

## Section 8.2 Conditional Probability and Bayes Theorem

Often it is required to compute the probability of an event given that another event has occurred. We call that **conditional probability**.

### Conditional Probability

The probability the event  $B$  occurs, given that event  $A$  has happened, is represented as  $P(B | A)$

This is read as “the probability of  $B$  given  $A$ ”

### Example 1

What is the probability that two cards drawn at random from a deck of playing cards will both be aces?

It might seem that you could use the formula for the probability of two independent events and simply multiply  $\frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$ . This would be incorrect, however, because the two events are not independent. If the first card drawn is an ace, then the probability that the second card is also an ace would be lower because there would only be three aces left in the deck.

Once the first card chosen is an ace, the probability that the second card chosen is also an ace is called the **conditional probability** of drawing an ace. In this case the "condition" is that the first card is an ace. Symbolically, we write this as:

$P(\text{ace on second draw} | \text{an ace on the first draw})$ .

The vertical bar "|" is read as "given," so the above expression is short for "The probability that an ace is drawn on the second draw given that an ace was drawn on the first draw."

What is this probability? After an ace is drawn on the first draw, there are 3 aces out of 51 total cards left. This means that the conditional probability of drawing an ace after one ace

has already been drawn is  $\frac{3}{51} = \frac{1}{17}$ .

Thus, the probability of both cards being aces is  $\frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}$ .

### Example 2

Find the probability that a die rolled shows a 6, given that a flipped coin shows a head.

These are two independent events, so the probability of the die rolling a 6 is  $\frac{1}{6}$ , regardless of the result of the coin flip.

## Example 3

The table below shows the number of survey subjects who have received and not received a speeding ticket in the last year, and the color of their car. Find the probability that a randomly chosen person:

- a) Has a speeding ticket *given* they have a red car  
 b) Has a red car *given* they have a speeding ticket

	Speeding ticket	No speeding ticket	Total
Red car	15	135	150
Not red car	45	470	515
Total	60	605	665

- a) Since we know the person has a red car, we are only considering the 150 people in the first row of the table. Of those, 15 have a speeding ticket, so

$$P(\text{ticket} \mid \text{red car}) = \frac{15}{150} = \frac{1}{10} = 0.1$$

- b) Since we know the person has a speeding ticket, we are only considering the 60 people in the first column of the table. Of those, 15 have a red car, so

$$P(\text{red car} \mid \text{ticket}) = \frac{15}{60} = \frac{1}{4} = 0.25.$$

Notice from the last example that  $P(B \mid A)$  is **not** equal to  $P(A \mid B)$ .

These kinds of conditional probabilities are what insurance companies use to determine your insurance rates. They look at the conditional probability of you having accident, given your age, your car, your car color, your driving history, etc., and price your policy based on that likelihood.

## Conditional Probability Formula

If Events  $A$  and  $B$  are not independent, then

$$P(A \text{ and } B) = P(A) \cdot P(B \mid A)$$

## Example 4

If you pull 2 cards out of a deck, what is the probability that both are spades?

The probability that the first card is a spade is  $\frac{13}{52}$ .

The probability that the second card is a spade, given the first was a spade, is  $\frac{12}{51}$ , since there is one less spade in the deck, and one less total cards.

The probability that both cards are spades is  $\frac{13}{52} \cdot \frac{12}{51} = \frac{156}{2652} \approx 0.0588$

## Example 5

If you draw two cards from a deck, what is the probability that you will get the Ace of Diamonds and a black card?

You can satisfy this condition by having Case A or Case B, as follows:

Case A) you can get the Ace of Diamonds first and then a black card or

Case B) you can get a black card first and then the Ace of Diamonds.

Let's calculate the probability of Case A. The probability that the first card is the Ace of Diamonds is  $\frac{1}{52}$ . The probability that the second card is black given that the first card is

the Ace of Diamonds is  $\frac{26}{51}$  because 26 of the remaining 51 cards are black. The

probability is therefore  $\frac{1}{52} \cdot \frac{26}{51} = \frac{1}{102}$ .

Now for Case B: the probability that the first card is black is  $\frac{26}{52} = \frac{1}{2}$ . The probability that

the second card is the Ace of Diamonds given that the first card is black is  $\frac{1}{51}$ . The

probability of Case B is therefore  $\frac{1}{2} \cdot \frac{1}{51} = \frac{1}{102}$ , the same as the probability of Case 1.

Recall that the probability of A or B is  $P(A) + P(B) - P(A \text{ and } B)$ . In this problem,  $P(A \text{ and } B) = 0$  since the first card cannot be the Ace of Diamonds and be a black card.

Therefore, the probability of Case A or Case B is  $\frac{1}{102} + \frac{1}{102} = \frac{2}{102} = \frac{1}{51}$ . The probability

that you will get the Ace of Diamonds and a black card when drawing two cards from a deck is  $\frac{1}{51}$ .

## Try it Now

1. In your drawer you have 10 pairs of socks, 6 of which are white. If you reach in and randomly grab two pairs of socks, what is the probability that both are white?

## Example 6

A home pregnancy test was given to women, then pregnancy was verified through blood tests. The following table shows the home pregnancy test results. Find

a)  $P(\text{not pregnant} \mid \text{positive test result})$

b)  $P(\text{positive test result} \mid \text{not pregnant})$

	Positive test	Negative test	Total
Pregnant	70	4	74
Not Pregnant	5	14	19
Total	75	18	93

- a) Since we know the test result was positive, we're limited to the 75 women in the first column, of which 5 were not pregnant.  $P(\text{not pregnant} \mid \text{positive test result}) = \frac{5}{75} \approx 0.067$ .
- b) Since we know the woman is not pregnant, we are limited to the 19 women in the second row, of which 5 had a positive test.  $P(\text{positive test result} \mid \text{not pregnant}) = \frac{5}{19} \approx 0.263$

The second result is what is usually called a false positive: A positive result when the woman is not actually pregnant.

### Bayes Theorem

Bayes Theorem is a formulaic approach to complex conditional probability problems like the last example. However, using the formula is itself complicated, so we will focus on a more intuitive approach.

#### Example 7

Suppose a certain disease has an incidence rate of 0.1% (that is, it afflicts 0.1% of the population). A test has been devised to detect this disease. The test does not produce false negatives (that is, anyone who has the disease will test positive for it), but the false positive rate is 5% (that is, about 5% of people who take the test will test positive, even though they do not have the disease). Suppose a randomly selected person takes the test and tests positive. What is the probability that this person actually has the disease?

There are two ways to approach the solution to this problem. One involves an important result in probability theory called Bayes' theorem. We will discuss this theorem a bit later, but for now we will use an alternative and, we hope, much more intuitive approach.

Let's break down the information in the problem piece by piece.

*Suppose a certain disease has an incidence rate of 0.1% (that is, it afflicts 0.1% of the population).* The percentage 0.1% can be converted to a decimal number by moving the decimal place two places to the left, to get 0.001. In turn, 0.001 can be rewritten as a fraction: 1/1000. This tells us that about 1 in every 1000 people has the disease. (If we wanted we could write  $P(\text{disease})=0.001$ .)

*A test has been devised to detect this disease. The test does not produce false negatives (that is, anyone who has the disease will test positive for it). This part is fairly straightforward: everyone who has the disease will test positive, or alternatively everyone who tests negative does not have the disease. (We could also say  $P(\text{positive} | \text{disease})=1$ .)*

*The false positive rate is 5% (that is, about 5% of people who take the test will test positive, even though they do not have the disease). This is even more straightforward. Another way of looking at it is that of every 100 people who are tested and do not have the disease, 5 will test positive even though they do not have the disease. (We could also say that  $P(\text{positive} | \text{no disease})=0.05$ .)*

*Suppose a randomly selected person takes the test and tests positive. What is the probability that this person actually has the disease? Here we want to compute  $P(\text{disease} | \text{positive})$ . We already know that  $P(\text{positive} | \text{disease})=1$ , but remember that conditional probabilities are not equal if the conditions are switched.*

Rather than thinking in terms of all these probabilities we have developed, let's create a hypothetical situation and apply the facts as set out above. First, suppose we randomly select 1000 people and administer the test. How many do we expect to have the disease? Since about 1/1000 of all people are afflicted with the disease, 1/1000 of 1000 people is 1. (Now you know why we chose 1000.) Only 1 of 1000 test subjects actually has the disease; the other 999 do not.

We also know that 5% of all people who do not have the disease will test positive. There are 999 disease-free people, so we would expect  $(0.05)(999)=49.95$  (so, about 50) people to test positive who do not have the disease.

Now back to the original question, computing  $P(\text{disease} | \text{positive})$ . There are 51 people who test positive in our example (the one unfortunate person who actually has the disease, plus the 50 people who tested positive but don't). Only one of these people has the disease, so

$$P(\text{disease} | \text{positive}) \approx \frac{1}{51} \approx 0.0196$$

or less than 2%. Does this surprise you? This means that of all people who test positive, over 98% *do not have the disease*.

The answer we got was slightly approximate, since we rounded 49.95 to 50. We could redo the problem with 100,000 test subjects, 100 of whom would have the disease and  $(0.05)(99,900)=4995$  test positive but do not have the disease, so the exact probability of having the disease if you test positive is

$$P(\text{disease} | \text{positive}) \approx \frac{100}{5095} \approx 0.0196$$

which is pretty much the same answer.

But back to the surprising result. *Of all people who test positive, over 98% do not have the disease.* If your guess for the probability a person who tests positive has the disease was wildly different from the right answer (2%), don't feel bad. The exact same problem was posed to doctors and medical students at the Harvard Medical School 25 years ago and the results revealed in a 1978 *New England Journal of Medicine* article. Only about 18% of the participants got the right answer. Most of the rest thought the answer was closer to 95% (perhaps they were misled by the false positive rate of 5%).

So at least you should feel a little better that a bunch of doctors didn't get the right answer either (assuming you thought the answer was much higher). But the significance of this finding and similar results from other studies in the intervening years lies not in making math students feel better but in the possibly catastrophic consequences it might have for patient care. If a doctor thinks the chances that a positive test result nearly guarantees that a patient has a disease, they might begin an unnecessary and possibly harmful treatment regimen on a healthy patient. Or worse, as in the early days of the AIDS crisis when being HIV-positive was often equated with a death sentence, the patient might take a drastic action and commit suicide.

As we have seen in this hypothetical example, the most responsible course of action for treating a patient who tests positive would be to counsel the patient that they most likely do *not* have the disease and to order further, more reliable, tests to verify the diagnosis.

One of the reasons that the doctors and medical students in the study did so poorly is that such problems, when presented in the types of statistics courses that medical students often take, are solved by use of Bayes' theorem, which is stated as follows:

#### Bayes' Theorem

$$P(A | B) = \frac{P(A)P(B | A)}{P(A)P(B | A) + P(\bar{A})P(B | \bar{A})}$$

In our earlier example, this translates to

$$P(\text{disease} | \text{positive}) = \frac{P(\text{disease})P(\text{positive} | \text{disease})}{P(\text{disease})P(\text{positive} | \text{disease}) + P(\text{no disease})P(\text{positive} | \text{no disease})}$$

Plugging in the numbers gives

$$P(\text{disease} | \text{positive}) = \frac{(0.001)(1)}{(0.001)(1) + (0.999)(0.05)} \approx 0.0196$$

which is exactly the same answer as our original solution.

The problem is that you (or the typical medical student, or even the typical math professor) are much more likely to be able to remember the original solution than to remember Bayes' theorem. Psychologists, such as Gerd Gigerenzer, author of *Calculated Risks: How to Know When Numbers Deceive You*, have advocated that the method involved in the original solution (which Gigerenzer calls the method of "natural frequencies") be employed in place

of Bayes' Theorem. Gigerenzer performed a study and found that those educated in the natural frequency method were able to recall it far longer than those who were taught Bayes' theorem. When one considers the possible life-and-death consequences associated with such calculations it seems wise to heed his advice.

### Example 8

A certain disease has an incidence rate of 2%. If the false negative rate is 10% and the false positive rate is 1%, compute the probability that a person who tests positive actually has the disease.

Imagine 10,000 people who are tested. Of these 10,000, 200 will have the disease; 10% of them, or 20, will test negative and the remaining 180 will test positive. Of the 9800 who do not have the disease, 1% of them, or 98, will test positive.

	Positive test	Negative test	Total
Have disease	180	20	200
Do not have disease	98	9702	9,800
Total	278	9822	10,000

So of the 278 total people who test positive, 180 will have the disease. Thus

$$P(\text{disease} \mid \text{positive}) = \frac{180}{278} \approx 0.647$$

so about 65% of the people who test positive will have the disease.

Using Bayes theorem directly would give the same result:

$$P(\text{disease} \mid \text{positive}) = \frac{(0.02)(0.90)}{(0.02)(0.90) + (0.98)(0.01)} = \frac{0.018}{0.0278} \approx 0.647$$

### Example 9

A company has found that 80% of its new management hires are meeting expectations, while 20% are not. Of the satisfactory hires, 75% had sales experience, while of the unsatisfactory hires, 55% had sales experience. What is the probability that a new hire with sales experience will meet expectations?

We can imagine 100 new hires. Of them, 80%, or 80, will meet expectations, and 20 will not. Of the 80 who meet expectations, 75%, or 60, had sales experience, and 20 did not. Of the 20 who did not meet expectations, 55%, or 11, had sales experience, and 9 did not.

Summarizing that in a table:

	Sales Experience	No Sales Experience	Total
Meeting expectations	60	20	80
Not Meeting expectations	11	9	20
Total	71	29	100

Now we can answer the question.

$$P(\text{meet expectations} \mid \text{sales experience}) = \frac{60}{71} \approx 0.845$$

So about 84.5% of new hires with sales experience will meet expectations.

### Try it Now

2. A certain disease has an incidence rate of 0.5%. If there are no false negatives and if the false positive rate is 3%, compute the probability that a person who tests positive actually has the disease.

### Important Topics of this Section

Conditional probability  
Probability of “and” for conditional events  
Bayes Theorem

### Try it Now Answers

$$1. \frac{6}{10} \cdot \frac{5}{9} = \frac{30}{90} = \frac{1}{3}$$

2. Out of 100,000 people, 500 would have the disease. Of those, all 500 would test positive. Of the 99,500 without the disease, 2,985 would falsely test positive and the other 96,515 would test negative.

$$P(\text{disease} \mid \text{positive}) = \frac{500}{500 + 2985} = \frac{500}{3485} \approx 14.3\%$$