Law of Total Probability). *If* $E_1, \ldots, E_n$ *are disjoint events whose union is the whole sample space, then:*

$$\Pr[A] = \sum_{i=1}^{n} \Pr\left[A \mid E_i\right] \cdot \Pr[E_i].$$

add up all disjoint events



large

Figure 16.15 A counterexample to Equation 16.5. Event A is the gray rectangle, event B is the rectangle with vertical stripes, and event C is the rectangle with horizontal stripes. $B \cap C$ lies entirely within A while $B - C$ and $C - B$ are entirely outside of A.

¿what are the stripes

16.5.8 Conditioning on a Single Event

The probability rules that we derived in Chapter 16 extend to probabilities conditioned on the same event. For example, the Inclusion-Exclusion formula for two sets holds when all probabilities are conditioned on an event C:

$$\Pr[A \cup B \mid C] = \Pr[A \mid C] + \Pr[B \mid C] - \Pr[A \cap B \mid C].$$

just extend

This is easy to verify by plugging in the definition of conditional probability (16.5.1).[5]

It is important not to mix up events before and after the conditioning bar. For example, the following is *not* a valid identity:

False Claim.

No!

$$\Pr[A \mid B \cup C] = \Pr[A \mid B] + \Pr[A \mid C] - \Pr[A \mid B \cap C]. \tag{16.5}$$

A counterexample is shown in Figure 16.15. In this case, $\Pr[A \mid B] = 1/2$, $\Pr[A \mid C] = 1/2$, $\Pr[A \mid B \cap C] = 1$, and $\Pr[A \mid B \cup C] = 1/3$. However, since $1/3 \neq 1/2 + 1/2 - 1$, Equation 16.5 does not hold.

So you're convinced that this equation is false in general, right? Let's see if you *really* believe that.

[5] Problem 16.11 explains why this and similar conditional identities follow on general principles from the corresponding unconditional identifies.

16.5.9 Discrimination Lawsuit

Several years ago there was a sex discrimination lawsuit against a famous university. A female math professor was denied tenure, allegedly because she was a woman. She argued that in every one of the university's 22 departments, the percentage of male applicants accepted was greater than the percentage of female applicants accepted. This sounds very suspicious!

What I talked about in 15.279 somewhat

However, the university's lawyers argued that across the university as a whole, the percentage of male applicants accepted was actually *lower* than the percentage of female applicants accepted. This suggests that if there was any sex discrimination, then it was against men! Surely, at least one party in the dispute must be lying.

Let's simplify the problem and express both arguments in terms of conditional probabilities. To simplify matters, suppose that there are only two departments, EE and CS, and consider the experiment where we pick a random applicant. Define the following events:

- Let A be the event that the applicant is accepted.
- Let F_{EE} the event that the applicant is a female applying to EE.
- Let F_{CS} the event that the applicant is a female applying to CS.
- Let M_{EE} the event that the applicant is a male applying to EE.
- Let M_{CS} the event that the applicant is a male applying to CS.

Assume that all applicants are either male or female, and that no applicant applied to both departments. That is, the events F_{EE}, F_{CS}, M_{EE}, and M_{CS} are all disjoint.

In these terms, the plaintiff is making the following argument:

$$\Pr[A \mid F_{EE}] < \Pr[A \mid M_{EE}] \quad \text{and}$$
$$\Pr[A \mid F_{CS}] < \Pr[A \mid M_{CS}].$$

That is, in both departments, the probability that a woman is accepted for tenure is less than the probability that a man is accepted. The university retorts that overall, a woman applicant is *more* likely to be accepted than a man; namely that

$$\Pr[A \mid F_{EE} \cup F_{CS}] > \Pr[A \mid M_{EE} \cup M_{CS}].$$

It is easy to believe that these two positions are contradictory. In fact, we might even try to prove this by adding the plaintiff's two inequalities and then arguing as

CS	0 females accepted, 1 applied	0%
	50 males accepted, 100 applied	50%
EE	70 females accepted, 100 applied	70%
	1 male accepted, 1 applied	100%
Overall	70 females accepted, 101 applied	$\approx 70\%$
	51 males accepted, 101 applied	$\approx 51\%$

Table 16.1 A scenario where females are less likely to be admitted than males in each department, but more likely to be admitted overall.

follows:

write in a diff way

$$\begin{aligned} & \Pr[A \mid F_{EE}] + \Pr[A \mid F_{CS}] < \Pr[A \mid M_{EE}] + \Pr[A \mid M_{CS}] \\ \Rightarrow \quad & \Pr[A \mid F_{EE} \cup F_{CS}] < \Pr[A \mid M_{EE} \cup M_{CS}]. \end{aligned}$$

rearrange

?

The second line exactly contradicts the university's position! But there is a big problem with this argument; the second inequality follows from the first only if we accept the false identity (16.5). This argument is bogus! Maybe the two parties do not hold contradictory positions after all!

In fact, Table 16.1 shows a set of application statistics for which the assertions of both the plaintiff and the university hold. In this case, a higher percentage of males were accepted in both departments, but overall a higher percentage of females were accepted! Bizarre!

¿ So who was right ?

16.6 Independence

Suppose that we flip two fair coins simultaneously on opposite sides of a room. Intuitively, the way one coin lands does not affect the way the other coin lands. The mathematical concept that captures this intuition is called *independence*.

Definition 16.6.1. An event with probability 0 is defined to be independent of every event (including itself). If $\Pr[B] \neq 0$, then event A is independent of event B iff

$$\Pr[A \mid B] = \Pr[A]. \tag{16.6}$$

B has independent

In other words, A and B are independent if knowing that B happens does not alter the probability that A happens, as is the case with flipping two coins on opposite sides of a room.

B has nothing to do w/A

16.6.1 Potential Pitfall

Students sometimes get the idea that disjoint events are independent. The *opposite* is true: if $A \cap B = \emptyset$, then knowing that A happens means you know that B does not happen. So disjoint events are *never* independent—unless one of them has probability zero.

16.6.2 Alternative Formulation

Sometimes it is useful to express independence in an alternate form which follows immediately from Definition 16.6.1:

Theorem 16.6.2. *A is independent of B if and only if*

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]. \tag{16.7}$$

no extra info

Notice that Theorem 16.6.2 makes apparent the symmetry between A being independent of B and B being independent of A:

Corollary 16.6.3. *A is independent of B iff B is independent of A.*

16.6.3 Independence Is an Assumption

Generally, independence is something that you *assume* in modeling a phenomenon. For example, consider the experiment of flipping two fair coins. Let A be the event that the first coin comes up heads, and let B be the event that the second coin is heads. If we assume that A and B are independent, then the probability that both coins come up heads is:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

In this example, the assumption of independence is reasonable. The result of one coin toss should have negligible impact on the outcome of the other coin toss. And if we were to repeat the experiment many times, we would be likely to have $A \cap B$ about 1/4 of the time.

There are, of course, many examples of events where assuming independence is *not* justified, For example, let C be the event that tomorrow is cloudy and R be the event that tomorrow is rainy. Perhaps $\Pr[C] = 1/5$ and $\Pr[R] = 1/10$ in Boston. If these events were independent, then we could conclude that the probability of a rainy, cloudy day was quite small:

$$\Pr[R \cap C] = \Pr[R] \cdot \Pr[C] = \frac{1}{5} \cdot \frac{1}{10} = \frac{1}{50}.$$

Unfortunately, these events are definitely not independent; in particular, every rainy day is cloudy. Thus, the probability of a rainy, cloudy day is actually $1/10$.

Deciding when to *assume* that events are independent is a tricky business. In practice, there are strong motivations to assume independence since many useful formulas (such as Equation 16.7) only hold if the events are independent. But you need to be careful: we'll describe several famous examples where (false) assumptions of independence led to trouble. This problem gets even trickier when there are more than two events in play.

16.6.4 Mutual Independence

We have defined what it means for two events to be independent. What if there are more than two events? For example, how can we say that the flips of n coins are all independent of one another? A set of events is said to be *mutually independent* if, the probability of each event in the set is the same no matter which of the other events has occurred. We could formalize this with conditional probabilities as in Definition 16.6.1, but we'll juno directly to the cleaner definition based on products of probabilities as in Theorem 16.6.2: (jump)

Definition 16.6.4. A set of events $E_1, E_2, \ldots, E_n$ is mutually independent iff for all subsets $S \subseteq [1, n]$,

$$\Pr\left[\bigcap_{j \in S} E_j\right] = \prod_{j \in S} \Pr[E_j].$$

Compare whole set

Definition 16.6.4 says that $E_1, E_2, \ldots, E_n$ are mutually independent if and only if all of the following equations hold for all distinct i, j, k, and l:

$$\begin{aligned}
\Pr[E_i \cap E_j] &= \Pr[E_i] \cdot \Pr[E_j] \\
\Pr[E_i \cap E_j \cap E_k] &= \Pr[E_i] \cdot \Pr[E_j] \cdot \Pr[E_k] \\
\Pr[E_i \cap E_j \cap E_k \cap E_l] &= \Pr[E_i] \cdot \Pr[E_j] \cdot \Pr[E_k] \cdot \Pr[E_l] \\
&\vdots \\
\Pr[E_1 \cap \cdots \cap E_n] &= \Pr[E_1] \cdots \Pr[E_n].
\end{aligned}$$

For example, if we toss n fair coins, the tosses are mutually independent iff for every subset of m coins, the probability that every coin in the subset comes up heads is 2^{-m}.

16.6.5 DNA Testing

Assumptions about independence are routinely made in practice. Frequently, such assumptions are quite reasonable. Sometimes, however, the reasonableness of an independence assumption is not so clear, and the consequences of a faulty assumption can be severe.

For example, consider the following testimony from the O. J. Simpson murder trial on May 15, 1995:

Mr. Clarke: When you make these estimations of frequency—and I believe you touched a little bit on a concept called independence?

Dr. Cotton: Yes, I did.

Mr. Clarke: And what is that again?

Dr. Cotton: It means whether or not you inherit one allele that you have is not—does not affect the second allele that you might get. That is, if you inherit a band at 5,000 base pairs, that doesn't mean you'll automatically or with some probability inherit one at 6,000. What you inherit from one parent is what you inherit from the other.

Mr. Clarke: Why is that important?

Dr. Cotton: Mathematically that's important because if that were not the case, it would be improper to multiply the frequencies between the different genetic locations.

Mr. Clarke: How do you—well, first of all, are these markers independent that you've described in your testing in this case?

Presumably, this dialogue was as confusing to you as it was for the jury. Essentially, the jury was told that genetic markers in blood found at the crime scene matched Simpson's. Furthermore, they were told that the probability that the markers would be found in a randomly-selected person was at most 1 in 170 million. This astronomical figure was derived from statistics such as:

- 1 person in 100 has marker A.
- 1 person in 50 marker B.
- 1 person in 40 has marker C.
- 1 person in 5 has marker D.
- 1 person in 170 has marker E.

Then these numbers were multiplied to give the probability that a randomly-selected person would have all five markers:

$$\begin{aligned}\Pr[A \cap B \cap C \cap D \cap E] &= \Pr[A] \cdot \Pr[B] \cdot \Pr[C] \cdot \Pr[D] \cdot \Pr[E] \\ &= \frac{1}{100} \cdot \frac{1}{50} \cdot \frac{1}{40} \cdot \frac{1}{5} \cdot \frac{1}{170} \\ &= \frac{1}{170{,}000{,}000}.\end{aligned}$$

The defense pointed out that this assumes that the markers appear mutually independently. Furthermore, all the statistics were based on just a few hundred blood samples.

5/1

After the trial, the jury was widely mocked for failing to "understand" the DNA evidence. If you were a juror, would *you* accept the 1 in 170 million calculation?

oh yeah pointless

16.6.6 Pairwise Independence

The definition of mutual independence seems awfully complicated—there are so many subsets of events to consider! Here's an example that illustrates the subtlety of independence when more than two events are involved. Suppose that we flip three fair, mutually-independent coins. Define the following events:

- A_1 is the event that coin 1 matches coin 2.
- A_2 is the event that coin 2 matches coin 3.
- A_3 is the event that coin 3 matches coin 1.

Are A_1, A_2, A_3 mutually independent?

was a ~~pset~~ inclass problem

The sample space for this experiment is:

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Every outcome has probability $(1/2)^3 = 1/8$ by our assumption that the coins are mutually independent.

To see if events A_1, A_2, and A_3 are mutually independent, we must check a sequence of equalities. It will be helpful first to compute the probability of each event A_i:

$$\begin{aligned}\Pr[A_1] &= \Pr[HHH] + \Pr[HHT] + \Pr[TTH] + \Pr[TTT] \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{1}{2}.\end{aligned}$$

By symmetry, $\Pr[A_2] = \Pr[A_3] = 1/2$ as well. Now we can begin checking all the equalities required for mutual independence in Definition 16.6.4:

$$\begin{aligned}
\Pr[A_1 \cap A_2] &= \Pr[HHH] + \Pr[TTT] \\
&= \frac{1}{8} + \frac{1}{8} \\
&= \frac{1}{4} \\
&= \frac{1}{2} \cdot \frac{1}{2} \\
&= \Pr[A_1]\Pr[A_2].
\end{aligned}$$

By symmetry, $\Pr[A_1 \cap A_3] = \Pr[A_1] \cdot \Pr[A_3]$ and $\Pr[A_2 \cap A_3] = \Pr[A_2] \cdot \Pr[A_3]$ must hold also. Finally, we must check one last condition:

$$\begin{aligned}
\Pr[A_1 \cap A_2 \cap A_3] &= \Pr[HHH] + \Pr[TTT] \\
&= \frac{1}{8} + \frac{1}{8} \\
&= \frac{1}{4} \\
&\neq \Pr[A_1]\Pr[A_2]\Pr[A_3] = \frac{1}{8}.
\end{aligned}$$

is 2-wise ind.

The three events A_1, A_2, and A_3 are not mutually independent even though any two of them are independent! This not-quite mutual independence seems weird at first, but it happens. It even generalizes:

Definition 16.6.5. A set $A_1, A_2, \ldots,$ of events is *k-way independent* iff every set of k of these events is mutually independent. The set is *pairwise independent* iff it is 2-way independent.

So the sets A_1, A_2, A_3 above are pairwise independent, but not mutually independent. Pairwise independence is a much weaker property than mutual independence.

For example, suppose that the prosecutors in the O. J. Simpson trial were wrong and markers A, B, C, D, and E appear only *pairwise* independently. Then the probability that a randomly-selected person has all five markers is no more than:

$$\begin{aligned}\Pr[A \cap B \cap C \cap D \cap E] &\le \Pr[A \cap E] \\ &= \Pr[A] \cdot \Pr[E] \\ &= \frac{1}{100} \cdot \frac{1}{170} \\ &= \frac{1}{17{,}000}.\end{aligned}$$

The first line uses the fact that $A \cap B \cap C \cap D \cap E$ is a subset of $A \cap E$. (We picked out the A and E markers because they're the rarest.) We use pairwise independence on the second line. Now the probability of a random match is 1 in 17,000—a far cry from 1 in 170 million! And this is the strongest conclusion we can reach assuming only pairwise independence.

On the other hand, the 1 in 17,000 bound that we get by assuming pairwise independence is a lot better than the bound that we would have if there were no independence at all. For example, if the markers are dependent, then it is possible that

everyone with marker E has marker A,
everyone with marker A has marker B,
everyone with marker B has marker C, and
everyone with marker C has marker D.

In such a scenario, the probability of a match is

$$\Pr[E] = 1/170.$$

✓ The smallest restriction

? So how would you describe what the defense said?

So a stronger independence assumption leads to a smaller bound on the probability of a match. The trick is to figure out what independence assumption is reasonable. Assuming that the markers are *mutually* independent may well *not* be reasonable unless you have examined hundreds of millions of blood samples. Otherwise, how would you know that marker D does not show up more frequently whenever the other four markers are simultaneously present?

We will conclude our discussion of independence with a useful, and somewhat famous, example known as the Birthday Principle.

16.7 The Birthday Principle

oh this is fun

There are 85 students in a class. What is the probability that some birthday is shared by two people? Comparing 85 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 $1/d$, where $d = 365$ in this case. We'll also assume that a class is composed of n randomly and independently selected students, with $n = 85$ in this case. 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, the 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). Frate ''

Selecting a sequence of n students for a class yields a sequence of n birthdays. Under the assumptions above, the d^n possible birthday sequences are equally likely outcomes. Let's examin ethe consequences of this probability model by focussing on the ith and jth elements in a birthday sequence, where $1 \leq i \neq j \leq n$. It makes for a better story if we refer to the ith birthday as "Alice's" and the jth as "Bob's."

Now since Bob's birthday is assumed to be independent of Alice's, it follows that whichever of the d birthdays Alice's happens to be, the probability that Bob has the same birthday $1/d$. Next, If we look at two other birthdays —call them "Carol's" and "Don's" —then whether Alice and Bob have matching birthdays has nothing to do with whether Carol and Don have matching birthdays. That is, the event that Alice and Bob have matching birthdays is independent of the event that Carol and Don have matching birthdays. In fact, for any set of non-overlapping couples, the events that a couple has matching birthdays are mutually independent.

In fact, it's pretty clear that the probability that Alice and Bob have matching birthdays remains $1/d$ whether or not Carol and *Alice* have matching birthdays. That is, the event that Alice and Bob match is also independent of Alice and Carol matching. In short, the set of all events in which a couple has matching birthdays is *pairwise* independent, despite the overlapping couples. This will be important in a later chapter because pairwise independence will be enough to justify some conclusions about the expected number of matches. However, these matching birthday events are obviously *not* even 3-way independent: if Alice and Bob match, and also Alice and Carol match, then Bob and Carol will match. 4way?

It turns out that as long as the number of students is noticeably smaller than the number of possible birthdays, we can get a pretty good estimate of the birthday matching probabilities by *pretending* that the matching events are mutually independent. (An intuitive justification for this is that with only a small number of matching pairs, it's likely that none of the pairs overlap.) Then the probability of *no* matching birthdays would be the same as rth power of the probability that a couple does *not* have matching birthdays, where $r ::= \binom{n}{2}$ is the number of couples. That is, the probability of no matching birthdays would be

$$(1 - 1/d)^{\binom{n}{2}}. \tag{16.8}$$

Using the fact that $e^x > 1 + x$ for all x,[6] we would conclude that the probability of no matching birthdays is at most

to the?

multiplied?

$$e^{-\frac{\binom{n}{2}}{d}}. \tag{16.9}$$

The matching birthday problem fits in here so far as a nice example illustrating pairwise and mutual independence. But it's actually not hard to justify the bound (16.9) without any pretence or any explicit consideration of independence. Namely, there are $d(d-1)(d-2)\cdots(d-(n-1))$ length n sequences of distinct birthdays. So the probability that everyone has a different birthday is:

$$\begin{aligned}
&\frac{d(d-1)(d-2)\cdots(d-(n-1))}{d^n} \\
&= \frac{d}{d}\cdot\frac{d-1}{d}\cdot\frac{d-2}{d}\cdots\frac{d-(n-1)}{d} \\
&= \left(1-\frac{0}{d}\right)\left(1-\frac{1}{d}\right)\left(1-\frac{2}{d}\right)\cdots\left(1-\frac{n-1}{d}\right) \\
&< e^0\cdot e^{-1/d}\cdot e^{-2/d}\cdots e^{-(n-1)/d} \qquad \text{(since } 1+x<e^x\text{)} \\
&= e^{-\left(\sum_{i=1}^{n-1} i/d\right)} \\
&= e^{-(n(n-1)/2d)} \\
&= \text{the bound (16.9).}
\end{aligned}$$

[6] This approximation is obtained 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.

For $n = 85$ and $d = 365$, (16.9) is less than $1/17,000$, which means the probability of having some pair of matching birthdays actually is more than $1 - 1/17,000 > 0.9999$. So it would be pretty astonishing if there were no pair of students in the class with matching birthdays.

For $d \leq n^2/2$, the probability of no match turns out to be asymptotically equal to the upper bound (16.9). 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 $1 - 1/e \approx 0.632$.

So what was this again?

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 unless n^2 is a small fraction of d. The Birthday Principle also famously comes into play as the basis of "birthday attacks" that crack certain cryptographic systems.

Problems for Section 16.2

Exam Problems

Problem 16.1.

What's the probability that 0 doesn't appear among k digits chosen independently and uniformly at random?

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?

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?

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?

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

Problem 16.2.
We consider a variation of Monty Hall's game. The contestant still picks one of three doors, with a prize randomly placed behind one door and goats behind the other two. But now, instead of always opening a door to reveal a goat, Monty instructs Carol to *randomly* open one of the two doors that the contestant hasn't picked. This means she may reveal a goat, or she may reveal the prize. If she reveals the prize, then the entire game is *restarted*, that is, the prize is again randomly placed behind some door, the contestant again picks a door, and so on until Carol finally picks a door with a goat behind it. Then the contestant can choose to *stick* with his original choice of door or *switch* to the other unopened door. He wins if the prize is behind the door he finally chooses.

To analyze this setup, we define two events:

GP: The event that the contestant **g**uesses the door with the **p**rize behind it on his first guess.

OP: The event that the game is restarted at least once. Another way to describe this is as the event that the door Carol first **o**pens has a **p**rize behind it.

(a) What is $\Pr[GP]$? 0.5in. $\Pr\left[OP \mid \overline{GP}\right]$?0.5in.

(b) What is $\Pr[OP]$? 0.5in

(c) Let R be the number of times the game is restarted before Carol picks a goat.
What is $\text{Ex}[R]$?0.8in
You may express the answer as a simple closed form in terms of $p ::= \Pr[OP]$.

(d) What is the probability the game will continue forever?0.5in

(e) When Carol finally picks the goat, the contestant has the choice of sticking or switching. Let's say that the contestant adopts the strategy of sticking. Let W be the event that the contestant wins with this strategy, and let $w ::= \Pr[W]$. Express the following conditional probabilities as simple closed forms in terms of w.

i) $\Pr\left[W \mid GP\right] =$ 0.5in
ii) $\Pr\left[W \mid \overline{GP} \cap OP\right] =$ 0.5in
iii) $\Pr\left[W \mid \overline{GP} \cap \overline{OP}\right] =$ 0.5in

(f) What is $\Pr[W]$? 0.5in

(g) For any final outcome where the contestant wins with a "stick" strategy, he would lose if he had used a "switch" strategy, and vice versa. In the original Monty Hall game, we concluded immediately that the probability that he would win with a "switch" strategy was $1 - \Pr[W]$. Why isn't this conclusion quite as obvious for this new, restartable game? Is this conclusion still sound? Briefly explain.

Problem 16.3.
Graphs, Logic & Probability

Let G be an undirected simple graph with $n > 3$ vertices. Let $E(x, y)$ mean that G has an edge between vertices x and y, and let $P(x, y)$ mean that there is a length 2 path in G between x and y.

(a) Explain why $E(x, y)$ implies $P(x, x)$.

(b) Circle the mathematical formula that best expresses the definition of $P(x, y)$.

- $P(x, y) ::= \exists z.\ E(x, z) \land E(y, z)$
- $P(x, y) ::= x \neq y \land \exists z.\ E(x, z) \land E(y, z)$
- $P(x, y) ::= \forall z.\ E(x, z) \lor E(y, z)$
- $P(x, y) ::= \forall z.\ x \neq y \longrightarrow [E(x, z) \lor E(y, z)]$

For the following parts (c)–(e), let V be a fixed set of $n > 3$ vertices, and let G be a graph with these vertices constructed randomly as follows: for all distinct vertices $x, y \in V$, independently include edge $\langle x\text{—}y\rangle$ as an edge of G with probability p. In particular, $\Pr[E(x, y)] = p$ for all $x \neq y$.

(c) For distinct vertices w, x, y and z in V, circle the event pairs that are independent.

1. $E(w, x)$ versus $E(x, y)$
2. $(E(w, x) \land E(w, y))$ versus $(E(z, x) \land E(z, y))$
3. $E(x, y)$ versus $P(x, y)$

4. $P(w, x)$ versus $P(x, y)$

5. $P(w, x)$ versus $P(y, z)$

(d) Write a simple formula in terms of n and p for Pr[not $P(x, y)$], for distinct vertices x and y in V.

Hint: Use part (c), item 2.

(e) What is the probability that two distinct vertices x and y lie on a three-cycle in G? Answer with a simple expression in terms of p and r, where $r ::=$ Pr[not $P(x, y)$] is the correct answer to part (d).

Hint: Express x and y being on a three-cycle as a simple formula involving $E(x, y)$ and $P(x, y)$.

Class Problems

Problem 16.4.
[A Baseball Series]
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 16.5.
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 whole the 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?

Suggestions: 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 difficult. 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.

Problem 16.6.
[The Four-Door Deal]

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.

Use The Four Step Method of Section 16.2 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?

Problem 16.7.
[Simulating a fair coin] Suppose you need a fair coin to decide which door to choose in the 6.042 Monty Hall game. After making everyone in your group empty their pockets, all you managed to turn up is some crumpled bubble gum wrappers, a few used tissues, and one penny. However, the penny was from Prof. Meyer's pocket, so it is **not** safe to assume that it is a fair coin.

How can we use a coin of unknown bias to get the same effect as a fair coin of bias $1/2$? Draw the tree diagram for your solution, but since it is infinite, draw only enough to see a pattern.

Suggestion: A neat trick allows you to sum all the outcome probabilities that cause you to say "Heads": Let s be the sum of all "Heads" outcome probabilities in the whole tree. Notice that *you can write the sum of all the "Heads" outcome probabilities in certain subtrees as a function of* s. Use this observation to write an equation in s and then solve.

Homework Problems

Problem 16.8.
I have a deck of 52 regular playing cards, 26 red, 26 black, randomly shuffled. They all lie face down in the deck so that you can't see them. I will draw a card off the top of the deck and turn it face up so that you can see it and then put it aside. I will continue to turn up cards like this but at some point while there are still cards left in the deck, you have to declare that you want the next card in the deck to be turned up. If that next card turns up black you win and otherwise you lose. Either way, the game is then over.

(a) Show that if you take the first card before you have seen any cards, you then have probability $1/2$ of winning the game.

(b) Suppose you don't take the first card and it turns up red. Show that you have then have a probability of winning the game that is greater than $1/2$.

(c) If there are r red cards left in the deck and b black cards, show that the probability of winning in you take the next card is $b/(r+b)$.

(d) Either,

1. come up with a strategy for this game that gives you a probability of winning strictly greater than $1/2$ and prove that the strategy works, or,
2. come up with a proof that no such strategy can exist.

Problems for Section 16.4

Class Problems

Problem 16.9.
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 $1 \leq i \leq n$. 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,

$$p \le \Pr[F] \le np. \tag{16.10}$$

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 $\mathcal{S} ::= \mathcal{P}(\{1, \dots, n\})$ 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 emptyset, $\emptyset$.

(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 (16.10).

Problem 16.10.
Here are some handy rules for reasoning about probabilities that all follow directly from the Disjoint Sum Rule. Prove them.

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

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

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

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

$$\text{If } A \subseteq B, \text{ then } \Pr[A] \le \Pr[B]. \qquad \text{(Monotonicity)}$$

Problem 16.11.
Suppose $\Pr[] : \mathcal{S} \to [0, 1]$ is a probability function on a sample space, $\mathcal{S}$, and let B be an event such that $\Pr[B] > 0$. Define a function $\Pr_{\cdot}[B]$ on events outcomes $w \in \mathcal{S}$ by the rule:

$$\Pr_w[B] ::= \begin{cases} \Pr[w]/\Pr[B] & \text{if } w \in B, \\ 0 & \text{if } w \notin B. \end{cases} \tag{16.11}$$

(a) Prove that $\Pr_{\cdot}[B]$ is also a probability function on $\mathcal{S}$ according to Definition 16.4.2.

(b) Prove that

$$\Pr_A[B] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

for all $A \subseteq \mathcal{S}$.

Homework Problems

Problem 16.12.
Prove the following probabilistic identity, referred to as the **Union Bound**. You may assume the theorem that the probability of a union of *disjoint* sets is the sum of their probabilities.

Let $A_1, \ldots, A_n$ be a collection of events. Then

$$\Pr[A_1 \cup A_2 \cup \cdots \cup A_n] \le \sum_{i=1}^{n} \Pr[A_i].$$

Hint: Induction.

Problems for Section 16.5

Exam Problems

Problem 16.13.
There are two decks of cards, the red deck and the blue deck. They differ slightly in a way that makes drawing the eight of hearts slightly more likely from the red deck than from the blue deck.

One of the decks is randomly chosen and hidden in a box. You reach in the box and randomly pick a card that turns out to be the eight of hearts. You believe intuitively that this makes the red deck more likely to be in the box than the blue deck.

Your intuitive judgment about the red deck can be formalized and verified using some inequalities between probabilities and conditional probabilities involving the events

$$\begin{aligned} R &::= \textbf{R}\text{ed deck is in the box,} \\ B &::= \textbf{B}\text{lue deck is in the box,} \\ E &::= \textbf{E}\text{ight of hearts is picked from the deck in the box.} \end{aligned}$$

(a) State an inequality between probabilities and/or conditional probabilities that formalizes the assertion, "picking the eight of hearts from the red deck is more likely than from the blue deck."

(b) State a similar inequality that formalizes the assertion "picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck."

(c) Assuming the each deck is equally likely to be the one in the box, prove that the inequality of part (a) implies the inequality of part (b).

(d) Suppose you couldn't be sure that the red deck and blue deck were equally likely to be in the box. Could you still conclude that picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck? Briefly explain.

Practice Problems

Problem 16.14.

Dirty Harry places two bullets in the six-shell cylinder of his revolver. He gives the cylinder a random spin and says "Feeling lucky?" as he holds the gun against your heart.

(a) What is the probability that you will get shot if he pulls the trigger?

(b) Suppose he pulls the trigger and you don't get shot. What is the probability that you will get shot if he pulls the trigger a second time?

(c) Suppose you noticed that he placed the two shells next to each other in the cylinder. How does this change the answers to the previous two questions?

Class Problems

Problem 16.15.

There are two decks of cards. One is complete, but the other is missing the ace of spades. Suppose you pick one of the two decks with equal probability and then select a card from that deck uniformly at random. What is the probability that you picked the complete deck, given that you selected the eight of hearts? Use the four-step method and a tree diagram.

Problem 16.16.
There are three prisoners in a maximum-security prison for fictional villains: the Evil Wizard Voldemort, the Dark Lord Sauron, and Little Bunny Foo-Foo. The parole board has declared that it will release two of the three, chosen uniformly at random, but has not yet released their names. Naturally, Sauron figures that he will be released to his home in Mordor, where the shadows lie, with probability $2/3$.

A guard offers to tell Sauron the name of one of the other prisoners who will be released (either Voldemort or Foo-Foo). Sauron knows the guard to be a truthful fellow. However, Sauron declines this offer. He reasons that if the guard says, for example, "Little Bunny Foo-Foo will be released", then his own probability of release will drop to $1/2$. This is because he will then know that either he or Voldemort will also be released, and these two events are equally likely.

Using a tree diagram and the four-step method, either prove that the Dark Lord Sauron has reasoned correctly or prove that he is wrong. Assume that if the guard has a choice of naming either Voldemort or Foo-Foo (because both are to be released), then he names one of the two uniformly at random.

Homework Problems

Problem 16.17.
Outside of their hum-drum duties as Math for Computer Science Teaching Assistants, Oscar is trying to learn to levitate using only intense concentration and Liz is trying to become the world champion flaming torch juggler. Suppose that Oscar's probability of success is $1/6$, Liz's chance of success is $1/4$, and these two events are independent.

(a) If at least one of them succeeds, what is the probability that Oscar learns to levitate?

(b) If at most one of them succeeds, what is the probability that Liz becomes the world flaming torch juggler champion?

(c) If exactly one of them succeeds, what is the probability that it is Oscar?

Problem 16.18.
There is a course—not Math for Computer Science, naturally—in which 10% of the assigned problems contain errors. If you ask a Teaching Assistant (TA) whether a problem has an error, then they will answer correctly 80% of the time. This 80% accuracy holds regardless of whether or not a problem has an error. Likewise when you ask a lecturer, but with only 75% accuracy.

We formulate this as an experiment of choosing one problem randomly and asking a particular TA and Lecturer about it. Define the following events:

$$E ::= \text{"the problem has an error,"}$$
$$T ::= \text{"the TA says the problem has an error,"}$$
$$L ::= \text{"the lecturer says the problem has an error."}$$

(a) Translate the description above into a precise set of equations involving conditional probabilities among the events E, T, and L

(b) Suppose you have doubts about a problem and ask a TA about it, and they tell you that the problem is correct. To double-check, you ask a lecturer, who says that the problem has an error. Assuming that *the correctness of the lecturers' answer and the TA's answer are independent of each other, regardless of whether there is an error*[7], what is the probability that there is an error in the problem?

(c) Is the event that "the TA says that there is an error", independent of the event that "the lecturer says that there is an error"?

Problem 16.19. (a) Suppose you repeatedly flip a fair coin until you see the sequence HHT or the sequence TTH. What is the probability you will see HHT first? *Hint:* Symmetry between Heads and Tails.

(b) What is the probability you see the sequence HTT before you see the sequence HHT? *Hint:* Try to find the probability that HHT comes before HTT conditioning on whether you first toss an H or a T. The answer is not $1/2$.

Problem 16.20.
A 52-card deck is thoroughly shuffled and you are dealt a hand of 13 cards.

(a) If you have one ace, what is the probability that you have a second ace?

[7] This assumption is questionable: by and large, we would expect the lecturer and the TA's to spot the same glaring errors and to be fooled by the same subtle ones.

(b) If you have the ace of spades, what is the probability that you have a second ace?

Remarkably, the two answers are different. This problem will test your counting ability!

Problem 16.21.
You are organizing a neighborhood census and instruct your census takers to knock on doors and note the sex of any child that answers the knock. Assume that there are two children in a household and that girls and boys are equally likely to be children and to open the door.

A sample space for this experiment has outcomes that are triples whose first element is either B or G for the sex of the elder child, likewise for the second element and the sex of the younger child, and whose third coordinate is E or Y indicating whether the *e*lder child or *y*ounger child opened the door. For example, (B, G, Y) is the outcome that the elder child is a boy, the younger child is a girl, and the girl opened the door.

(a) Let T be the event that the household has two girls, and O be the event that a girl opened the door. List the outcomes in T and O.

(b) What is the probability $\Pr[T \mid O]$, that both children are girls, given that a girl opened the door?

(c) Where is the mistake in the following argument?

> If a girl opens the door, then we know that there is at least one girl in the household. The probability that there is at least one girl is
>
> $$1 - \Pr[\text{both children are boys}] = 1 - (1/2 \times 1/2) = 3/4. \tag{16.12}$$
>
> So,
>
> $$\Pr[T \mid \text{there is at least one girl in the household}] \tag{16.13}$$
> $$= \frac{\Pr[T \cap \text{there is at least one girl in the household}]}{\Pr[\text{there is at least one girl in the household}]} \tag{16.14}$$
> $$= \frac{\Pr[T]}{\Pr[\text{there is at least one girl in the household}]} \tag{16.15}$$
> $$= (1/4)/(3/4) = 1/3. \tag{16.16}$$
>
> Therefore, given that a girl opened the door, the probability that there are two girls in the household is 1/3.

Problems for Section 16.6

Exam Problems

Problem 16.22.
Sally Smart just graduated from high school. She was accepted to three top colleges.

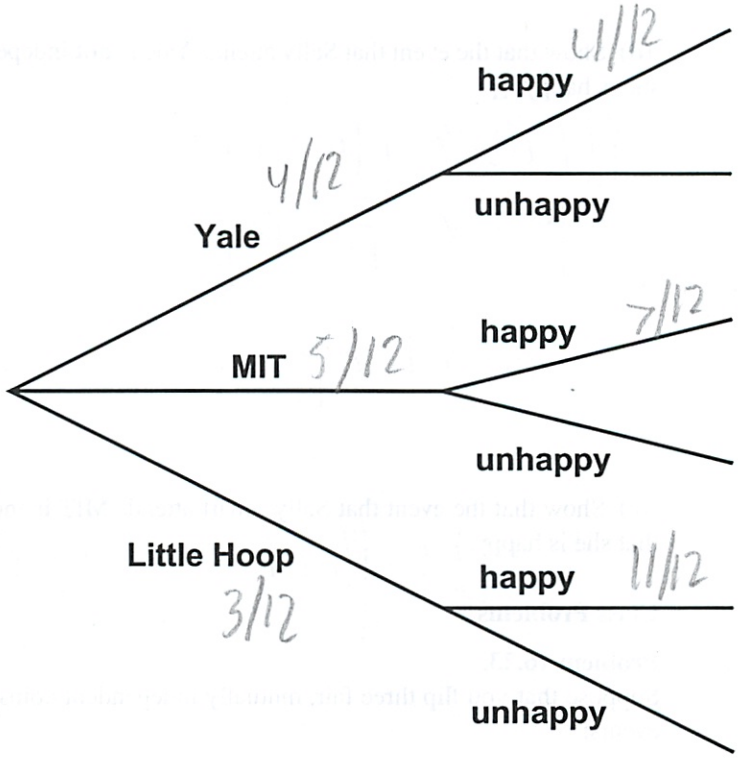
- With probability $4/12$, she attends Yale.
- With probability $5/12$, she attends MIT.
- With probability $3/12$, she attends Little Hoop Community College.

Sally will either be happy or unhappy in college.

- If she attends Yale, she is happy with probability $4/12$.
- If she attends MIT, she is happy with probability $7/12$.
- If she attends Little Hoop, she is happy with probability $11/12$.

leaves space

(a) A tree diagram for Sally's situation is shown below. On the diagram, fill in the edge probabilities and at each leaf write the probabilty of that outcome.



(b) What is the probability that Sally is happy in college? []

P Add $\frac{4}{12}\cdot\frac{4}{12}+\frac{5}{12}\cdot\frac{7}{12}+\frac{3}{12}\cdot\frac{11}{12}$

(c) What is the probability that Sally Smart attends Yale, given that she is happy in college? []

$P(B/A) = \frac{P(A\cap B)}{P(A)} = P\left(\frac{\frac{4}{12}\cdot\frac{4}{12}}{\text{prev ans}}\right.$

(what I thought)

no way to verify

(d) Show that the event that Sally attends Yale is **not** independent of the event that she is happy. []

$P(A \cap B) \neq P(A) \cdot P(B)$

$\frac{4}{12} \cdot \frac{4}{12} \neq \frac{4}{12} \cdot$ ans to b

ans to b $\neq \frac{4}{12}$

(e) Show that the event that Sally Smart attends MIT **is** independent of the event that she is happy. hmm same,

$P(A \cap B) = P(A) \cdot P(B)$

$\frac{5}{12} \cdot \frac{7}{12} = \frac{5}{12} \cdot$ ans to b

ans b $= \frac{7}{12}$? I guess

Class Problems

Problem 16.23.
Suppose that you flip three fair, mutually independent coins. Define the following events:

- Let A be the event that *the first* coin is heads.
- Let B be the event that *the second* coin is heads.
- Let C be the event that *the third* coin is heads.
- Let D be the event that *an even number of* coins are heads.

(a) Use the four step method to determine the probability space for this experiment and the probability of each of A, B, C, D.

(b) Show that these events are not mutually independent.

(c) Show that they are 3-way independent.

17 Random Variables

Thus far, we have focused on probabilities of events. For example, we computed the probability that you win the Monty Hall game or that you have a rare medical condition given that you tested positive. But, in many cases we would like to know more. For example, *how many* contestants must play the Monty Hall game until one of them finally wins? *How long* will this condition last? *How much* will I lose gambling with strange dice all night? To answer such questions, we need to work with random variables.

17.1 Random Variable Examples

Definition 17.1.1. A *random variable* R on a probability space is a total function whose domain is the sample space.

The codomain of R can be anything, but will usually be a subset of the real numbers. Notice that the name "random variable" is a misnomer; random variables are actually functions!

For example, suppose we toss three independent, unbiased coins. Let C be the number of heads that appear. Let $M = 1$ if the three coins come up all heads or all tails, and let $M = 0$ otherwise. Now every outcome of the three coin flips uniquely determines the values of C and M. For example, if we flip heads, tails, heads, then $C = 2$ and $M = 0$. If we flip tails, tails, tails, then $C = 0$ and $M = 1$. In effect, C counts the number of heads, and M indicates whether all the coins match.

Since each outcome uniquely determines C and M, we can regard them as functions mapping outcomes to numbers. For this experiment, the sample space is:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Now C is a function that maps each outcome in the sample space to a number as follows:

$$\begin{aligned} C(HHH) &= 3 & \quad C(THH) &= 2 \\ C(HHT) &= 2 & \quad C(THT) &= 1 \\ C(HTH) &= 2 & \quad C(TTH) &= 1 \\ C(HTT) &= 1 & \quad C(TTT) &= 0. \end{aligned}$$

Similarly, M is a function mapping each outcome another way:

$$\begin{array}{rclrcl} M(HHH) &=& 1 & M(THH) &=& 0 \\ M(HHT) &=& 0 & M(THT) &=& 0 \\ M(HTH) &=& 0 & M(TTH) &=& 0 \\ M(HTT) &=& 0 & M(TTT) &=& 1. \end{array}$$

So C and M are random variables.

17.1.1 Indicator Random Variables

An *indicator random variable* is a random variable that maps every outcome to either 0 or 1. Indicator random variables are also called *Bernoulli variables.* The random variable M is an example. If all three coins match, then $M = 1$; otherwise, $M = 0$.

1 or 0

Indicator random variables are closely related to events. In particular, an indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0. For example, the indicator M partitions the sample space into two blocks as follows:

$$\underbrace{HHH \quad TTT}_{M=1} \quad \underbrace{HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH}_{M=0}.$$

In the same way, an event E partitions the sample space into those outcomes in E and those not in E. So E is naturally associated with an indicator random variable, I_E, where $I_E(\omega) = 1$ for outcomes $\omega \in E$ and $I_E(\omega) = 0$ for outcomes $\omega \notin E$. Thus, $M = I_E$ where E is the event that all three coins match.

? so how are they different

17.1.2 Random Variables and Events

There is a strong relationship between events and more general random variables as well. A random variable that takes on several values partitions the sample space into several blocks. For example, C partitions the sample space as follows:

$$\underbrace{TTT}_{C=0} \quad \underbrace{TTH \quad THT \quad HTT}_{C=1} \quad \underbrace{THH \quad HTH \quad HHT}_{C=2} \quad \underbrace{HHH}_{C=3}.$$

Each block is a subset of the sample space and is therefore an event. So the assertion that $C = 2$ defines the event

$$[C = 2] = \{THH, HTH, HHT\},$$

and this event has probability

$$\Pr[C = 2] = \Pr[THH] + \Pr[HTH] + \Pr[HHT] = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = 3/8.$$

Likewise $[M = 1]$ is the event $\{TTT + HHH\}$ and has probability $1/4$.

More generally, any assertion about the values of random variables defines an event. For example, the assertion that $C \leq 1$ defines

$$[C \leq 1] = \{TTT, TTH, THT, HTT\},$$

and so $\Pr[C \leq 1] = 1/2$.

Another example is the assertion that $C \cdot M$ is an odd number. This is an obscure way of saying that all three coins came up heads, namely,

$$[C \cdot M \text{ is odd}] = \{TTT\}.$$

Think about it!

? How to prove

17.2 Independence

The notion of independence carries over from events to random variables as well. Random variables R_1 and R_2 are *independent* iff for all x_1, x_2, the two events

$$[R_1 = x_1] \quad \text{and} \quad [R_2 = x_2]$$

are independent.

~~For example~~, are C and M independent? Intuitively, the answer should be "no." ? how to prove? The number of heads, C, completely determines whether all three coins match; that is, whether $M = 1$. But, to verify this intuition, we must find some $x_1, x_2 \in \mathbb{R}$ such that:

how to prove

$$\Pr[C = x_1 \text{ AND } M = x_2] \neq \Pr[C = x_1] \cdot \Pr[M = x_2].$$

One appropriate choice of values is $x_1 = 2$ and $x_2 = 1$. In this case, we have:

$$\Pr[C = 2 \text{ AND } M = 1] = 0 \neq \frac{1}{4} \cdot \frac{3}{8} = \Pr[M = 1] \cdot \Pr[C = 2].$$

The first probability is zero because we never have exactly two heads ($C = 2$) when all three coins match ($M = 1$). The other two probabilities were computed earlier.

On the other hand, let H_1 be the indicator variable for event that the first flip is a Head, so

$$[H_1 = 1] = \{HHH, HTH, HHT, HTT\}.$$

Then H_1 is independent of M, since

$$\Pr[M = 1] = 1/4 = \Pr\big[M = 1 \mid H_1 = 1\big] = \Pr\big[M = 1 \mid H_1 = 0\big]$$
$$\Pr[M = 0] = 3/4 = \Pr\big[M = 0 \mid H_1 = 1\big] = \Pr\big[M = 0 \mid H_1 = 0\big]$$

This example is an instance of:

∴ so can prove ind w/ indicator RV's

Lemma 17.2.1. *Two events are independent iff their indicator variables are independent.*

The simple proof is left to Problem 17.2.

As with events, the notion of independence generalizes to more than two random variables.

Definition 17.2.2. Random variables $R_1, R_2, \ldots, R_n$ are *mutually independent* iff for all $x_1, x_2, \ldots, x_n$, the n events

$$[R_1 = x_1], [R_2 = x_2], \ldots, [R_n = x_n]$$

are mutually independent.

17.3 Distribution Functions

So RV is just diff way of repres[en]ting; easier for certain cases;

A random variable maps outcomes to values. Often, random variables that show up for different spaces of outcomes wind up behaving in much the same way because they have the same probability of taking any given value. Hence, random variables on different probability spaces may wind up having the same *probability density function.*

PDF

Definition 17.3.1. Let R be a random variable with codomain V. The *probability density function (pdf)* of R is a function $\text{PDF}_R : V \to [0, 1]$ defined by:

$$\text{PDF}_R(x) ::= \begin{cases} \Pr[R = x] & \text{if } x \in \text{range}(R), \\ 0 & \text{if } x \notin \text{range}(R). \end{cases}$$

A consequence of this definition is that

$$\sum_{x \in \text{range}(R)} \text{PDF}_R(x) = 1.$$

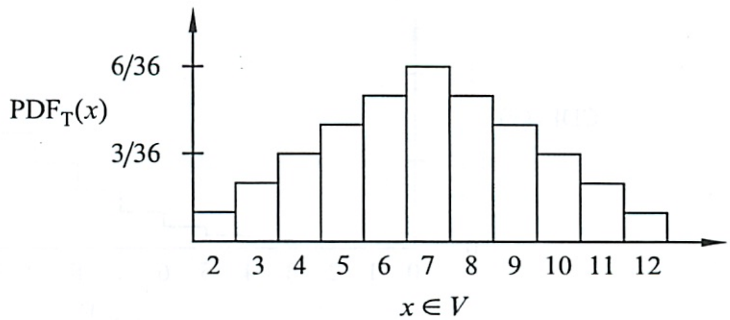


Figure 17.1 The probability density function for the sum of two 6-sided dice.

So the sum is like a RV

This is because R has a value for each outcome, so summing the probabilities over all outcomes is the same as summing over the probabilities of each value in the range of R.

As an example, suppose that you roll two unbiased, independent, 6-sided dice. Let T be the random variable that equals the sum of the two rolls. This random variable takes on values in the set $V = \{2, 3, \ldots, 12\}$. A plot of the probability density function for T is shown in Figure 17.1: The lump in the middle indicates that sums close to 7 are the most likely. The total area of all the rectangles is 1 since the dice must take on exactly one of the sums in $V = \{2, 3, \ldots, 12\}$.

Cumulative distribution functions (cdf's) are closely-related to pdf's. The cdf for a random variable R whose codomain is a subset of real numbers is the function $\mathrm{CDF}_R : \mathbb{R} \to [0, 1]$ defined by:

CDF

$$\mathrm{CDF}_R(x) ::= \Pr[R \le x].$$

As an example, the cumulative distribution function for the random variable T is shown in Figure 17.2: The height of the ith bar in the cumulative distribution function is equal to the *sum* of the heights of the leftmost i bars in the probability density function. This follows from the definitions of pdf and cdf:

$$\begin{aligned}
\mathrm{CDF}_R(x) &= \Pr[R \le x] \\
&= \sum_{y \le x} \Pr[R = y] \\
&= \sum_{y \le x} \mathrm{PDF}_R(y).
\end{aligned}$$

In summary, $\mathrm{PDF}_R(x)$ measures the probability that $R = x$ and $\mathrm{CDF}_R(x)$ measures the probability that $R \le x$. Both PDF_R and CDF_R capture the same

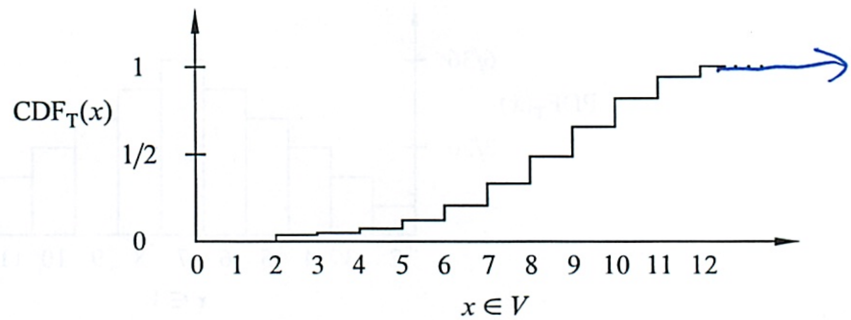


Figure 17.2 The cumulative distribution function for the sum of two 6-sided dice.

information about the random variable R —obviously each one determines the other —but sometimes one is more convenient. The key point here is that neither the probability density function nor the cumulative distribution function involves the sample space of an experiment.

One of the really interesting things about density functions and distribution functions is that many random variables turn out to have the *same* pdf and cdf. In other words, even though R and S are different random variables on different probability spaces, it is often the case that

$$\text{PDF}_R = \text{PDF}_S.$$

In fact, some pdf's are so common that they are given special names. For example, the three most important distributions in computer science are the *Bernoulli distribution*, the *uniform distribution*, and the *binomial distribution*. We look more closely at these common distributions in the next several sections.

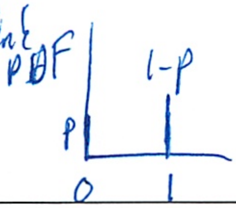
17.3.1 Bernoulli Distributions

The Bernoulli distribution is the simplest and most common distribution function. That's because it is the distribution function for an indicator random variable. Specifically, the *Bernoulli distribution* has a probability density function of the form $f_p : \{0, 1\} \to [0, 1]$ where

$$f_p(0) = p, \quad \text{and}$$
$$f_p(1) = 1 - p,$$

1 or 0

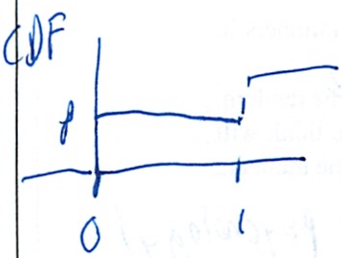
I think



remembered
- yeah makes sense

CDF

for some $p \in [0, 1]$. The corresponding cumulative distribution function is $F_p : \mathbb{R} \to [0, 1]$ where



$$F_p(x) ::= \begin{cases} 0 & \text{if } x < 0 \\ p & \text{if } 0 \le x < 1 \\ 1 & \text{if } 1 \le x. \end{cases}$$

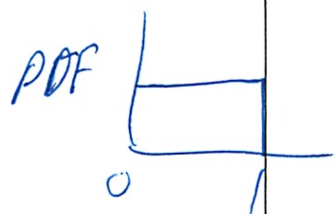
← just defines base cases,
← actual cases
← picture

17.3.2 Uniform Distributions

A random variable that takes on each possible value in its codomain with the same probability is said to be *uniform*. If the codomain V has n elements, then the *uniform distribution* has a pdf of the form

new notation
- not from 6.041

$$f : V \to [0, 1]$$

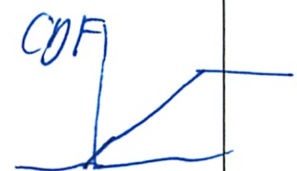


where

$$f(v) = \frac{1}{n}$$

for all $v \in V$.

If $V = \{1, 2, \ldots, n\}$, the cumulative distribution function would be $F : \mathbb{R} \to [0, 1]$ where



$$F(x) ::= \begin{cases} 0 & \text{if } x < 1 \\ k/n & \text{if } k \le x < k+1 \text{ for } 1 \le k < n \\ 1 & \text{if } n \le x. \end{cases}$$

Uniform distributions come up all the time. For example, the number rolled on a fair die is uniform on the set $\{1, 2, \ldots, 6\}$. An indicator variable is uniform when its pdf is $f_{1/2}$.

continuous though

17.3.3 The Numbers Game

Enough definitions —let's play a game! We have two envelopes. Each contains an integer in the range $0, 1, \ldots, 100$, and the numbers are distinct. To win the game, you must determine which envelope contains the larger number. To give you a fighting chance, we'll let you peek at the number in one envelope selected at random. Can you devise a strategy that gives you a better than 50% chance of winning?

For example, you could just pick an envelope at random and guess that it contains the larger number. But this strategy wins only 50% of the time. Your challenge is to do better.

So you might try to be more clever. Suppose you peek in one envelope and see the number 12. Since 12 is a small number, you might guess that the number in the sh did something like this in class as well

other envelope is larger. But perhaps we've been tricky and put small numbers in *both* envelopes. Then your guess might ~~not be so~~ good!

An important point here is that the numbers in the envelopes may *not* be random. We're picking the numbers and we're choosing them in a way that we think will defeat your guessing strategy. We'll only use randomization to choose the numbers if that serves our purpose, which is making you lose!

gets into psychology/ game theory

Intuition Behind the Winning Strategy

Amazingly, there is a strategy that wins more than 50% of the time, regardless of what numbers we put in the envelopes!

Suppose that you somehow knew a number x that was in between the numbers in the envelopes. Now you peek in one envelope and see a number. If it is bigger than x, then you know you're peeking at the higher number. If it is smaller than x, then you're peeking at the lower number. In other words, if you know a number x between the numbers in the envelopes, then you are certain to win the game.

The only flaw with this brilliant strategy is that you do *not* know such an x. Oh well.

But what if you try to *guess* x? There is some probability that you guess correctly. In this case, you win 100% of the time. On the other hand, if you guess incorrectly, then you're no worse off than before; your chance of winning is still 50%. Combining these two cases, your overall chance of winning is better than 50%!

how

Informal arguments about probability, like this one, often sound plausible, but do not hold up under close scrutiny. In contrast, this argument sounds completely implausible —but is actually correct!

Analysis of the Winning Strategy

For generality, suppose that we can choose numbers from the set $\{0, 1, \ldots, n\}$. Call the lower number L and the higher number H.

Your goal is to guess a number x between L and H. To avoid confusing equality cases, you select x at random from among the half-integers:

$$\{\frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}, \ldots, n - \frac{1}{2}\}$$

But what probability distribution should you use?

The uniform distribution turns out to be your best bet. An informal justification is that if we figured out that you were unlikely to pick some number —say $50\frac{1}{2}$ —then we'd always put 50 and 51 in the envelopes. Then you'd be unlikely to pick an x between L and H and would have less chance of winning.

After you've selected the number x, you peek into an envelope and see some number T. If $T > x$, then you guess that you're looking at the larger number. If $T < x$, then you guess that the other number is larger.

All that remains is to determine the probability that this strategy succeeds. We can do this with the usual four step method and a tree diagram.

Step 1: Find the sample space.
You either choose x too low ($< L$), too high ($> H$), or just right ($L < x < H$). Then you either peek at the lower number ($T = L$) or the higher number ($T = H$). This gives a total of six possible outcomes, as show in Figure 17.3.

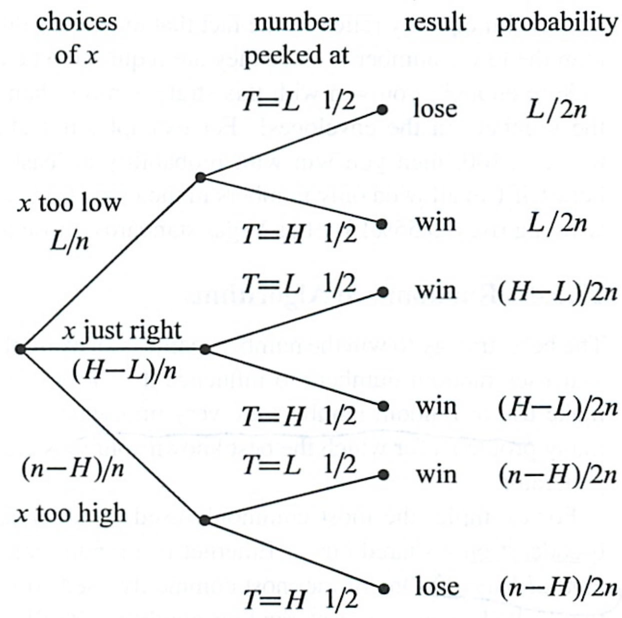


Figure 17.3 The tree diagram for the numbers game.

Step 2: Define events of interest.
The four outcomes in the event that you win are marked in the tree diagram.

Step 3: Assign outcome probabilities.
First, we assign edge probabilities. Your guess x is too low with probability L/n, too high with probability $(n - H)/n$, and just right with probability $(H - L)/n$. Next, you peek at either the lower or higher number with equal probability. Multiplying along root-to-leaf paths gives the outcome probabilities.

W8W

Step 4: Compute event probabilities.
The probability of the event that you win is the sum of the probabilities of the four outcomes in that event:

$$\begin{aligned}\Pr[\text{win}] &= \frac{L}{2n} + \frac{H-L}{2n} + \frac{H-L}{2n} + \frac{n-H}{2n} \\ &= \frac{1}{2} + \frac{H-L}{2n} \\ &\geq \frac{1}{2} + \frac{1}{2n}\end{aligned}$$

The final inequality relies on the fact that the higher number H is at least 1 greater than the lower number L since they are required to be distinct.

Sure enough, you win with this strategy more than half the time, regardless of the numbers in the envelopes! For example, if I choose numbers in the range $0, 1, \ldots, 100$, then you win with probability at least $\frac{1}{2} + \frac{1}{200} = 50.5\%$. Even better, if I'm allowed only numbers in the range $0, \ldots, 10$, then your probability of winning rises to 55%! By Las Vegas standards, those are great odds!

17.3.4 Randomized Algorithms

The best strategy to win the numbers game is an example of ~~a~~ *randomized algorithm* —it uses random numbers to influence decisions. Protocols and algorithms that make use of random numbers are very important in computer science. There are many problems for which the best known solutions are based on a random number generator.

?, how

For example, the most commonly-used protocol for deciding when to send a broadcast on a shared bus or Ethernet is a randomized algorithm known as *exponential backoff*. One of the most commonly-used sorting algorithms used in practice, called *quicksort*, uses random numbers. You'll see many more examples if you take an algorithms course. In each case, randomness is used to improve the probability that the algorithm runs quickly or otherwise performs well.

Oh, just that

17.3.5 Binomial Distributions

The third commonly-used distribution in computer science is the *binomial distribution*. The standard example of a random variable with a binomial distribution is the number of heads that come up in n independent flips of a coin. If the coin is fair, then the number of heads has an *unbiased binomial distribution*, specified by the pdf

$$f_n : \{1, 2, \ldots, n\} \to [0, 1].$$

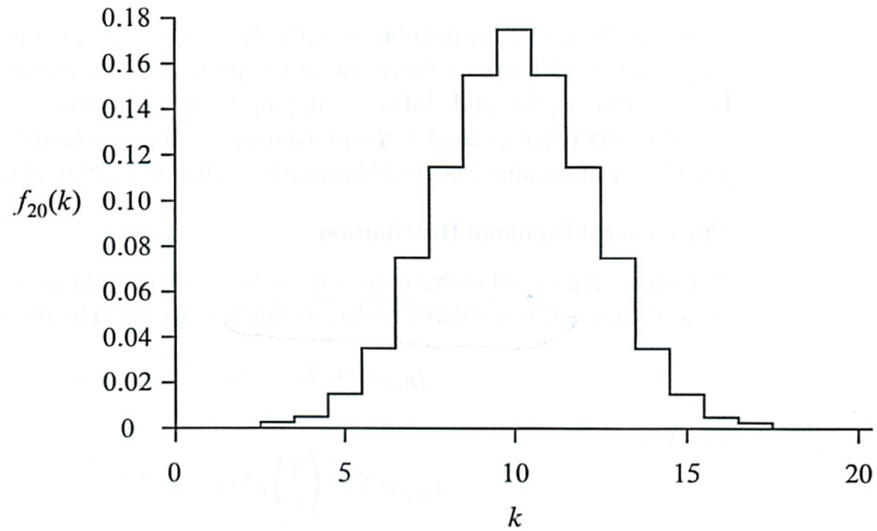


Figure 17.4 The pdf for the unbiased binomial distribution for $n = 20$, $f_{20}(k)$.

This is because there are $\binom{n}{k}$ sequences of n coin tosses with exactly k heads, and each such sequence has probability 2^{-n}.

A plot of $f_{20}(k)$ is shown in Figure 17.4. The most likely outcome is $k = 10$ heads, and the probability falls off rapidly for larger and smaller values of k. The falloff regions to the left and right of the main hump are called the *tails of the distribution.*

The cumulative distribution function for the unbiased binomial distribution is $F_n : \mathbb{R} \to [0, 1]$ where

CDF

$$F_n(x) = \begin{cases} 0 & \text{if } x < 1 \\ \sum_{i=0}^{k} \binom{n}{i} 2^{-n} & \text{if } k \le x < k+1 \text{ for } 1 \le k < n \\ 1 & \text{if } n \le x. \end{cases}$$

did we write stuff like this in 6.041?

In many fields, including Computer Science, probability analyses come down to getting small bounds on the tails of the binomial distribution. In the context of a problem, this typically means that there is very small probability that something *bad* happens, which could be a server or communication link overloading or a randomized algorithm running for an exceptionally long time or producing the wrong result.

As an example, we can calculate the probability of flipping at most 25 heads in 100 tosses of a fair coin and see that it is very small, namely, less than 1 in 3,000,000.

In fact, the tail of the distribution falls off so rapidly that the probability of flipping exactly 25 heads is nearly twice the probability of flipping fewer than 25 heads! That is, the probability of flipping exactly 25 heads —small as it is — is still nearly twice as large as the probability of flipping exactly 24 heads *plus* the probability of flipping exactly 23 heads *plus* ... the probability of flipping no heads.

The General Binomial Distribution

If the coins are biased so that each coin is heads with probability p, then the number of heads has a *general binomial density function* specified by the pdf

$$f_{n,p} : \{1, 2, \ldots, n\} \to [0, 1]$$

where

$$f_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

for some $n \in \mathbb{N}^+$ and $p \in [0, 1]$. This is because there are $\binom{n}{k}$ sequences with k heads and $n - k$ tails, but now the probability of each such sequence is $p^k(1 - p)^{n-k}$.

n= # of flips
p= prob of heads

For example, the plot in Figure 17.5 shows the probability density function $f_{n,p}(k)$ corresponding to flipping $n = 20$ independent coins that are heads with probability $p = 0.75$. The graph shows that we are most likely to get $k = 15$ heads, as you might expect. Once again, the probability falls off quickly for larger and smaller values of k.

The cumulative distribution function for the general binomial distribution is $F_{n,p} : \mathbb{R} \to [0, 1]$ where

CDF

$$F_{n,p}(x) = \begin{cases} 0 & \text{if } x < 1 \\ \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} & \text{if } k \le x < k+1 \text{ for } 1 \le k < n \\ 1 & \text{if } n \le x. \end{cases} \tag{17.1}$$

17.4 Great Expectations

The *expectation* or *expected value* of a random variable is a single number that tells reveals a lot about the behavior of the variable. The expectation of a random variable is also known as its *mean* or *average*. It is the average value of the variable where each value is weighted according to its probability.

For example, suppose we select a student uniformly at random ~~from the~~ class, and let R be the student's quiz score. Then $\text{Ex}[R]$ is just the class average —the

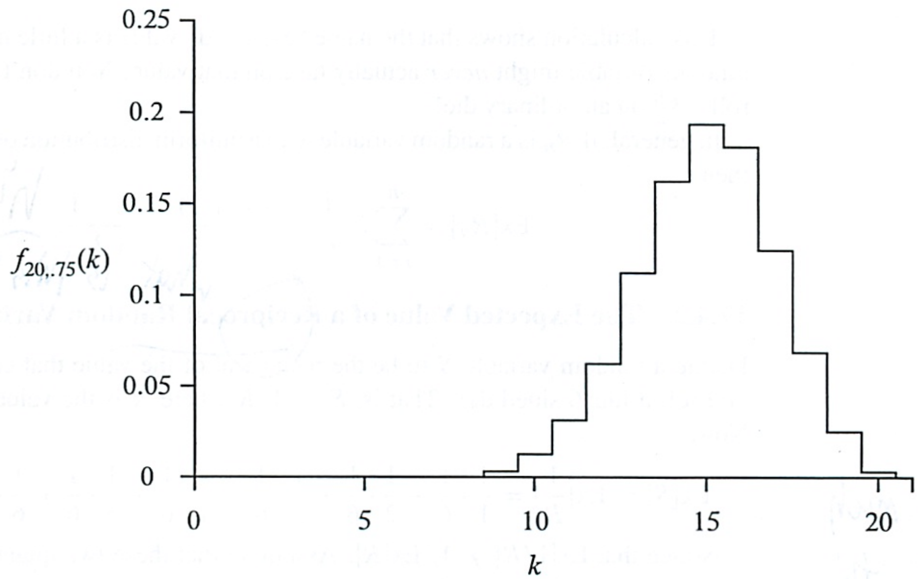


Figure 17.5 The pdf for the general binomial distribution $f_{n,p}(k)$ for $n = 20$ and $p = .75$.

first thing everyone wants to know after getting their test back! For similar reasons, the first thing you usually want to know about a random variable is its expected value.

Formally, the expected value of a random variable is defined as follows:

Definition 17.4.1. If R is a random variable defined on a sample space S, then the expectation of R is

$$\text{Ex}[R] ::= \sum_{\omega \in S} R(\omega) \Pr[\omega]. \tag{17.2}$$

this is on sets

Let's work through some examples.

17.4.1 The Expected Value of a Uniform Random Variable

Rolling a 6-sided die provides an example of a uniform random variable. Let R be the value that comes up when you roll a fair 6-sided die. Then by (17.2), the expected value of R is

$$\text{Ex}[R] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

This calculation shows that the name "expected" value is a little misleading; the random variable might *never* actually take on that value. You don't ever expect to roll a $3\frac{1}{2}$ on an ordinary die!

In general, if R_n is a random variable with a uniform distribution on $\{1, 2, \ldots, n\}$, then

$$\text{Ex}[R_n] = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}.$$

this formula $\frac{b+a}{2}$

what is this? not defined

17.4.2 The Expected Value of a Reciprocal Random Variable

Define a random variable S to be the reciprocal of the value that comes up when you roll a fair 6-sided die. That is, $S = 1/R$ where R is the value that you roll. Now,

$$\text{Ex}[S] = \text{Ex}[\frac{1}{R}] = \frac{1}{1} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{49}{120}.$$

There is so much to learn in this class!

Notice that $\text{Ex}[1/R] \neq 1/\text{Ex}[R]$. Assuming that these two quantities are equal is a common mistake.

17.4.3 The Expected Value of an Indicator Random Variable

The expected value of an indicator random variable for an event is just the probability of that event.

Oh! wow!

Lemma 17.4.2. *If I_A is the indicator random variable for event A, then*

$$\text{Ex}[I_A] = \Pr[A].$$

Proof.

$$\begin{aligned} \text{Ex}[I_A] &= 1 \cdot \Pr[I_A = 1] + 0 \cdot \Pr[I_A = 0] \\ &= \Pr[I_A = 1] \\ &= \Pr[A]. && \text{(def of } I_A\text{)} \end{aligned}$$

For example, if A is the event that a coin with bias p comes up heads, then $\text{Ex}[I_A] = \Pr[I_A = 1] = p$.

17.4.4 Alternate Definition of Expectation

There is another standard way to define expectation.

Theorem 17.4.3. *For any random variable R,*

$$\text{Ex}[R] = \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]. \tag{17.3}$$

set

The proof of Theorem 17.4.3, like many of the elementary proofs about expectation in this chapter, follows by judicious regrouping of terms in equation 17.2:

Proof. Suppose R is defined on a sample space $\mathcal{S}$. Then,

$$\begin{aligned}
\text{Ex}[R] &= \sum_{\omega \in \mathcal{S}} R(\omega) \Pr[\omega] && \text{(Def 17.4.1 of expectation)} \\
&= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R=x]} R(\omega) \Pr[\omega] \\
&= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R=x]} x \Pr[\omega] && \text{(def of the event } [R = x]) \\
&= \sum_{x \in \text{range}(R)} x \left(\sum_{\omega \in [R=x]} \Pr[\omega] \right) && \text{(distributing } x \text{ over the inner sum)} \\
&= \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]. && \text{(def of } \Pr[R = x])
\end{aligned}$$

look at closly

The first equality follows because the events $[R = x]$ for $x \in \text{range}(R)$ partition the sample space $\mathcal{S}$, so summing over the outcomes in $[R = x]$ for $x \in \text{range}(R)$ is the same as summing over $\mathcal{S}$. ■

In general, equation (17.3) is more useful than the defining equation (17.2) for calculating expected values. It also has the advantage that it does not depend on the sample space, but only on the density function of the random variable. On the other hand, summing over all outcomes as in equation (17.2) sometimes yields easier proofs about general properties of expectation.

Medians

The mean of a random variable is not the same as the *median*. The median is the *midpoint* of a distribution.

Definition 17.4.4. The *median* of a random variable R is the value $x \in \text{range}(R)$ such that

$$\Pr[R \leq x] \leq \frac{1}{2} \quad \text{and}$$

$$\Pr[R > x] < \frac{1}{2}.$$

middle

We won't devote much attention to the median. The expected value is more useful and has much more interesting properties.

17.4.5 Conditional Expectation

Just like event probabilities, expectations can be conditioned on some event. Given a random variable R, the expected value of R conditioned on an event A is the probability-weighted average value of R over outcomes in A. More formally:

Definition 17.4.5. The *conditional expectation* $\text{Ex}[R \mid A]$ of a random variable R given event A is:

$$\text{Ex}[R \mid A] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr\big[R = r \mid A\big]. \tag{17.4}$$

For example, we can compute the expected value of a roll of a fair die, given that the number rolled is at least 4. We do this by letting R be the outcome of a roll of the die. Then by equation (17.4),

$$\text{Ex}[R \mid R \geq 4] = \sum_{i=1}^{6} i \cdot \Pr\big[R = i \mid R \geq 4\big] = 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot \tfrac{1}{3} + 5 \cdot \tfrac{1}{3} + 6 \cdot \tfrac{1}{3} = 5.$$

Conditional expectation is useful in dividing complicated expectation calculations into simpler cases. We can find a desired expectation by calculating the conditional expectation in each simple case and averaging them, weighing each case by its probability.

For example, suppose that 49.8% of the people in the world are male and the rest female —which is more or less true. Also suppose the expected height of a randomly chosen male is $5'\,11''$, while the expected height of a randomly chosen female is $5'\,5''$. What is the expected height of a randomly chosen person? We can calculate this by averaging the heights of men and women. Namely, let H be the height (in feet) of a randomly chosen person, and let M be the event that the person is male and F the event that the person is female. Then

$$\begin{aligned}
\text{Ex}[H] &= \text{Ex}[H \mid M]\Pr[M] + \text{Ex}[H \mid F]\Pr[F] \\
&= (5 + 11/12) \cdot 0.498 + (5 + 5/12) \cdot 0.502 \\
&= 5.665
\end{aligned}$$

which is a little less than 5' 8".

This method is justified by:

Theorem 17.4.6 (Law of Total Expectation). *Let R be a random variable on a sample space $\mathcal{S}$, and suppose that $A_1, A_2, \ldots,$ is a partition of $\mathcal{S}$. Then*

$$\text{Ex}[R] = \sum_i \text{Ex}[R \mid A_i]\Pr[A_i].$$

Proof.

$$
\begin{aligned}
\text{Ex}[R] &= \sum_{r \in \text{range}(R)} r \cdot \Pr[R = r] && \text{(by 17.3)} \\
&= \sum_r r \cdot \sum_i \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(Law of Total Probability)} \\
&= \sum_r \sum_i r \cdot \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(distribute constant } r\text{)} \\
&= \sum_i \sum_r r \cdot \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(exchange order of summation)} \\
&= \sum_i \Pr[A_i] \sum_r r \cdot \Pr\left[R = r \mid A_i\right] && \text{(factor constant } \Pr[A_i]\text{)} \\
&= \sum_i \Pr[A_i] \, \text{Ex}[R \mid A_i]. && \text{(Def 17.4.5 of cond. expectation)}
\end{aligned}
$$

■

17.4.6 Mean Time to Failure

A computer program crashes at the end of each hour of use with probability p, if it has not crashed already. What is the expected time until the program crashes? This will be easy to figure out using the Law of Total Expectation (Theorem 17.4.6). Specifically, we want to find $\text{Ex}[C]$ where C is the number of hours until the first crash. We'll do this by conditioniing on whether or not the crash occurs in the first hour.

So let A to be the event that the system fails on the first step and $\overline{A}$ to be the complementary event that the system does not fail on the first step. Then the mean time to failure $\text{Ex}[C]$ is

$$\text{Ex}[C] = \text{Ex}[C \mid A] \Pr[A] + \text{Ex}[C \mid \overline{A}] \Pr[\overline{A}]. \tag{17.5}$$

Since A is the condition that the system crashes on the first step, we know that

$$\text{Ex}[C \mid A] = 1. \tag{17.6}$$

Since $\overline{A}$ is the condition that the system does *not* crash on the first step, conditioning on $\overline{A}$ is equivalent to taking a first step without failure and then starting over without conditioning. Hence,

$$\text{Ex}[C \mid \overline{A}] = 1 + \text{Ex}[C]. \tag{17.7}$$

Plugging (17.6) and (17.7) into (17.5):

$$\begin{aligned}\text{Ex}[C] &= 1 \cdot p + (1 + \text{Ex}[C])(1-p) \\ &= p + 1 - p + (1-p)\,\text{Ex}[C] \\ &= 1 + (1-p)\,\text{Ex}[C].\end{aligned}$$

Then, rearranging terms gives

$$1 = \text{Ex}[C] - (1-p)\,\text{Ex}[C] = p\,\text{Ex}[C],$$

and thus

$$\text{Ex}[C] = \frac{1}{p}.$$

The general principle here is well-worth remembering.

Mean Time to Failure

If a system fails at each time step with probability p, then the expected number of steps up to the first failure is $1/p$.

So, for example, if there is a 1% chance that the program crashes at the end of each hour, then the expected time until the program crashes is $1/0.01 = 100$ hours.

As a further example, suppose a couple wants to have a baby girl. For simplicity assume there is a 50% chance that each child they have is a girl, and the genders of their children are mutually independent. If the couple insists on having children until they get a girl, then how many baby boys should they expect first?

This is really a variant of the previous problem. The question, "How many hours until the program crashes?" is mathematically the same as the question, "How many children must the couple have until they get a girl?" In this case, a crash corresponds to having a girl, so we should set $p = 1/2$. By the preceding analysis, the couple should expect a baby girl after having $1/p = 2$ children. Since the last of these will be the girl, they should expect just one boy.

Oh if 1st is boy won't have more! gets complex!

Something to think about: If every couple follows the strategy of having children until they get a girl, what will eventually happen to the fraction of girls born in this world?

Using the Law of Total Expectation to find expectations is worthwhile approach to keep in mind, but it's good review to derive the same formula directly from the definition of expectation. Namely, the probability that the first crash occurs in the ith hour for some $i > 0$ is the probability, $(1-p)^{i-1}$, that it does not crash in each of the first $i-1$ hours, times the probability, p, that it does crash in the ith hour. So

$$\begin{aligned}
\text{Ex}[C] &= \sum_{i\in\mathbb{N}} i \cdot \Pr[C=i] && \text{(by (17.3))}\\
&= \sum_{i\in\mathbb{N}} i(1-p)^{i-1}p\\
&= \frac{p}{1-p}\cdot\sum_{i\in\mathbb{N}} i(1-p)^i. && (17.8)
\end{aligned}$$

But we've already seen a sum like this last one, namely, equation (14.13):

$$\sum_{i\in\mathbb{N}} ix^i = \frac{x}{(1-x)^2}.$$

Combining (14.13) with (17.8) gives

$$\text{Ex}[C] = \frac{p}{1-p}\cdot\frac{1-p}{(1-(1-p))^2} = \frac{1}{p}$$

as expected.

17.4.7 Expected Returns in Gambling Games

Some of the most interesting examples of expectation can be explained in terms of gambling games. For straightforward games where you win w dollars with probability p and you lose x dollars with probability $1-p$, it is easy to compute your *expected return* or *winnings*. It is simply

$$pw-(1-p)x \text{ dollars.}$$

For example, if you are flipping a fair coin and you win \$1 for heads and you lose \$1 for tails, then your expected winnings are

$$\frac{1}{2}\cdot 1 - \left(1-\frac{1}{2}\right)\cdot 1 = 0.$$

In such cases, the game is said to be *fair* since your expected return is zero.

Some gambling games are more complicated and thus more interesting. The following game where the winners split a pot is representative of many poker games, betting pools, and lotteries.

Splitting the Pot

After your last encounter with biker dude, one thing led to another and you have dropped out of school and become a Hell's Angel. It's late on a Friday night and, feeling nostalgic for the old days, you drop by your old hangout, where you encounter two of your former TAs, Eric and Nick. Eric and Nick propose that you join them in a simple wager. Each player will put \$2 on the bar and secretly write "heads" or "tails" on their napkin. Then one player will flip a fair coin. The \$6 on the bar will then be divided equally among the players who correctly predicted the outcome of the coin toss.

After your life-altering encounter with strange dice, you are more than a little skeptical. So Eric and Nick agree to let you be the one to flip the coin. This certainly seems fair. How can you lose?

But you have learned your lesson and so before agreeing, you go through the four-step method and write out the tree diagram to compute your expected return. The tree diagram is shown in Figure 17.6.

The "payoff" values in Figure 17.6 are computed by dividing the \$6 pot[1] among those players who guessed correctly and then subtracting the \$2 that you put into the pot at the beginning. For example, if all three players guessed correctly, then your payoff is \$0, since you just get back your \$2 wager. If you and Nick guess correctly and Eric guessed wrong, then your payoff is

$$\frac{6}{2} - 2 = 1.$$

In the case that everyone is wrong, you all agree to split the pot and so, again, your payoff is zero.

To compute your expected return, you use equation (17.3):

$$\begin{aligned}\text{Ex[payoff]} &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} \\ &\quad + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + 0 \cdot \frac{1}{8} \\ &= 0.\end{aligned}$$

This confirms that the game is fair. So, for old time's sake, you break your solemn vow to never ever engage in strange gambling games.

The Impact of Collusion

Needless to say, things are not turning out well for you. The more times you play the game, the more money you seem to be losing. After 1000 wagers, you have lost

[1] The money invested in a wager is commonly referred to as the *pot*.

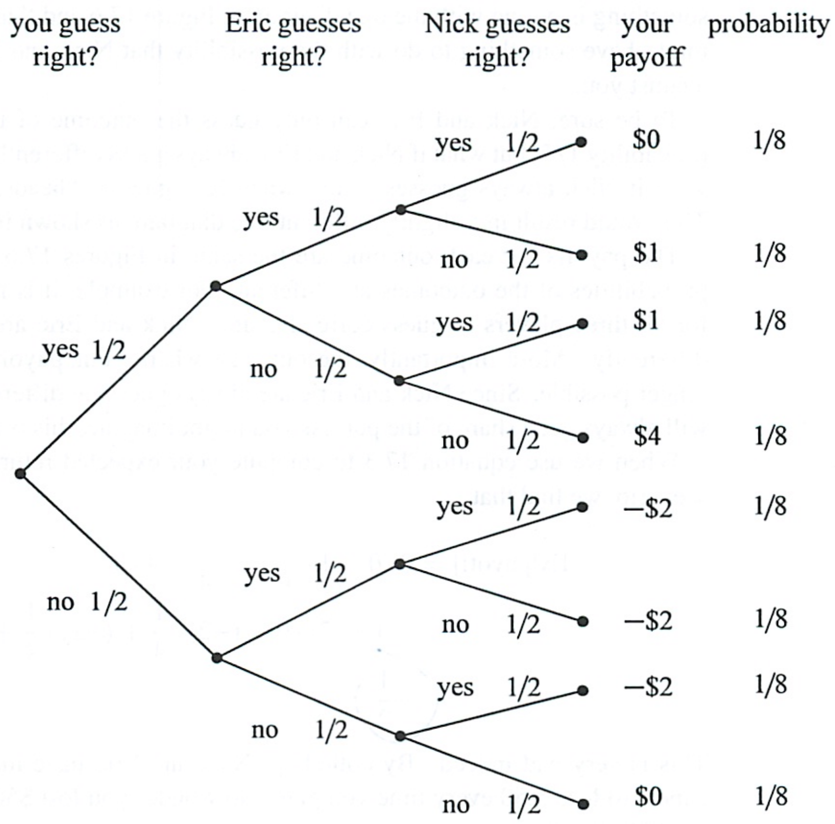


Figure 17.6 The tree diagram for the game where three players each wager $2 and then guess the outcome of a fair coin toss. The winners split the pot.

over \$500. As Nick and Eric are consoling you on your "bad luck," you remember how rapidly the tails of the binomial distribute decrease, suggesting that the probability of losing \$500 in 1000 fair \$2 wagers is less than the probability of being struck by lightning while playing poker and being dealt four Aces. How can this be?

It is possible that you are truly very very unlucky. But it is more likely that something is wrong with the tree diagram in Figure 17.6 and that "something" just might have something to do with the possibility that Nick and Eric are ~~colluding~~ against you. not ind!

To be sure, Nick and Eric can only guess the outcome of the coin toss with probability $1/2$, but what if Nick and Eric always guess differently? In other words, what if Nick always guesses "tails" when Eric guesses "heads," and vice-versa? This would result in a slightly different tree diagram, as shown in Figure 17.7.

The payoffs for each outcome are the same in Figures 17.6 and 17.7, but the probabilities of the outcomes are different. For example, it is no longer possible for all three players to guess correctly, since Nick and Eric are always guessing differently. More importantly, the outcome where your payoff is \$4 is also no longer possible. Since Nick and Eric are always guessing differently, one of them will always get a share of the pot. As you might imagine, this is not good for you!

When we use equation 17.3 to compute your expected return in the collusion scenario, we find that

$$\begin{aligned}\text{Ex}[\text{payoff}] &= 0 \cdot 0 + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 4 \cdot 0 \\ &\quad + (-2) \cdot 0 + (-2) \cdot \frac{1}{4} + (-2) \cdot \frac{1}{4} + 0 \cdot 0 \\ &= -\frac{1}{2}.\end{aligned}$$

This is very bad indeed. By colluding, Nick and Eric have made it so that you expect to lose \$.50 every time you play. No wonder you lost \$500 over the course of 1000 wagers.

Maybe it would be a good idea to go back to school —your Hell's Angels buds may not be too happy that you just lost their \$500.

How to Win the Lottery

Similar opportunities to "collude" arise in many betting games. For example, consider the typical weekly football betting pool, where each participant wagers \$10 and the participants that pick the most games correctly split a large pot. The pool seems fair if you think of it as in Figure 17.6. But, in fact, if two or more players collude by guessing differently, they can get an "unfair" advantage at your expense!

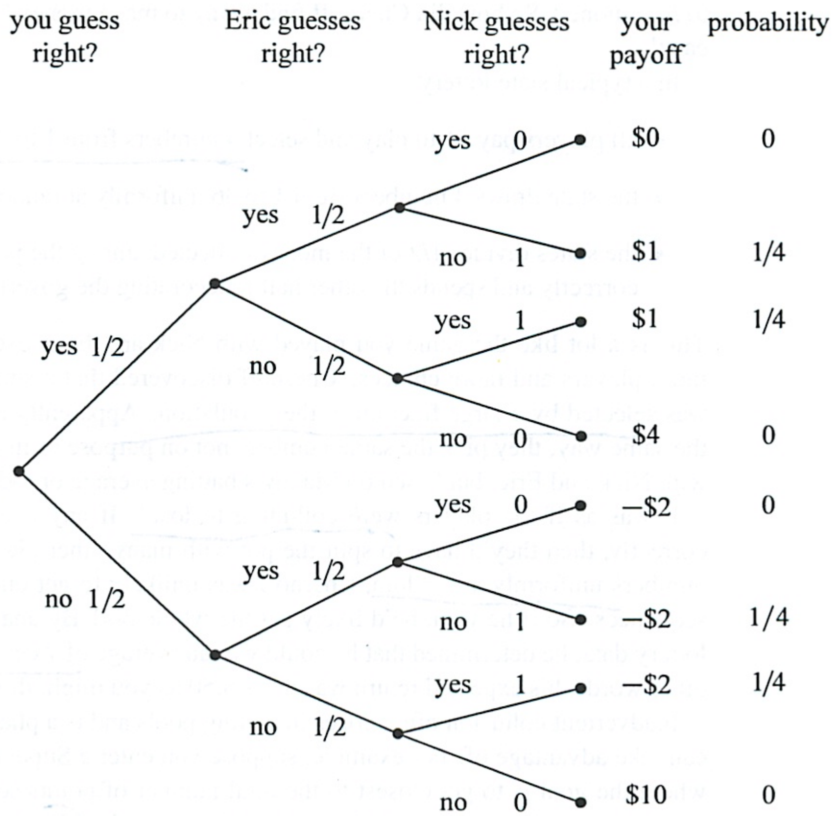


Figure 17.7 The revised tree diagram reflecting the scenario where Nick always guesses the opposite of Eric.

In some cases, the collusion is inadvertent and you can profit from it. For example, many years ago, a former MIT Professor of Mathematics named Herman Chernoff figured out a way to make money by playing the state lottery. This was surprising since state lotteries typically have very poor expected returns. That's because the state usually takes a large share of the wagers before distributing the rest of the pot among the winners. Hence, anyone who buys a lottery ticket is expected to *lose* money. So how did Chernoff find a way to make money? It turned out to be easy!

In a typical state lottery,

- all players pay \$1 to play and select 4 numbers from 1 to 36,
- the state draws 4 numbers from 1 to 36 uniformly at random,
- the states divides 1/2 of the money collected among the people who guessed correctly and spends the other half redecorating the governor's residence.

This is a lot like the game you played with Nick and Eric, except that there are more players and more choices. Chernoff discovered that a small set of numbers was selected by a large fraction of the population. Apparently many people think the same way; they pick the same numbers not on purpose as in the previous game with Nick and Eric, but based on Manny's batting average or today's date.

It was as if the players were colluding to lose! If any one of them guessed correctly, then they'd have to split the pot with many other players. By selecting numbers uniformly at random, Chernoff was unlikely to get one of these favored sequences. So if he won, he'd likely get the whole pot! By analyzing actual state lottery data, he determined that he could win an average of 7 cents on the dollar. In other words, his expected return was not −\$.50 as you might think, but +\$.07.[2] ← not squared footnote

Inadvertent collusion often arises in betting pools and is a phenomenon that you can take advantage of. For example, suppose you enter a Super Bowl betting pool where the goal is to get closest to the total number of points scored in the game. Also suppose that the average Super Bowl has a total of 30 point scored and that everyone knows this. Then most people will guess around 30 points. Where should you guess? Well, you should guess just outside of this range because you get to cover a lot more ground and you don't share the pot if you win. Of course, if you are in a pool with math students and they all know this strategy, then maybe you should guess 30 points after all.

psychology/ game theory

[2] Most lotteries now offer randomized tickets to help smooth out the distribution of selected sequences.

17.4.8 Linearity of Expectation

Expected values obey a simple, very helpful rule called *Linearity of Expectation.* Its simplest form says that the expected value of a sum of random variables is the sum of the expected values of the variables.

Theorem 17.4.7. *For any random variables R_1 and R_2,*

$$\mathrm{Ex}[R_1 + R_2] = \mathrm{Ex}[R_1] + \mathrm{Ex}[R_2].$$

add them up

Proof. Let $T ::= R_1 + R_2$. The proof follows straightforwardly by rearranging terms in equation (17.2) in the definition of expectation:

$$\begin{aligned}
\mathrm{Ex}[T] &= \sum_{\omega \in S} T(\omega) \cdot \Pr[\omega] && \text{(by (17.2))} \\
&= \sum_{\omega \in S} (R_1(\omega) + R_2(\omega)) \cdot \Pr[\omega] && \text{(def of } T\text{)} \\
&= \sum_{\omega \in S} R_1(\omega) \Pr[\omega] + \sum_{\omega \in S} R_2(\omega) \Pr[\omega] && \text{(rearranging terms)} \\
&= \mathrm{Ex}[R_1] + \mathrm{Ex}[R_2]. && \text{(by 17.2)}
\end{aligned}$$

■

A small extension of this proof, which we leave to the reader, implies

Theorem 17.4.8. *For random variables R_1, R_2 and constants $a_1, a_2 \in \mathbb{R}$,*

$$\mathrm{Ex}[a_1 R_1 + a_2 R_2] = a_1 \mathrm{Ex}[R_1] + a_2 \mathrm{Ex}[R_2].$$

In other words, expectation is a linear function. A routine induction extends the result to more than two variables:

Corollary 17.4.9 (Linearity of Expectation)**.** *For any random variables $R_1, \dots, R_k$ and constants $a_1, \dots, a_k \in \mathbb{R}$,*

$$\mathrm{Ex}[\sum_{i=1}^{k} a_i R_i] = \sum_{i=1}^{k} a_i \mathrm{Ex}[R_i].$$

The great thing about linearity of expectation is that *no independence is required.* This is really useful, because dealing with independence is a pain, and we often need to work with random variables that are not known to be independent.

As an example, let's compute the expected value of the sum of two fair dice.

Expected Value of Two Dice

What is the expected value of the sum of two fair dice?

Let the random variable R_1 be the number on the first die, and let R_2 be the number on the second die. We observed earlier that the expected value of one die is 3.5. We can find the expected value of the sum using linearity of expectation:

$$\text{Ex}[R_1 + R_2] = \text{Ex}[R_1] + \text{Ex}[R_2] = 3.5 + 3.5 = 7.$$

Notice that we did *not* have to assume that the two dice were independent. The expected sum of two dice is 7, even if they are glued together (provided each individual die remains fair after the gluing). Proving that this expected sum is 7 with a tree diagram would be a bother: there are 36 cases. And if we did not assume that the dice were independent, the job would be really tough!

17.4.9 Sums of Indicator Random Variables

Another class problem

(I should read earlier!)

Linearity of expectation is especially useful when you have a sum of indicator random variables. As an example, suppose there is a dinner party where n men check their hats. The hats are mixed up during dinner, so that afterward each man receives a random hat. In particular, each man gets his own hat with probability $1/n$. What is the expected number of men who get their own hat?

Letting G be the number of men that get their own hat, we want to find the expectation of G. But all we know about G is that the probability that a man gets his own hat back is $1/n$. There are many different probability distributions of hat permutations with this property, so we don't know enough about the distribution of G to calculate its expectation directly. But linearity of expectation makes the problem really easy.

The trick[3] is to express G as a sum of indicator variables. In particular, let G_i be an indicator for the event that the ith man gets his own hat. That is, $G_i = 1$ if the ith man gets his own hat, and $G_i = 0$ otherwise. The number of men that get their own hat is then the sum of these indicator random variables:

sum b/c easy way to compare

$$G = G_1 + G_2 + \cdots + G_n. \tag{17.9}$$

These indicator variables are *not* mutually independent. For example, if $n - 1$ men all get their own hats, then the last man is certain to receive his own hat. But, since we plan to use linearity of expectation, we don't have worry about independence!

Since G_i is an indicator random variable, we know from Lemma 17.4.2 that

$$\text{Ex}[G_i] = \Pr[G_i = 1] = 1/n. \tag{17.10}$$

[3] We are going to use this trick a lot so it is important to understand it.

By Linearity of Expectation and equation (17.9), this means that

$$\begin{aligned}\mathrm{Ex}[G] &= \mathrm{Ex}[G_1 + G_2 + \cdots + G_n] \\ &= \mathrm{Ex}[G_1] + \mathrm{Ex}[G_2] + \cdots + \mathrm{Ex}[G_n] \\ &= \overbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}^{n} \\ &= 1.\end{aligned}$$

So even though we don't know much about how hats are scrambled, we've figured out that on average, just one man gets his own hat back!

More generally, Linearity of Expectation provides a very good method for computing the expected number of events that will happen. Cool

Theorem 17.4.10. *Given any collection of events $A_1, A_2, \ldots, A_n$, the expected number of events that will occur is*

$$\sum_{i=1}^{n} \Pr[A_i].$$

For example, A_i could be the event that the ith man gets the right hat back. But in general, it could be any subset of the sample space, and we are asking for the expected number of events that will contain a random sample point.

Proof. Define R_i to be the indicator random variable for A_i, where $R_i(\omega) = 1$ if $w \in A_i$ and $R_i(\omega) = 0$ if $w \notin A_i$. Let $R = R_1 + R_2 + \cdots + R_n$. Then

$$\begin{aligned}\mathrm{Ex}[R] &= \sum_{i=1}^{n} \mathrm{Ex}[R_i] && \text{(by Linearity of Expectation)} \\ &= \sum_{i=1}^{n} \Pr[R_i = 1] && \text{(by Lemma 17.4.2)} \\ &= \sum_{i=1}^{n} \Pr[A_i]. && \text{(def of indicator variable)}\end{aligned}$$

So whenever you are asked for the expected number of events that occur, all you have to do is sum the probabilities that each event occurs. Independence is not needed.

another example?

17.4.10 Expectation of a Binomial Distribution

Suppose that we independently flip n biased coins, each with probability p of coming up heads. What is the expected number of heads?

Let J be the random variable denoting the number of heads. Then J has a binomial distribution with parameters n, p, and

$$\Pr[J = k] = \binom{n}{k} k^p (n-k)^{1-p}.$$

Applying Equation 17.3, this means that

$$\begin{aligned} \text{Ex}[J] &= \sum_{k=0}^{n} k \Pr[J = k] \\ &= \sum_{k=0}^{n} k \binom{n}{k} k^p (n-k)^{1-p}. \end{aligned} \tag{17.11}$$

Ouch! This is one nasty looking sum. Let's try another approach.

Since we have just learned about linearity of expectation for sums of indicator random variables, maybe Theorem 17.4.10 will be helpful. But how do we express J as a sum of indicator random variables? It turns out to be easy. Let J_i be the indicator random variable for the ith coin. In particular, define

$$J_i = \begin{cases} 1 & \text{if the } i\text{th coin is heads} \\ 0 & \text{if the } i\text{th coin is tails.} \end{cases}$$

Then the number of heads is simply

$$J = J_1 + J_2 + \cdots + J_n.$$

By Theorem 17.4.10,

$$\begin{aligned} \text{Ex}[J] &= \sum_{i=1}^{n} \Pr[J_i] \\ &= np. \end{aligned} \tag{17.12}$$

That really was easy. If we flip n mutually independent coins, we expect to get pn heads. Hence the expected value of a binomial distribution with parameters n and p is simply pn.

But what if the coins are not mutually independent? It doesn't matter—the answer is still pn because Linearity of Expectation and Theorem 17.4.10 do not assume any independence.

If you are not yet convinced that Linearity of Expectation and Theorem 17.4.10 are powerful tools, consider this: without even trying, we have used them to prove a very complicated identity, namely, combining equations (17.11) and (17.12) gives:

$$\sum_{k=0}^{n} k \binom{n}{k} k^p (n-k)^{1-p} = pn.$$

If you are still not convinced, then take a look at the next problem.

17.4.11 The Coupon Collector Problem

Every time we purchase a kid's meal at Taco Bell, we are graciously presented with a miniature "Racin' Rocket" car together with a launching device which enables us to project our new vehicle across any tabletop or smooth floor at high velocity. Truly, our delight knows no bounds.

There are n different types of Racin' Rocket cars (blue, green, red, gray, etc.). The type of car awarded to us each day by the kind woman at the Taco Bell register appears to be selected uniformly and independently at random. What is the expected number of kid's meals that we must purchase in order to acquire at least one of each type of Racin' Rocket car?

The same mathematical question shows up in many guises: for example, what is the expected number of people you must poll in order to find at least one person with each possible birthday? Here, instead of collecting Racin' Rocket cars, you're collecting birthdays. The general question is commonly called the *coupon collector problem* after yet another interpretation.

A clever application of linearity of expectation leads to a simple solution to the coupon collector problem. Suppose there are five different types of Racin' Rocket cars, and we receive this sequence:

blue green green red blue orange blue orange gray.

Let's partition the sequence into 5 segments:

$$\underbrace{\text{blue}}_{X_0}\quad \underbrace{\text{green}}_{X_1}\quad \underbrace{\text{green}\quad \text{red}}_{X_2}\quad \underbrace{\text{blue}\quad \text{orange}}_{X_3}\quad \underbrace{\text{blue}\quad \text{orange}\quad \text{gray}}_{X_4}.$$

The rule is that a segment ends whenever we get a new kind of car. For example, the middle segment ends when we get a red car for the first time. In this way, we can

why do this?

break the problem of collecting every type of car into stages. Then we can analyze each stage individually and assemble the results using linearity of expectation.

Let's return to the general case where we're collecting n Racin' Rockets. Let X_k be the length of the kth segment. The total number of kid's meals we must purchase to get all n Racin' Rockets is the sum of the lengths of all these segments:

$$T = X_0 + X_1 + \cdots + X_{n-1}$$

why do I never think of this stuff - even after 6.041 - skills are not improving

Now let's focus our attention on X_k, the length of the kth segment. At the beginning of segment k, we have k different types of car, and the segment ends when we acquire a new type. When we own k types, each kid's meal contains a type that we already have with probability k/n. Therefore, each meal contains a new type of car with probability $1 - k/n = (n-k)/n$. Thus, the expected number of meals until we get a new kind of car is $n/(n-k)$ by the Mean Time to Failure rule. This means that

$$\text{Ex}[X_k] = \frac{n}{n-k}.$$

Linearity of expectation, together with this observation, solves the coupon collector problem:

$$\begin{aligned}
\text{Ex}[T] &= \text{Ex}[X_0 + X_1 + \cdots + X_{n-1}] \\
&= \text{Ex}[X_0] + \text{Ex}[X_1] + \cdots + \text{Ex}[X_{n-1}] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \cdots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1} \\
&= n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&= n\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} + \frac{1}{n}\right) \\
&= nH_n \qquad (17.13) \\
&\sim n \ln n.
\end{aligned}$$

Wow! It's those Harmonic Numbers again!

We can use Equation 17.13 to answer some concrete questions. For example, the expected number of die rolls required to see every number from 1 to 6 is:

$$6H_6 = 14.7\ldots.$$

And the expected number of people you must poll to find at least one person with each possible birthday is:

$$365H_{365} = 2364.6\ldots.$$

17.4.12 Infinite Sums

Linearity of expectation also works for an infinite number of random variables provided that the variables satisfy some stringent absolute convergence criteria.

Theorem 17.4.11 (Linearity of Expectation). *Let R_0, $R_1, \ldots,$ be random variables such that*

$$\sum_{i=0}^{\infty} \text{Ex}[|R_i|]$$

converges. Then

$$\text{Ex}\left[\sum_{i=0}^{\infty} R_i\right] = \sum_{i=0}^{\infty} \text{Ex}[R_i].$$

Proof. Let $T ::= \sum_{i=0}^{\infty} R_i$.

We leave it to the reader to verify that, under the given convergence hypothesis, all the sums in the following derivation are absolutely convergent, which justifies rearranging them as follows:

$$\begin{aligned}
\sum_{i=0}^{\infty} \text{Ex}[R_i] &= \sum_{i=0}^{\infty} \sum_{s \in \mathcal{S}} R_i(s) \cdot \Pr[s] && \text{(Def. 17.4.1)} \\
&= \sum_{s \in \mathcal{S}} \sum_{i=0}^{\infty} R_i(s) \cdot \Pr[s] && \text{(exchanging order of summation)} \\
&= \sum_{s \in \mathcal{S}} \left[\sum_{i=0}^{\infty} R_i(s)\right] \cdot \Pr[s] && \text{(factoring out } \Pr[s]) \\
&= \sum_{s \in \mathcal{S}} T(s) \cdot \Pr[s] && \text{(Def. of } T) \\
&= \text{Ex}[T] && \text{(Def. 17.4.1)} \\
&= \text{Ex}[\sum_{i=0}^{\infty} R_i]. && \text{(Def. of } T). \blacksquare
\end{aligned}$$

17.4.13 Expectations of Products

While the expectation of a sum is the sum of the expectations, the same is usually not true for products. For example, suppose that we roll a fair 6-sided die and denote the outcome with the random variable R. Does $\text{Ex}[R \cdot R] = \text{Ex}[R] \cdot \text{Ex}[R]$?

We know that $\text{Ex}[R] = 3\frac{1}{2}$ and thus $\text{Ex}[R]^2 = 12\frac{1}{4}$. Let's compute $\text{Ex}[R^2]$ to see if we get the same result.

$$\begin{aligned}
\text{Ex}[R^2] &= \sum_{\omega \in S} R^2(\omega) \Pr[w] \\
&= \sum_{i=1}^{6} i^2 \cdot \Pr[R_i = i] \\
&= \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} \\
&= 15\ 1/6 \\
&\neq 12\ 1/4.
\end{aligned}$$

That is,

$$\text{Ex}[R \cdot R] \neq \text{Ex}[R] \cdot \text{Ex}[R].$$

So the expectation of a product is not always equal to the product of the expectations.

There is a special case when such a relationship *does* hold however; namely, when the random variables in the product are *independent*.

Theorem 17.4.12. *For any two* independent *random variables R_1, R_2,*

$$\text{Ex}[R_1 \cdot R_2] = \text{Ex}[R_1] \cdot \text{Ex}[R_2].$$

The proof follows by judicious rearrangement of terms in the sum that defines $\text{Ex}[R_1 \cdot R_2]$. Details appear in Problem 17.12.

Theorem 17.4.12 extends routinely to a collection of mutually independent variables.

Corollary 17.4.13 (Expectation of Independent Product). *If random variables $R_1, R_2, \ldots, R_k$ are mutually independent, then*

$$\text{Ex}[\prod_{i=1}^{k} R_i] = \prod_{i=1}^{k} \text{Ex}[R_i].$$

Problems for Section 17.3

Practice Problems

Problem 17.1.

Suppose X_1, X_2, and X_3 are three mutually independent random variables, each having the uniform distribution

$$\Pr[X_i = k] \text{ equal to } 1/3 \text{ for each of } k = 1, 2, 3.$$

Let M be another random variable giving the maximum of these three random variables. What is the density function of M?

Problem 17.2. (a) Prove that if A and B are independent events, then so are A and $\overline{B}$.

(b) Let I_A and I_B be the indicator variables for events A and B. Prove that I_A and I_B are independent iff A and B are independent.
Hint: For any event, E, let $E^1 ::= E$ and $E^0 ::= \overline{E}$.

Class Problems

> Guess the Bigger Number Game
>
> Team 1:
>
> - Write different integers between 0 and 7 on two pieces of paper.
> - Put the papers face down on a table.
>
> Team 2:
>
> - Turn over one paper and look at the number on it.
> - Either stick with this number or switch to the unseen other number.
>
> Team 2 wins if it chooses the larger number.

Problem 17.3.
The analysis in section 17.3.3 implies that Team 2 has a strategy that wins 4/7 of the time no matter how Team 1 plays. Can Team 2 do better? The answer is "no," because Team 1 has a strategy that guarantees that it wins at least 3/7 of the time, no matter how Team 2 plays. Describe such a strategy for Team 1 and explain why it works.

Problem 17.4.
Suppose you have a biased coin that has probability p of flipping heads. Let J be the number of heads in n independent coin flips. So J has the general binomial

distribution:

$$\text{PDF}_J(k) = \binom{n}{k} p^k q^{n-k}$$

where $q ::= 1 - p$.

(a) Show that

$$\begin{aligned} \text{PDF}_J(k) &< \text{PDF}_J(k+1) && \text{for } k < np + p, \\ \text{PDF}_J(k) &> \text{PDF}_J(k+1) && \text{for } k > np + p. \end{aligned}$$

(b) Conclude that the maximum value of PDF_J is asymptotically equal to

$$\frac{1}{\sqrt{2\pi npq}}.$$

Hint: For the asymptotic estimate, it's ok to assume that np is an integer, so by part (a), the maximum value is $\text{PDF}_J(np)$. Use Stirling's formula (14.30).

Homework Problems

Problem 17.5.
A drunken sailor wanders along main street, which conveniently consists of the points along the x axis with integral coordinates. In each step, the sailor moves one unit left or right along the x axis. A particular *path* taken by the sailor can be described by a sequence of "left" and "right" steps. For example, ⟨left,left,right⟩ describes the walk that goes left twice then goes right.

We model this scenario with a random walk graph whose vertices are the integers and with edges going in each direction between consecutive integers. All edges are labelled $1/2$.

The sailor begins his random walk at the origin. This is described by an initial distribution which labels the origin with probability 1 and all other vertices with probability 0. After one step, the sailor is equally likely to be at location 1 or -1, so the distribution after one step gives label 1/2 to the vertices 1 and -1 and labels all other vertices with probability 0.

(a) Give the distributions after the 2nd, 3rd, and 4th step by filling in the table of probabilities below, where omitted entries are 0. For each row, write all the nonzero entries so they have the same denominator.

	location								
	-4	-3	-2	-1	0	1	2	3	4
initially					1				
after 1 step				1/2	0	1/2			
after 2 steps			?	?	?	?	?		
after 3 steps		?	?	?	?	?	?	?	
after 4 steps	?	?	?	?	?	?	?	?	?

(b)

1. What is the final location of a t-step path that moves right exactly i times?
2. How many different paths are there that end at that location?
3. What is the probability that the sailor ends at this location?

(c) Let L be the random variable giving the sailor's location after t steps, and let $B ::= (L+t)/2$. Use the answer to part (b) to show that B has an unbiased binomial density function.

(d) Again let L be the random variable giving the sailor's location after t steps, where t is even. Show that

$$\Pr[|L| < \frac{\sqrt{t}}{2}] < \frac{1}{2}.$$

So there is a better than even chance that the sailor ends up at least $\sqrt{t}/2$ steps from where he started.

Hint: Work in terms of B. Then you can use an estimate that bounds the binomial distribution. Alternatively, observe that the origin is the most likely final location and then use the asymptotic estimate

$$\Pr[L = 0] = \Pr[B = t/2] \sim \sqrt{\frac{2}{\pi t}}.$$

Problems for Section 17.4

Practice Problems

Problem 17.6.
MIT students sometimes delay laundry for a few days. Assume all random values described below are mutually independent.

(a) A *busy* student must complete 3 problem sets before doing laundry. Each problem set requires 1 day with probability $2/3$ and 2 days with probability $1/3$. Let B be the number of days a busy student delays laundry. What is $\text{Ex}[B]$?

Example: If the first problem set requires 1 day and the second and third problem sets each require 2 days, then the student delays for $B = 5$ days.

(b) A *relaxed* student rolls a fair, 6-sided die in the morning. If he rolls a 1, then he does his laundry immediately (with zero days of delay). Otherwise, he delays for one day and repeats the experiment the following morning. Let R be the number of days a relaxed student delays laundry. What is $\text{Ex}[R]$?

Example: If the student rolls a 2 the first morning, a 5 the second morning, and a 1 the third morning, then he delays for $R = 2$ days.

(c) Before doing laundry, an *unlucky* student must recover from illness for a number of days equal to the product of the numbers rolled on two fair, 6-sided dice. Let U be the expected number of days an unlucky student delays laundry. What is $\text{Ex}[U]$?

Example: If the rolls are 5 and 3, then the student delays for $U = 15$ days.

(d) A student is *busy* with probability $1/2$, *relaxed* with probability $1/3$, and *unlucky* with probability $1/6$. Let D be the number of days the student delays laundry. What is $\text{Ex}[D]$?

Problem 17.7.

Each Math for Computer Science final exam will be graded according to a rigorous procedure:

- With probability $\frac{4}{7}$ the exam is graded by a *TA*,with probability $\frac{2}{7}$ it is graded by a *lecturer*, and with probability $\frac{1}{7}$, it is accidentally dropped behind the radiator and arbitrarily given a score of 84.
- *TAs* score an exam by scoring each problem individually and then taking the sum.
  - There are ten true/false questions worth 2 points each. For each, full credit is given with probability $\frac{3}{4}$, and no credit is given with probability $\frac{1}{4}$.
  - There are four questions worth 15 points each. For each, the score is determined by rolling two fair dice, summing the results, and adding 3.

  - The single 20 point question is awarded either 12 or 18 points with equal probability.

- *Lecturers* score an exam by rolling a fair die twice, multiplying the results, and then adding a "general impression"score.

  - With probability $\frac{4}{10}$, the general impression score is 40.
  - With probability $\frac{3}{10}$, the general impression score is 50.
  - With probability $\frac{3}{10}$, the general impression score is 60.

Assume all random choices during the grading process are independent.

(a) What is the expected score on an exam graded by a TA?

(b) What is the expected score on an exam graded by a lecturer?

(c) What is the expected score on a Math for Computer Science final exam?

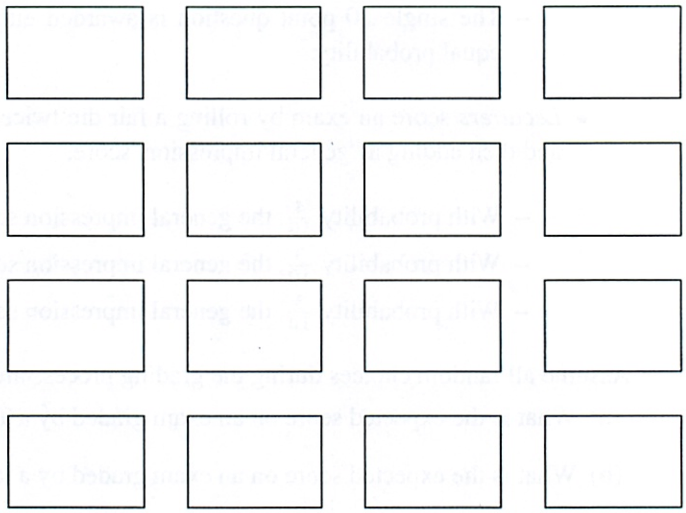
Class Problems

Problem 17.8.
Let's see what it takes to make Carnival Dice fair. Here's the game with payoff parameter k: make three independent rolls of a fair die. If you roll a six

- no times, then you lose 1 dollar.
- exactly once, then you win 1 dollar.
- exactly twice, then you win two dollars.
- all three times, then you win k dollars.

For what value of k is this game fair?

Problem 17.9.
A classroom has sixteen desks arranged as shown below.



If there is a girl in front, behind, to the left, or to the right of a boy, then the two of them *flirt*. One student may be in multiple flirting couples; for example, a student in a corner of the classroom can flirt with up to two others, while a student in the center can flirt with as many as four others. Suppose that desks are occupied by boys and girls with equal probability and mutually independently. What is the expected number of flirting couples? *Hint:* Linearity.

Problem 17.10.
Here are seven propositions:

$$\begin{array}{ccccc}
x_1 & \text{OR} & x_3 & \text{OR} & \overline{x_7} \\
\overline{x_5} & \text{OR} & x_6 & \text{OR} & x_7 \\
x_2 & \text{OR} & \overline{x_4} & \text{OR} & x_6 \\
\overline{x_4} & \text{OR} & x_5 & \text{OR} & \overline{x_7} \\
x_3 & \text{OR} & \overline{x_5} & \text{OR} & \overline{x_8} \\
x_9 & \text{OR} & \overline{x_8} & \text{OR} & x_2 \\
\overline{x_3} & \text{OR} & x_9 & \text{OR} & x_4
\end{array}$$

Note that:

1. Each proposition is the disjunction (OR) of three terms of the form x_i or the form $\overline{x_i}$.
2. The variables in the three terms in each proposition are all different.

Suppose that we assign true/false values to the variables $x_1, \ldots, x_9$ independently and with equal probability.

(a) What is the expected number of true propositions?

Hint: Let T_i be an indicator for the event that the i-th proposition is true.

(b) Use your answer to prove that for *any* set of 7 propositions satisfying the conditions 1. and 2., there is an assignment to the variables that makes all 7 of the propositions true.

Problem 17.11. **(a)** Suppose we flip a fair coin until two Tails in a row come up. What is the expected number, N_{TT}, of flips we perform? *Hint:* Let D be the tree diagram for this process. Explain why $D = H \cdot D + T \cdot (H \cdot D + T)$. Use the Law of Total Expectation 17.4.6

(b) Suppose we flip a fair coin until a Tail immediately followed by a Head come up. What is the expected number, N_{TH}, of flips we perform?

(c) Suppose we now play a game: flip a fair coin until either TT or TH first occurs. You win if TT comes up first, lose if TH comes up first. Since TT takes 50% longer on average to turn up, your opponent agrees that he has the advantage. So you tell him you're willing to play if you pay him \$5 when he wins, but he merely pays you a 20% premium, that is, \$6, when you win.

If you do this, you're sneakily taking advantage of your opponent's untrained intuition, since you've gotten him to agree to unfair odds. What is your expected profit per game?

Problem 17.12.
Justify each line of the following proof that if R_1 and R_2 are *independent*, then

$$\text{Ex}[R_1 \cdot R_2] = \text{Ex}[R_1] \cdot \text{Ex}[R_2].$$

Proof.

$$
\begin{aligned}
&\mathrm{Ex}[R_1 \cdot R_2] \\
&= \sum_{r \in \mathrm{range}(R_1 \cdot R_2)} r \cdot \Pr[R_1 \cdot R_2 = r] \\
&= \sum_{r_i \in \mathrm{range}(R_i)} r_1 r_2 \cdot \Pr[R_1 = r_1 \text{ and } R_2 = r_2] \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \sum_{r_2 \in \mathrm{range}(R_2)} r_1 r_2 \cdot \Pr[R_1 = r_1 \text{ and } R_2 = r_2] \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \sum_{r_2 \in \mathrm{range}(R_2)} r_1 r_2 \cdot \Pr[R_1 = r_1] \cdot \Pr[R_2 = r_2] \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \left(r_1 \Pr[R_1 = r_1] \cdot \sum_{r_2 \in \mathrm{range}(R_2)} r_2 \Pr[R_2 = r_2] \right) \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} r_1 \Pr[R_1 = r_1] \cdot \mathrm{Ex}[R_2] \\
&= \mathrm{Ex}[R_2] \cdot \sum_{r_1 \in \mathrm{range}(R_1)} r_1 \Pr[R_1 = r_1] \\
&= \mathrm{Ex}[R_2] \cdot \mathrm{Ex}[R_1].
\end{aligned}
$$

■

Problem 17.13.
A *literal* is a propositional variable or its negation. A k*-clause* is an OR of k literals, with no variable occurring more than once in the clause. For example,

$$P \text{ OR } \overline{Q} \text{ OR } \overline{R} \text{ OR } V,$$

is a 4-clause, but

$$\overline{V} \text{ OR } \overline{Q} \text{ OR } \overline{X} \text{ OR } V,$$

is not, since V appears twice.

Let $\mathcal{S}$ be a set of n distinct k-clauses involving v variables. The variables in different k-clauses may overlap or be completely different, so $k \le v \le nk$.

A random assignment of true/false values will be made independently to each of the v variables, with true and false assignments equally likely. Write formulas in n, k, and v in answer to the first two parts below.

(a) What is the probability that the last k-clause in $\mathcal{S}$ is true under the random assignment?

(b) What is the expected number of true k-clauses in $\mathcal{S}$?

(c) A set of propositions is *satisfiable* iff there is an assignment to the variables that makes all of the propositions true. Use your answer to part (b) to prove that if $n < 2^k$, then $\mathcal{S}$ is satisfiable.

Problem 17.14.
A gambler bets \$10 on "red" at a roulette table (the odds of red are 18/38 which slightly less than even) to win \$10. If he wins, he gets back twice the amount of his bet and he quits. Otherwise, he doubles his previous bet and continues.

(a) What is the expected number of bets the gambler makes before he wins?

(b) What is his probability of winning?

(c) What is his expected final profit (amount won minus amount lost)?

(d) The fact that the gambler's expected profit is positive, despite the fact that the game is biased against him, is known as the *St. Petersberg paradox*. The paradox arises from an unrealistic, implicit assumption about the gambler's money. Explain.

Hint: What is the expected size of his last bet?

Homework Problems

Problem 17.15.
Let R and S be independent random variables, and f and g be any functions such that $\text{domain}(f) = \text{codomain}(R)$ and $\text{domain}(g) = \text{codomain}(S)$. Prove that $f(R)$ and $g(S)$ are independent random variables. *Hint:* The event $[f(R) = a]$ is the disjoint union of all the events $[R = r]$ for r such that $f(r) = a$.