Chapter 5: Probability: What are the Chances?

Section 5.3

Conditional Probability and Independence

What is Conditional Probability?

The probability we assign to an event can change if we know that some other event has occurred. This idea is the key to many applications of probability.

When we are trying to find the probability that one event will happen under the *condition* that some other event is already known to have occurred, we are trying to determine a **conditional probability.**

Definition:

The probability that one event happens given that another event is already known to have happened is called a **conditional probability.** Suppose we know that event *A* has happened. Then the probability that event *B* happens *given* that event *A* has happened is denoted by $P(B \mid A)$.

Read | as "given that" or "under the condition that"

Calculating Conditional Probabilities

To find the conditional probability $P(A \mid B)$, use the formula

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

The conditional probability $P(B \mid A)$ is given by

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

Example 1: Here is the two-way table from the pierced ears example.

	Gender	
Pierced Ears?	Male	Female
Yes	19	84
No	71	4

1. *P*(is male given has pierced ears) = 19/103, or about 18.4%

2. If we know that a randomly selected student is male, what's the probability that the student has pierced ears? This time, our attention is focused on the males in the class. Because 19 of the 90 males in the class have pierced ears,

P(has pierced ears given is male) = 19/90, or about 21.1%

These two questions sound alike, but they're asking about two very different things.

Example 2: Students at the University of New Harmony received 10,000 course grades last semester. The two-way table below breaks down these grades by which school of the university taught the course. The schools are Liberal Arts, Engineering and Physical Sciences, and Health and Human Services.

Define events:

E: the grade comes from an EPS course, and

L: the grade is lower than a B.

	Grade Level			
School	A	B	Below B	Total
Liberal Arts	2,142	1,890	2,268	6300
Engineering and Physical Sciences	368	432	800	1600
Health and Human Services	882	630	588	2100
Total	3392	2952	3656	10000

a) Find *P*(*L*)

P(*L*) = 3656 / 10000 = 0.3656

b) Find *P*(*E* | *L*)

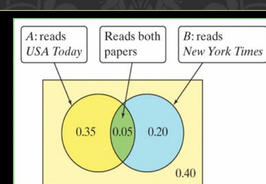
P(*E* | *L*) = 800 / 3656 = 0.2188

c) Find *P*(*L* | *E*)

P(*L*| *E*) = 800 / 1600 = 0.5000

Example 3: Let's classify the residents of a large apartment complex based on the events A: reads USA Today, and B: reads the New York Times. The completed Venn diagram is shown below:

What's the probability that a randomly selected resident who reads USA Today also reads the New York Times?



We want to find *P*(reads New York Times | reads USA Today).

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

Because $P(B \cap A) = 0.05$ and $P(A) = 0.40$, we have $P(B|A) = \frac{0.05}{0.40} = 0.125$

There's a 12.5% chance that a randomly selected resident who reads *USA Today* also reads the *New York Times*.

General Multiplication Rule

How do we find the probability that both events happen, $P(A \text{ and } B)$? Start with the conditional probability formula

This formula is known as the **general multiplication rule**.

P(A and B) = P(A) x P(B)

General Multiplication Rule

Example 4: The Pew Internet and American Life Project finds that 93% of teenagers (ages 12 to 17) use the Internet, and that 55% of online teens have created a profile on a social-networking site. Find the probability that a randomly selected teen uses the Internet and has created a profile. Show your work.

We know that *P*(online) = 0.93 and *P*(profile | online) = 0.55. Use the general multiplication rule:

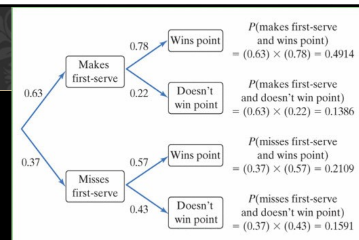
.93 x .55 = .51

There is about a 51% chance that a randomly selected teen uses the Internet and has created a profile on a social-networking site.

Tree Diagrams

Tennis great Roger Federer made 63% of his first serves in the 2011 season. When Federer made his first serve, he won 78% of the points. When Federer missed his first serve and had to serve again, he won only 57% of the points. Suppose we randomly choose a point on which Federer served.

The figure on the next slide shows a tree diagram for this chance process. There are only two possible outcomes on Federer's first serve, a make or a miss. The first set of branches in the tree diagram displays these outcomes with their probabilities. The second set of branches shows the two possible results of the point for Federer—win or lose—and the chance of each result based on the outcome of the first serve. Note that the probabilities on the second set of branches are conditional probabilities, like *P*(win point | make first serve) = 0.78.



What's the probability that Federer makes the first serve and wins the point? From the tree diagram, Federer makes 63% of his first serves. Of this 63%, he wins the point 78% of the time. Because 78% of 63% = (0.63)(0.78) = 0.4914, Federer makes his first serve and wins the point about 49.14% of the time.

The previous calculation amounts to multiplying probabilities along the branches of the tree diagram. Why does this work? The general multiplication rule provides the answer:

$$P(\text{make first serve and win point}) = P(\text{make first serve}) \cdot P(\text{win point} \mid \text{make first serve}) = (0.63)(0.78) = 0.4914$$

When Federer is serving, what's the probability that he wins the point? From the tree diagram, there are two ways Federer can win the point. He can make the first serve and win the point, or he can miss the first serve and win the point. Because these outcomes are mutually exclusive,

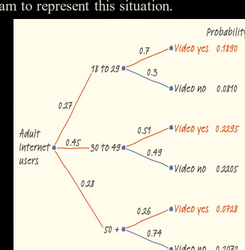
$$P(\text{win point}) = P(\text{makes first serve and wins point}) + P(\text{misses first serve and wins point}) = 0.4914 + 0.2109 = 0.7023$$

Federer wins about 70% of the points when he is serving.

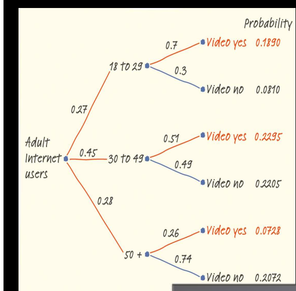
Example 5: Video-sharing sites, led by YouTube, are popular destinations on the Internet. Let's look only at adult Internet users, aged 18 and over. About 27% of adult Internet users are 18 to 29 years old, another 45% are 30 to 49 years old, and the remaining 28% are 50 and over. The Pew Internet and American Life Project finds that 70% of Internet users aged 18 to 29 have visited a video-sharing site, along with 51% of those aged 30 to 49 and 26% of those 50 or older. Do most Internet users visit YouTube and similar sites?

Suppose we select an adult Internet user at random.

a) Draw a tree diagram to represent this situation.



b) Find the probability that this person has visited a video-sharing site. Show your work.



$P(\text{video yes}) = 0.1890 + 0.2295 + 0.0728 = 0.4913$

About 49% of all adult Internet users have visited a video-sharing site.

c) Given that this person has visited a video-sharing site, find the probability that he or she is aged 18 to 29. Show your work.

$$P(\text{18 to 29} \mid \text{visits video-sharing site}) = \frac{P(\text{18 to 29 and visits video-sharing site})}{P(\text{visits video-sharing site})} = \frac{0.1890}{0.4913} = 0.3847$$

Given that the person visits video-sharing sites, there is about a 38.5% chance that he or she is aged 18 to 29.

Example 6: Many women choose to have annual mammograms to screen for breast cancer after age 40. A mammogram isn't foolproof. Sometimes, the test suggests that a woman has breast cancer when she really doesn't (a "false positive"). Other times, the test says that a woman doesn't have breast cancer when she actually does (a "false negative").

Suppose that we know the following information about breast cancer and mammograms in a particular region:

*One percent of the women aged 40 or over in this region have breast cancer.

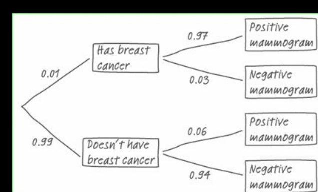
*For women who have breast cancer, the probability of a negative mammogram is 0.03.

*For women who don't have breast cancer, the probability of a positive mammogram is 0.06.

A randomly selected woman aged 40 or over from this region tests positive for breast cancer in a mammogram. Find the probability that she actually has breast cancer. Show your work.

$$P(\text{breast cancer} \mid \text{positive mammogram}) = \frac{P(\text{breast cancer and positive mammogram})}{P(\text{positive mammogram})}$$

The tree diagram below summarizes the situation. Because 1% of women in this region have breast cancer, 99% don't. Of those women who do have breast cancer, 3% would test negative on a mammogram. The remaining 97% would (correctly) test positive. Among the women who don't have breast cancer, 6% would test positive on a mammogram. The remaining 94% would (correctly) test negative.



$$P(\text{breast cancer} \mid \text{positive mammogram}) = \frac{P(\text{breast cancer and positive mammogram})}{P(\text{positive mammogram})}$$

We want to find P(breast cancer | positive mammogram). By the conditional probability formula,

$$P(\text{breast cancer} \mid \text{positive mammogram}) = \frac{P(\text{breast cancer and positive mammogram})}{P(\text{positive mammogram})}$$

To find P(breast cancer and positive mammogram), we use the general multiplication rule along with the information displayed in the tree diagram:

$$P(\text{breast cancer and positive mammogram}) = P(\text{breast cancer}) \cdot P(\text{positive mammogram} \mid \text{breast cancer}) = (0.01)(0.97) = 0.0097$$

To find P(positive mammogram), we need to calculate the probability that a randomly selected woman aged 40 or over from this region gets a positive mammogram. There are two ways this can happen: (1) if the woman has breast cancer and the test result is positive, and (2) if the woman doesn't have cancer, but the mammogram gives a false positive. From the tree diagram, the desired probability is

P(positive mammogram) = (0.01)(0.97) + (0.99)(0.06) = 0.0691

Using these two results, we can find the conditional probability:

$$P(\text{breast cancer} \mid \text{positive mammogram}) = \frac{P(\text{breast cancer and positive mammogram})}{P(\text{positive mammogram})} = \frac{0.0097}{0.0691} = 0.14$$

Given that a randomly selected woman from the region has a positive mammogram, there is only about a 14% chance that she actually has breast cancer!

Conditional Probability and Independence
When knowledge that one event has happened does not change the likelihood that another event will happen, we say the two events are **independent.**

Definition:
Two events A and B are **independent** if the occurrence of one event has no effect on the chance that the other event will happen. In other words, events A and B are independent if
$P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

Example 7: Is there a relationship between gender and handedness? To find out, we used CensusAtSchool's Random Data Selector to choose an SRS of 100 Australian high school students who completed a survey. The two-way table displays data on the gender and dominant hand of each student. Are the events "male" and "left-handed" independent? Justify your answer.

	Gender		
Dominant Hand	Male	Female	Total
Right	39	51	**90**
Left	7	3	**10**
Total	**46**	**54**	**100**

***P*(left-handed | male) = 7/46 = 0.1522**

***P*(left-handed) = 10/100 = 0.10**

These probabilities are not equal, therefore the events "male" and "left-handed" are not independent.

Example 8: For each chance process below, determine whether the events are independent. Justify your answer.

a) Shuffle a standard deck of cards, and turn over the top card. Put it back in the deck, shuffle again, and turn over the top card. Define events A: first card is a heart, and B: second card is a heart.

Independent. Because we are replacing the cards, knowing what the first card was will not help us predict what the second card will be.

b) Shuffle a standard deck of cards, and turn over the top two cards, one at a time. Define events A: first card is a heart, and B: second card is a heart.

Not independent. Once we know the suit of the first card, then the probability of getting a heart on the second card will change depending on what the first card was.

c) The 28 students in Mr. Tabor's AP® Statistics class completed a brief survey. One of the questions asked whether each student was right or left-handed. The two-way table summarizes the class data. Choose a student from the class at random. The events of interest are "female" and "right-handed."

	Gender	
Handedness	**Female**	**Male**
Left	3	1
Right	18	6

Independent. *P*(right-handed) = 24/28 = 6/7 is the same as *P*(right-handed | female) = 18/21 = 6/7.

Independence: A Special Multiplication Rule
When events A and B are independent, we can simplify the general multiplication rule since $P(B \mid A) = P(B)$.

Definition:
Multiplication rule for independent events
If A and B are independent events, then the probability that A and B both occur is
$P(A \cap B) = P(A) \cdot P(B)$

Note that this rule only applies to independent events. Let's look at an example that uses the multiplication rule for independent events to analyze an important historical event.

Example 9: In January 28, 1986, Space Shuttle *Challenger* exploded on takeoff. All seven crew members were killed. Following the disaster, scientists and statisticians helped analyze what went wrong. They determined that the failure of O-ring joints in the shuttle's booster rockets was to blame. Under the cold conditions that day, experts estimated that the probability that an individual O-ring joint would function properly was 0.977. But there were six of these O-ring joints, and all six had to function properly for the shuttle to launch safely. Assuming that O-ring joints succeed or fail independently, find the probability that the shuttle would launch safely under similar conditions.

P(joint1 OK and joint 2 OK and joint 3 OK and joint 4 OK and joint 5 OK and joint 6 OK)

=*P*(joint 1 OK) • P(joint 2 OK) • … • P(joint 6 OK)

=(0.977)(0.977)(0.977)(0.977)(0.977)(0.977)

= 0.87

Example 10: Many people who come to clinics to be tested for HIV, the virus that causes AIDS, don't come back to learn the test results. Clinics now use "rapid HIV tests" that give a result while the client waits. In a clinic in Malawi, for example, use of rapid tests increased the percent of clients who learned their test results from 69% to 99.7%.

The trade-off for fast results is that rapid tests are less accurate than slower laboratory tests. Applied to people who have no HIV antibodies, one rapid test has probability about 0.004 of producing a false positive (that is, of falsely indicating that antibodies are present). If a clinic tests 200 randomly selected people who are free of HIV antibodies, what is the chance that at least one false positive will occur?

It is reasonable to assume that the test results for different individuals are independent. We have 200 independent events, each with probability 0.004. "at least one" combines many possible outcomes. It will be easier to use the fact that

P(at least one positive) = 1 − *P*(no positives)

We'll find *P*(no positives) first. The probability of a negative result for any one person is 1 − 0.004 = 0.996. To find the probability that all 200 people tested have negative results, use the multiplication rule for independent events:

$$\begin{aligned} P(\text{no positives}) &= P(\text{all 200 negative}) \\ &= (0.996)(0.996)\ldots(0.996) \\ &= 0.996^{200} = 0.4486 \end{aligned}$$

The probability we want is therefore

P(at least one positive) = 1 − 0.4486 = 0.5514

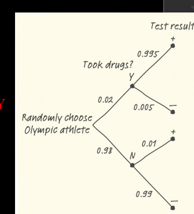
There is more than a 50% chance that at least 1 of the 200 people will test positive for HIV, even though no one has the virus.

Example 11: Lately, it seems that we've been hearing more in the news about athletes who have tested positive for performance-enhancing drugs. From Major League Baseball, to the Tour de France, to the Olympics, we have learned about athletes who took banned substances. Sports at all levels now have programs in place to test athletes' urine or blood samples. There's an important question about drug testing that probability can help answer: if an athlete tests positive for a banned substance, did the athlete necessarily attempt to cheat?

Over 10,000 athletes competed in the 2008 Olympic Games in Beijing. The International Olympic Committee wanted to ensure that the competition was as fair as possible. So the committee administered more than 5000 drug tests to athletes. All medal winners were tested, as well as other randomly selected competitors.

Suppose that 2% of athletes had actually taken (banned) drugs. No drug test is perfect. Sometimes the test says that an athlete took drugs, but the athlete actually didn't. We call this a false positive result. Other times, the drug test says an athlete is "clean," but the athlete actually took drugs. This is called a false negative result. Suppose that the testing procedure used at the Olympics has a false positive rate of 1% and a false negative rate of 0.5%.

The tree diagram to the right shows whether Olympic athletes took drugs and the likelihood of getting a positive or a negative test result from a drug test. What's the probability that an athlete who tests positive actually took drugs?



We want to find *P*(took drugs | test +). By the conditional probability formula,

$$P(\text{took drugs} \mid \text{test} +) = \frac{P(\text{took drugs} \cap \text{test} +)}{P(\text{test} +)}$$

To find *P*(took drugs ∩ test +), we multiply along the Y and + branches of the tree.

P(took drugs ∩ test +) = (0.02)(0.995) = 0.0199

To find *P*(test +), we need to calculate the probability that a randomly selected athlete gets a positive test result. there are two ways this can happen: (1) if the athlete took drugs and the test result is positive, and (2) if the athlete didn't take drugs, but the test gives a false positive. From the tree diagram, the desired probability is

P(test +) = (0.02)(0.995) + (0.98)(0.01) = 0.0297

Using these two results, we can find the conditional probability:

$$P(\text{took drugs} \mid \text{test} +) = \frac{P(\text{took drugs} \cap \text{test} +)}{P(\text{test} +)} = \frac{0.0199}{0.0297} = 0.67$$

That is, 67% of athletes who test positive actually took drugs.

Is there a connection between mutually exclusive and independent? Let's start with a new chance process. Choose a U.S. adult at random. Define event *A*: the person is male, and event *B*: the person is pregnant. It's pretty clear that these two events are mutually exclusive (can't happen together)! What about independence? If you know that event *A* has occurred, does this affect the probability that event *B* happens? Of course! If we know the person is male, then the chance that the person is pregnant is 0. Since $P(B \mid A) \neq P(B)$, the two events are not independent. **Two mutually exclusive events can never be independent, because if one event happens, the other event is guaranteed not to happen.**