

15 Conditional Probability

15.1 Definition

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 15.1 and asking how likely that point is to fall into region A or B .

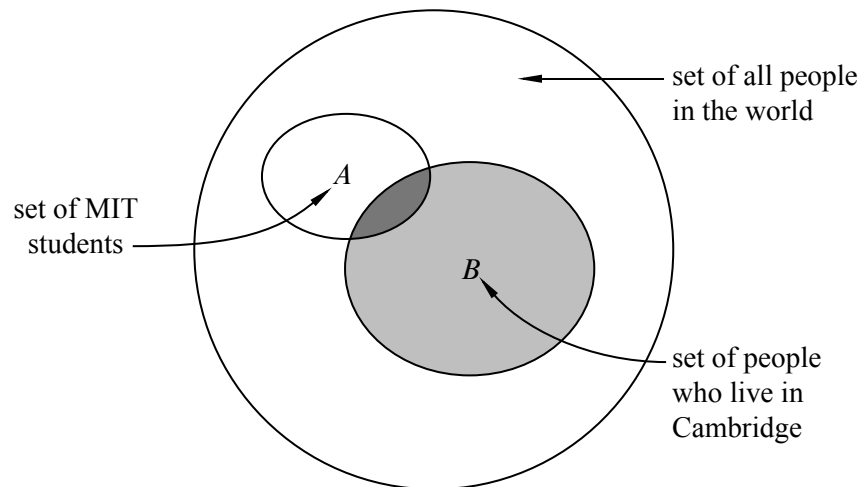


Figure 15.1 Selecting a random person. A is the event that the person is an MIT student. B is the event that the person lives in Cambridge.

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?

- ◀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 15.1) divided by the probability that the person is in set B (the lightly shaded region). This motivates the definition of conditional probability:

Definition 15.1.1.

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

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 formal reasoning and not intuition. The four-step method will also be very helpful as we will see in the next examples.

15.2 Using the Four-Step Method to Determine Conditional Probability

15.2.1 The “Halting Problem”

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.

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?

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 $\Pr[A | B]$.

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 15.2.

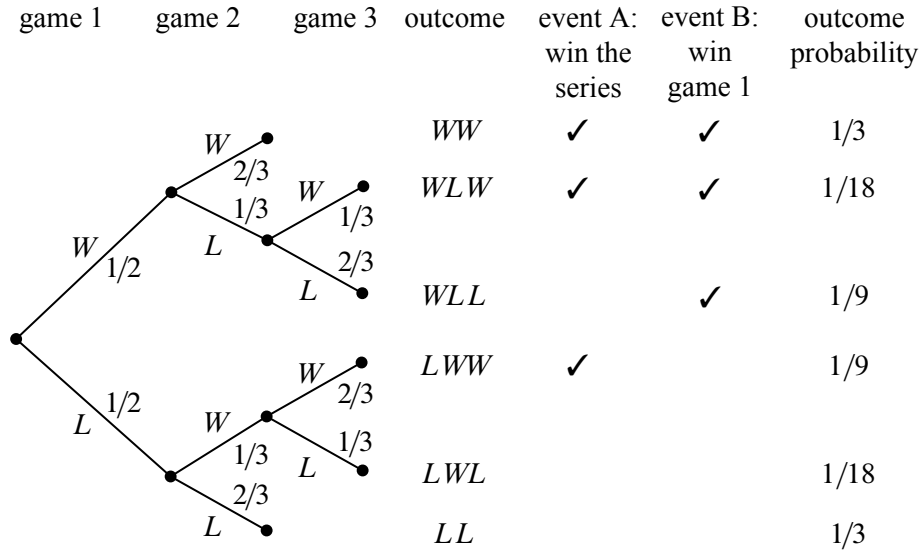


Figure 15.2 The tree diagram for computing the probability that the “Halting Problem” wins two out of three games given that they won the first game.

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 (la-

beled L). The complete sample space is:

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

Step 2: Define Events of Interest

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

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

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 15.2.

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:

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

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 $7/9$.

15.2.2 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?

The answer goes back to Definition 15.1.1 of conditional probability which could be written in a form called the *Product Rule* for probabilities:

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

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2 \mid E_1] \left($$

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

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

So the Product Rule is the formal justification for multiplying edge 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 for n Events).

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

provided that

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

This rule follows from the definition of conditional probability and induction on n .

15.2.3 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. 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 15.3.

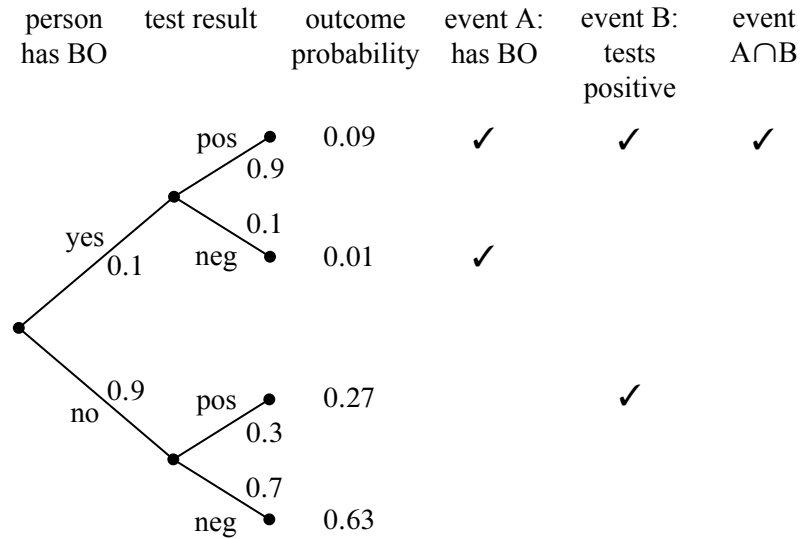


Figure 15.3 The tree diagram for the BO problem.

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 $\Pr[A | B]$, 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 15.3.

Step 4: Compute Event Probabilities

From Definition 15.1.1, we have

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

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 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!

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!

15.3 *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.

15.3.1 The “Halting Problem,” in Reverse

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

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 $\Pr[B | A]$. (We can compute this using the definition of conditional probability and the tree diagram in Figure 15.2:

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

This answer is suspicious! In the preceding section, we showed that $\Pr[A | B]$ was also $7/9$. Could it be true that $\Pr[A | B] = \Pr[B | A]$ 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.

Let’s work out the general conditions under which $\Pr[A | B] = \Pr[B | A]$. By the definition of conditional probability, this equation holds if and 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

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

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:

Theorem 15.3.1 (Bayes’ Rule). *If $\Pr[A]$ and $\Pr[B]$ are nonzero, then:*

$$\Pr[B | A] = \frac{\Pr[A | B] \Pr[B]}{\Pr[A]} \tag{15.1}$$

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

$$\Pr[A | B] \Pr[B] = \Pr[A \cap B] = \Pr[B | A] \Pr[A]$$

by definition of conditional probability. Dividing by $\Pr[A]$ gives (15.1). ■

Next, let’s look at a problem that even bothers us.

15.3.2 A Coin Problem

Suppose that someone hands you either a fair coin or a trick coin with heads on both sides. You flip the coin 100 times and see heads every time. What can you say about the probability that you flipped the fair coin? Remarkably, nothing!

In order to make sense out of this outrageous claim, let’s formalize the problem. The sample space is worked out in the tree diagram shown in Figure 15.4. We do not know the probability p that you were handed the fair coin initially—you were just given one coin or the other. Let A be the event that you were handed the

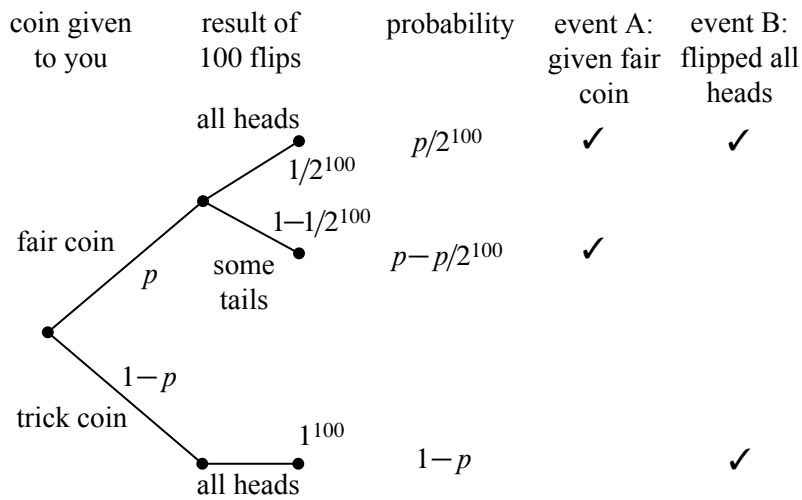


Figure 15.4 The tree diagram for the coin-flipping problem.

fair coin, and let B be the event that you flipped 100 straight heads. We’re looking for $\Pr[A | B]$, the probability that you were handed the fair coin, given that you flipped 100 heads. The outcome probabilities are worked out in Figure 15.4. Plugging the results into the definition of conditional probability gives:

$$\begin{aligned} \Pr[A | B] &= \frac{\Pr[A \cap B]}{\Pr[B]} \\ &= \frac{p/2^{100}}{1 - p + p/2^{100}} \\ &= \frac{p}{2^{100}(1 - p) + p}. \end{aligned}$$

This expression is very small for moderate values of p because of the 2^{100} term in the denominator. For example, if $p = 1/2$, then the probability that you were given the fair coin is essentially zero.

But we *do not know* the probability p that you were given the fair coin. And perhaps the value of p is *not* moderate; in fact, maybe $p = 1 - 2^{-100}$. Then there is nearly an even chance that you have the fair coin, given that you flipped 100 heads. In fact, maybe you were handed the fair coin with probability $p = 1$. Then the probability that you were given the fair coin is, well, 1!

Of course, it is extremely unlikely that you would flip 100 straight heads, but in this case, that is a given from the assumption of the conditional probability. And so if you really did see 100 straight heads, it would be very tempting to also assume that p is not close to 1 and hence that you are very likely to have flipped the trick coin.

We will encounter a very similar issue when we look at methods for estimation by sampling in Section 17.5.5.

15.4 Conditional Identities

15.4.1 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 $A \cap \bar{E}$. By the Sum Rule, the sum of these probabilities equals $\Pr[A]$. Expressing the intersection probabilities as conditional probabilities yields:

Rule 15.4.1 (Law of Total Probability, single event). *If $\Pr[E]$ and $\Pr[\bar{E}]$ are nonzero, then*

$$\Pr[A] = \Pr[A | E] \Pr[E] + \Pr[A | \bar{E}] \Pr[\bar{E}].$$

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, $\Pr[E] = \Pr[\bar{E}] = 1/2$. There are now two cases. If we flip heads, then we roll a 2 on a single die with probability $\Pr[A | E] = 1/6$. On the other hand, if we flip tails, then we get a sum of 2 on two dice with probability $\Pr[A | \bar{E}] = 1/36$. 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 15.4.2 (Law of Total Probability). *If E_1, \dots, E_n are disjoint events whose union is the whole sample space, then:*

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

15.4.2 Conditioning on a Single Event

The probability rules that we derived in Chapter 14 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 :

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

This follows from the fact that if $\Pr[C] \neq 0$, then

$$\begin{aligned} \Pr[A \cup B \mid C] &= \frac{\Pr[(A \cup B) \cap C]}{\Pr[C]} \\ &= \frac{\Pr[(A \cap C) \cup (B \cap C)]}{\Pr[C]} \\ &= \frac{\Pr[A \cap C] + \Pr[B \cap C] - \Pr[A \cap B \cap C]}{\Pr[C]} \\ &= \Pr[A \mid C] + \Pr[B \mid C] - \Pr[A \cap B \mid C] \end{aligned}$$

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.

$$\Pr[A \mid B \cup C] = \Pr[A \mid B] + \Pr[A \mid C] - \Pr[A \mid B \cap C] \tag{15.2}$$

A counterexample is shown in Figure 15.5. 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 15.2 does not hold.

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

15.4.3 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

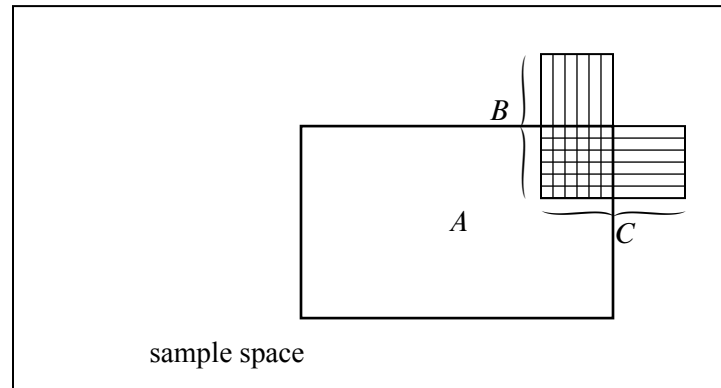


Figure 15.5 A counterexample to Equation 15.2. Event A is the gray rectangle, event B is the rectangle with vertical stripes, and event C is the rectangle with horizontal stripes. $B \cap C$ lies entirely within A while $B - C$ and $C - B$ are entirely outside of A .

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!

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.

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

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

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

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

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

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 follows:

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

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 (15.2). This argument is bogus! Maybe the two parties do not hold contradictory positions after all!

In fact, Table 15.1 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!

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

6.042J / 18.062J Mathematics for Computer Science
Fall 2010

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