Section 7.1 Basic Concepts of Probability

Question 1 – What is an event?

Question 2 – What is probability?

Question 3 - How is probability assigned?

Question 1 – What is an event?

Key Terms

Probability	Outcomes	Event	Experiment
Sample space			

Summary

Probability is used to measure the likelihood of something happening. Implicit in the idea of likelihood is chance. We are uncertain what will happen. In this section, we'll learn how to assign a number from 0 to 1 that reflects how likely an event will occur. A probability of 0 means the event will not occur. A probability of 1 means the event will occur. A probability between 0 and 1 reflects varying degrees to which the event might occur. If the chance of rain is 0.1 (often read as a 10% chance of rain), it probably won't rain. However, a probability of 0.9 (a 90% chance of rain) indicates that it probably will rain. A 50% chance of rain (probability equal to 0.5) means is just as likely to rain as not rain.

An **experiment** is a process that generates uncertain occurrences. These occurrences are called the **outcomes** of the experiment.

The collection of all possible outcomes of an experiment is called the **sample space**. An **event** is any collection of outcomes in the sample space.

For instance, suppose a manufacturer is producing batteries that are sold in a two pack. If a package of batteries is selected from the production line, the batteries in the package may be examined to determine whether they work or are defective. The process of examining whether the batteries in the package are defective is an experiment. The outcome of the experiment may be listed by indicating whether each battery is working (W) or defective (D).

First Battery	Second Battery
W	W
W	D
D	W
D	D

We can specify the first outcome of the experiment as (*W*, *W*). Other outcomes can be written in a similar manner. Written this way, this first letter indicates whether the first battery in the package is working or defective. The second letter indicates whether the second battery in the package is working or defective. We can refer to these outcomes collectively as

$$S = \{(W,W),(W,D),(D,W),(D,D)\}$$

The experiment is carried out many times with each outcome being uncertain. These repetitions of the experiment are called trials.

<u>Notes</u>

<table>
<tr><th><u>Guided Example 1</u></th><th><u>Practice</u></th></tr>
<tr><td>
A marketing specialist administers a three-question test. Each question is answered yes or no.

a. Find the sample space if we are interested in knowing how many questions were answered yes.

Solution The sample space consists of all of the possible ways to answer yes. Since there are three questions, it is possible to answer no questions, 1 question, 2 questions, or three questions yes. This makes the sample space,

$S = \{0, 1, 2, 3\}$

b. Find the sample space if we are interested in knowing how the tests were answered.

Solution Think of each survey as consisting a sequence of three Y’s or N’s. For instance, NYY would correspond to the outcome where the answer to the first question is no and the answer to the last two questions is yes. The sample space is

$S = \{YYY, YNN, NYN, NNY, NYY, YNY, YYN, NNN\}$

c. For the sample space in part b, list the outcomes in the event “more questions answered yes than no”.

Solution The event “more questions answered yes than no” consists of events in the sample space with more Y’s than N’s. Examining the sample space S in part b, the event is

$\{YYY, YYN, NYY, YNY\}$
</td><td>
A marketing specialist administers a four-question test. Each question is answered yes or no.

a. Find the sample space if we are interested in knowing how many questions were answered yes.

b. Find the sample space if we are interested in knowing how the tests were answered.

c. For the sample space in part b, list the outcomes in the event “more questions answered yes than no”.
</td></tr>
</table>

Question 2 – What is probability?

Key Terms

Probability

Summary

Probability is defined in terms of the outcomes in the sample space of an experiment. Suppose we have an experiment with a finite number of outcomes in the sample space. Let's represent the outcomes with the letter e followed by a subscript. If there are n outcomes from the experiment, then the sample space S is

$$S = \{e_1, e_2, \ldots, e_n\}$$

The probability of each outcome is symbolized by writing $P(e_1)$, $P(e_2)$, …, $P(e_n)$. We can assign a probability to each outcome as long as the probability satisfies certain requirements.

Each outcome of an experiment must meet two requirements.

1. The probability of each outcome is a number from 0 to 1,

$$0 \le P(e_1) \le 1, \quad 0 \le P(e_2) \le 1, \ \ldots, \quad 0 \le P(e_n) \le 1$$

2. The sum of the probabilities of all outcomes is equal to 1,

$$P(e_1) + P(e_2) + \cdots + P(e_n) = 1$$

Probability is a number that indicates the likelihood of an occurrence happening. This number may be as low as 0 indicating the occurrence will not happen. If the probability is equal to 1, the occurrence is certain to happen. Probabilities between 0 and 1 indicate the varying levels of uncertainty about the occurrence.

Notes

Guided Example 2	Practice
A company monitors the snack boxes of fig newtons coming off a production line. They measure the number of fig newtons in each package and assign the probabilities below. Determine if the probability assignment is valid.	A company monitors the graphics chips coming off a production line. They measure the number of defective chips and assign the probabilities below. Determine if the probability assignment is valid.
a. $P(0) = 10, P(1) = 10, P(2) = 80, P(3) = 20$	a. $P(0) = 1, P(2) = 99, P(3) = 5$
Solution To be a valid assignment, each outcome must be assigned a probability from 0 to 1. The outcomes in this assignment are greater than 1 so this is not a valid probability assignment.	
b. $P(0) = 0.01, P(1) = 0.01, P(2) = 0.95, P(3) = 0.03$	b. $P(0) = 0.98, P(2) = 0.01, P(3) = 0.01$
Solution Each probability is in the interval [0, 1] so the first conditional on probability assignments is met. The second condition indicates that the sum of the probabilities must be 1. Adding the four probabilities gives $$0.01 + 0.01 + 0.95 + 0.03 = 1$$ Since both conditions are satisfied, this is a valid probability assignment.	
c. $P(0) = 0.02, P(1) = 0.01, P(2) = 0.92, P(3) = 0.01$	c. $P(0) = 0.95, P(1) = 0.02, P(2) = 0.01$
Solution All probabilities are in the interval [0, 1]. Checking the sum of the probabilities, we see that $$0.02 + 0.01 + 0.92 + 0.01 = 0.98$$ The sum of the probabilities is not 1 so this is not a valid probability assignment.	

Question 3 – How is probability assigned?

Key Terms

Equally likely Empirical probability

Summary

There are several ways to assign probability to outcomes in an experiment. The simplest method is to assume that each outcome in the sample space is **equally likely**. In this case, the probability of each outcome in the sample space is the same as any other outcome in the sample space.

Probability of Equally Likely Outcomes

Suppose the outcomes from an experiment are equally likely. If the sample space for the experiment contains *n* outcomes,

$$S = \{e_1, e_2, \ldots, e_n\}$$

then the probabilities of the outcomes are

$$P(e_1) = P(e_2) = \cdots = P(e_n) = \frac{1}{n}$$

The assumption that the outcomes are equally likely is a powerful assumption. It allows us to roll a fair die with six sides and compute the probability of getting a six as $\frac{1}{6}$. We can also use this assumption to compute the probability of selecting the king of clubs from a 52-card deck as $\frac{1}{52}$. However, this assumption may lead to probabilities that are not realistic.

Suppose a factory worker tests randomly selected items from a production line to determine whether they are defective or not defective. If these two outcomes are assumed to be equally likely,

$$P(\text{defective}) = \frac{1}{2}, \quad P(\text{nondefective}) = \frac{1}{2}$$

This factory has a serious problem with quality control! The worker knows from experience that he is much more likely to find that the item is not defective. The equally likely assumption must not be valid.

To get an idea of how likely it is to test an item and find whether it is defective or not defective, the factory worker repeats the testing experiment many times. Out of 500 items, he finds 10 defective products and 490 not defective products. Based on these results, he calculates the probabilities

$$P(\text{defective}) = \frac{10}{500} = 0.02, \quad P(\text{nondefective}) = \frac{490}{500} = 0.98$$

These numbers are the relative frequencies of each outcome in the sample space. We can estimate probabilities of outcomes by repeating an experiment many times and calculating the relative frequency of each outcome. Probability estimated from relative frequencies in a sample of trials from an experiment is called **empirical probability**.

Empirical Probability

If an experiment is performed many times, the probability of an outcome to the experiment is

$$P(e_i) \approx \frac{\text{Number of times } e_i \text{ occurs}}{\text{Total number of trials}}$$

where e_i is any outcome in the sample space of the experiment.

<u>Notes</u>

Guided Example 3

In an analysis of airplane crashes, a researcher notes the primary causes of a crash.

Cause of Crash
Mechanical failure
Weather
Terrorism
Collision with another object
Pilot error

If each of the outcomes are assumed to be equally likely, what is the probability that a crash is caused by terrorism.

Solution Since the each of the outcomes are equally likely, the probability of all outcomes are the same and each is equal to $\frac{1}{n}$ where n is the number of outcomes in the sample space. Since there are 5 outcomes in the sample space, each has a probability of $\frac{1}{5}$. In particular,

$$P(\text{terrorism}) = \frac{1}{5}$$

Practice

A meteorologist notes that on a certain date with certain weather data, the following weather conditions are observed at noon.

Conditions
Sunny
Partly cloudy
Mostly cloudy
Rain

If each of the outcomes are assumed to be equally likely, what is the probability of rain on this date with the weather data observed.

.

Guided Example 4

In an analysis of airplane crashes, a researcher notes the primary cause of a crash and the corresponding frequency of those causes.

Cause of Crash	Frequency
Mechanical failure	123
Weather	52
Terrorism	1
Collision with another object	3
Pilot error	21

Use the frequencies to find the probability that the primary cause of an airplane crash is weather.

Solution To compute the relative frequency of a crash caused by weather we need to know the

Practice

A meteorologist notes that on a certain date with certain weather data, the following weather conditions are observed at noon.

Conditions	Frequency
Sunny	76
Partly cloudy	23
Mostly cloudy	5
Rain	4

Use the frequencies to find the probability of a sunny day on this date with the weather data observed.

total number of trials from the table. The sum of the frequencies is 200. The probability that a crash is caused by weather is $$P(\text{weather}) = \frac{\text{Number of weather crashes}}{\text{Total number of trials}}$$ $$= \frac{52}{200}$$ $$= 0.26$$ Based on the data in the table, the likelihood of a crash caused by weather is 26%.	

Section 7.2 Probability Rules

Question 1 – How do you find the probability of an event?

Question 2 – What is the complement of an event?

Question 3 - How do you find the probability of a compound event?

Question 4 – What is the difference between marginal and joint probability?

Question 1 – How do you find the probability of an event?

<u>Key Terms</u>

Event

<u>Summary</u>

An **event** is any collection of outcomes in the sample space. The probability of an event is the sum of the probabilities of the outcomes corresponding to the event.

Probability of an Event

If E is composed of a collection of the outcomes in the sample space S,

$$E = \{e_1, e_2, ..., e_n\}$$

then the probability of the event E is

$$P(E) = P(e_1) + P(e_2) + \cdots + P(e_n)$$

We can use the probabilities of the outcomes in the sample space to find the probability of any event.

The probability of events composed of equally likely outcomes may be calculated by counting the number of outcomes in the event. For instance, if an event contains M outcomes (each with probability $\frac{1}{N}$), the probability of the event must be

$$P(E) = \underbrace{\frac{1}{N} + \frac{1}{N} + \cdots + \frac{1}{N}}_{M \text{ terms}} = \frac{M}{N}$$

Probability of an Event With Equally Likely Outcomes

Suppose the sample space of an experiment contains N equally outcomes. If an event E contains M of those outcomes, the probability of the event is

$$P(E) = \frac{M}{N}$$

This expression must be used cautiously since it requires that each outcome in the sample space be equally likely. This expression is often written using the letter *n* to indicate the number of outcomes in a collection. In this case,

$$P(E) = \frac{n(E)}{n(S)}$$

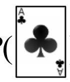
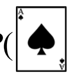
where the notations $n(E)$ and $n(S)$ are the number of outcomes in the event *E* and sample space *S*.

<u>Notes</u>

Guided Example 1

You are dealt one card from a standard 52-card deck. Find the probability of being dealt an ace.

Solution The likelihood of dealing any individual card in the deck is $\frac{1}{52}$. Since there are four aces in the deck, the probability of dealing an ace is

$P(\text{ace}) = P($$) + P($$) + P($$) + P($$)$

Each of these cards has a probability of $\frac{1}{52}$ and we can insert these probabilities to give

$$P(\text{ace}) = \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{4}{52}$$

We could also calculate this probability by counting the aces,

$$P(\text{ace}) = \frac{\text{number of aces in deck}}{\text{number of cards in deck}} = \frac{4}{52}$$

This probability simplifies to $\frac{1}{13}$ or approximately 7.7%.

Practice

You are dealt one card from a standard 52-card deck. Find the probability of being dealt a heart.

Guided Example 2

A marble is selected at random from a jar containing 3 red marbles, 2 yellow marbles, and 5 green marbles.

a. What is the probability that the marble is red?

Solution Each of ten marbles is equally likely. The probability is calculated by counting the outcomes,

$$P(\text{red}) = \frac{n(\text{red})}{n(S)} = \frac{3}{10}$$

The probability of selecting a red marble is 0.3 or 30%.

Practice

A marble is selected at random from a jar containing 5 red marbles, 13 purple marbles, and 2 green marbles.

a. What is the probability that the marble is purple?

b. What is the probability that the marble is green? **Solution** There are five green marbles, $$P(\text{green}) = \frac{n(\text{green})}{n(S)} = \frac{5}{10}$$ The probability of selecting a green marble is 0.5 or 50%.	b. What is the probability that the marble is green?

Guided Example 3

The table below lists the estimated number of preventable injuries in the United States in 2016 and their causes.

Cause	Injuries
Fall	8,591,683
Struck By/Against	3,685,012
Overexertion	2,569,850
Motor vehicle-Occupant	2,500,353
Other Specified	2,365,891
Cut/Pierce	1,823,358
Poisoning	1,755,044
Other Bite/Sting	1,142,130
Unknown/Unspecified	755,567
Foreign Body	557,650

Source: Center for Disease Control and Prevention, National Center for Injury Prevention and Control

a. What is the probability that a randomly selected person with a preventable injury was hurt by overexertion?

Solution To find the probability of preventable injury by overexertion, we must first find the total number of preventable injuries in the table. Adding the number of preventable injuries gives a total of 25,746,538 injuries. The probability that a preventable injury was caused by overexertion is the number of preventable overexertion injuries

Practice

The table below lists the estimated number of preventable deaths in the United States in 2016 and their causes.

Cause	Deaths
Poisoning	64,795
Motor vehicle	38,659
Fall	36,338
Suffocation	6,946
Drowning	3,709
Fire/burn	2,902
Natural/ Environment	1,750
Other Specified, classifiable	1,440
Other Land Transport	1,332
Other Specified	1,251

Source: Center for Disease Control and Prevention, National Center for Injury Prevention and Control

a. What is the probability that a randomly selected person with a preventable injury died by motor vehicle?

<table>
<tr>
<td>
divided by the total number of preventable injuries,

$$P(\text{overexertion}) = \frac{2569850}{25746538} \approx 0.100$$

Written as a percent, this means the likelihood that a preventable injury was caused by overexertion is 10.0%.

b. What is the probability that a randomly selected person with a preventable injury was hurt by “other”?

Solution Two of the injuries listed in the table include the term “other” (Other Specified and Other Bite/Sting). The number of “other” causes is 2,365,891 + 1,142,130 or 3,508,021. The probability is calculated by dividing this number by the total number of preventable injuries,

$$P(\text{other}) = \frac{3508021}{25746538} \approx 0.136$$

This means that the likelihood of a preventable injury being caused by “other” is 13.6%.
</td>
<td>
b. What is the probability that a randomly selected person with a preventable injury died “other”?
</td>
</tr>
</table>

Question 2 – What is the complement of an event?

Key Terms

Complement

Summary

The complement of an event E is all of the outcomes in the sample space that are not in the event E. The complement of an event E is represented by the symbols E'. In discussing the complement, we often refer to it as the outcomes not in E. Since the event E and the event not in E combine to give the entire sample space S,

$$P(E)+P(E')=P(S)$$

The likelihood of an outcomes in the sample space occurring is certain, so we simplify this to

$$P(E)+P(E')=1$$

This leads us to a convenient relationship for determining the likelihood that an outcome in the complement will occur.

Probability of the Complement of an Event

The probability that an outcome in the complement E' will occur is

$$P(E')=1-P(E)$$

In other words, the probability that an event will not occur is 1 minus the probability that it will occur.

Notes

Guided Example 4	Practice
You are dealt one card from a standard 52-card deck. Find the probability of being dealt a card that is not an ace. **Solution** From Guided Example 1 in this section, we know that $$P(\text{ace}) = \frac{\text{number of aces in deck}}{\text{number of cards in deck}} = \frac{4}{52}$$ To find the probability of not being dealt an ace, we realize that being dealt an ace and not being dealt an ace are compliments of each other. We can subtract the probability of being dealt an ace from 1 to get the probability of not being dealt an ace, $$\begin{aligned} P(\text{not ace}) &= 1 - P(\text{ace}) \\ &= 1 - \frac{4}{52} \\ &= \frac{48}{52} \end{aligned}$$ or approximately 92.3%.	You are dealt one card from a standard 52-card deck. Find the probability of being dealt a card that is not a heart.

<table>
<tr><th><u>Guided Example 5</u></th><th><u>Practice</u></th></tr>
<tr><td>
A marble is selected at random from a jar containing 3 red marbles, 2 yellow marbles, and 5 green marbles.

a. What is the probability that the marble is <u>not</u> red?

Solution From Guided Example 2 in this section,

$$P(\text{red}) = \frac{n(\text{red})}{n(S)} = \frac{3}{10}$$
Since the event “marble selected is red” and “marble selected is not red” are compliments of each other,
$$\begin{aligned} P(\text{not red}) &= 1 - P(\text{red}) \\ &= 1 - \frac{3}{10} \\ &= \frac{7}{10} \end{aligned}$$
b. What is the probability that the marble is <u>not</u> green?

Solution In Guided Example 2, we calculated

$$P(\text{green}) = \frac{n(\text{green})}{n(S)} = \frac{5}{10}$$
Since the event “marble selected is green” and “marble selected is not green” are compliments,

$$\begin{aligned} P(\text{not green}) &= 1 - P(\text{green}) \\ &= 1 - \frac{5}{10} \\ &= \frac{5}{10} \end{aligned}$$
</td><td>
A marble is selected at random from a jar containing 5 red marbles, 13 purple marbles, and 2 green marbles.

a. What is the probability that the marble is <u>not</u> purple?

b. What is the probability that the marble is <u>not</u> green?
</td></tr>
</table>

Guided Example 6

The table below lists the estimated number of preventable injuries in the United States in 2016 and their causes.

Cause	Injuries
Fall	8,591,683
Struck By/Against	3,685,012
Overexertion	2,569,850
Motor vehicle-Occupant	2,500,353
Other Specified	2,365,891
Cut/Pierce	1,823,358
Poisoning	1,755,044
Other Bite/Sting	1,142,130
Unknown/Unspecified	755,567
Foreign Body	557,650

Source: Center for Disease Control and Prevention, National Center for Injury Prevention and Control

What is the probability that a randomly selected person with a preventable injury was hurt by a cause other than overexertion?

Solution The event “selected person with a preventable injury was hurt by a cause other than overexertion” corresponds to all of the rows in the table except the overexertion table. This helps us to deduce that the event “selected person with a preventable injury was hurt by a cause other than overexertion” and the event “selected person with a preventable injury was hurt by overexertion” are compliments of each other. We can use the results from Guided Example 3 to calculate

$$\begin{aligned} P(\text{not overexertion}) &= 1 - P(\text{overexertion}) \\ &= 1 - \frac{2569850}{25746538} \\ &= \frac{23176688}{25746538} \\ &\approx 0.900 \end{aligned}$$

or approximately 90.0%.

Practice

The table below lists the estimated number of preventable deaths in the United States in 2016 and their causes.

Cause	Deaths
Poisoning	64,795
Motor vehicle	38,659
Fall	36,338
Suffocation	6,946
Drowning	3,709
Fire/burn	2,902
Natural/ Environment	1,750
Other Specified, classifiable	1,440
Other Land Transport	1,332
Other Specified	1,251

Source: Center for Disease Control and Prevention, National Center for Injury Prevention and Control

What is the probability that a randomly selected person with a preventable death by a cause other than motor vehicle?

Question 3 – How do you find the probability of a compound event?

Key Terms

Compound event Intersection Union

Union rule for probability

Summary

Events may be combined together in various ways. These combinations are called **compound events**. If we know the probability of the events that make up the compound event, we are often able to compute the probability of the compound event.

Outcomes that are in the event A as well as the event B are said to be in the compound event, A and B. The word "and" is used to indicate that the outcomes in this event are in both events simultaneously. Mathematicians describe outcomes in A and B with the intersection symbol $\cap$. An outcome in A and B are the same outcomes in the intersection of A with B, $A \cap B$. The probability of A and B occurring is often referred to as the joint probability of A and B.

Another type of compound event is denoted using the word "or". An outcome is in the event A or B if it is in A, B, or both events simultaneously. In the language of sets, the compound event A or B is the same as the union of the set A with the set B. The symbol $\cup$ represents the union of two sets. Using this symbol, we write the union of A with B as $A \cup B$. In this text we will use the word "or" instead of the union symbol to represent outcomes in A, in B, or in both events.

Compound probabilities are related to each other the following relationship.

The Probability of A or B

The likelihood of an event A occurring or an event B occurring is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

This rule is often referred to as the **union rule for probability**.

Notes

Guided Example 7

A product survey of form returned by 1000 consumers after the purchase of a new vacuum cleaner gives the following results.

Event	**Number of consumers**
Consumer feels the vacuum is a good value	950
Consumer feels vacuum's manual is easy to follow	430
Consumer feels the vacuum is a good value and vacuum's manual is easy to follow	400

Define the events

G: Consumer feels the vacuum is a good value
M: Consumer feels the vacuum's manual is easy to follow.

a. Write the compound event "Consumer feels the vacuum is a good value and vacuum's manual is easy to follow" in terms of the events defined above.

Solution The compound event uses the word "and" to connect the events. This corresponds to the joint event G and M.

b. Describe the event G or M in words.

Solution The event G or M corresponds to outcomes in which the consumer feels the vacuum is a good value or feels the manual is easy to follow.

Practice

A mechanic tracks 100 repairs with the following results.

Event	**Number of repairs**
Repair requires a new battery	50
Repair requires a new alternator	10
Repair requires a new battery and a new alternator	5

Define the events

B: Repair requires a new battery
A: Repair requires a new alternator

a. Write the compound event "needs a new battery and a new alternator" in terms of the events defined above.

b. Describe the event B or A in words.

c. What is the probability of G or M occurring? **Solution** To find $P(G \text{ or } M)$, apply $P(G \text{ or } M) = P(G) + P(M) - P(G \text{ and } M)$ Each of the probabilities on the right may be calculated from relative frequencies: $P(G) = \frac{950}{1000} = 0.95$ $P(M) = \frac{430}{1000} = 0.43$ $P(G \text{ and } M) = \frac{400}{1000} = 0.4$ Using these probabilities gives $P(G \text{ or } M) = 0.95 + 0.43 - 0.40 = 0.98$ or 98%	c. What is the probability of B or A occurring?

Guided Example 8

What is the probability of getting either a sum of 6 or doubles in the roll of a pair of dice?

Solution Let's define the events as follows:

A: sum of six
B: doubles

The event "getting either a sum of 6 or doubles" corresponds to A or B. The union rule for probability states that

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

A table of all possible rolls of two dice helps us to calculate each of the probabilities on the right side.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

In this table, the faces of each die are indicated in the shaded portion and the corresponding sums in the unshaded portion.

To find $P(A)$, we locate the combinations of the dice that sum to 6.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

There are 36 possible rolls of the dice and five of them sum to 6 so

Practice

What is the probability of getting either a sum of 8 or doubles in the roll of a pair of dice?

$$P(A) = \frac{n(A)}{n(S)} = \frac{5}{36}$$

The likelihood of doubles is found by counting the possibilities in the table.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

There are 6 ways to get doubles,

$$P(B) = \frac{n(B)}{n(S)} = \frac{6}{36}$$

The event A and B corresponds to all rolls of the dice with a sum of 6 and that are doubles. This event corresponds to one out come where each die is a three:

$$P(A \text{ and } B) = \frac{n(A \text{ and } B)}{n(S)} = \frac{1}{36}$$

Put these probabilities into the union rule to give

$$\begin{aligned} P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) \\ &= \frac{5}{36} + \frac{6}{36} - \frac{1}{36} \\ &= \frac{10}{36} \end{aligned}$$

or approximately 27.8%.

Question 4 – What is the difference between marginal and joint probability?

<u>Key Terms</u>

Marginal probability

<u>Summary</u>

In Question 3 we introduced the idea of joint probability. Joint probabilities are the likelihoods associated with compound events using “and”. The joint probability of A and B is the likelihood that both events will occur simultaneously. **Marginal probabilities** are the probabilities of the individual events that make up the joint probability. Marginal probabilities are often found by using relative frequencies or the relationship

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

where A and B are events.

<u>Notes</u>

Guided Example 9

Suppose a cellular provider collects the data below about a sample of users.

Amount of Data Used	Male Users	Female Users
Less than 1GB	425	300
1 GB up to, but not including 2GB	1505	1225
2 GB up to, but not including 5 GB	450	550
5 GB up to, but not including 10GB	330	625
10 GB or more	125	205

a. Find the likelihood that a female user will use more than 10 GB of data.

Solution To make the problem easier to complete, let's sum the row and columns of the table.

Amount of Data Used	Male Users	Female Users	Total
Less than 1GB	425	300	725
1 GB up to, but not including 2GB	1505	1225	2730
2 GB up to, but not including 5 GB	450	550	1000
5 GB up to, but not including 10GB	330	625	955
10 GB or more	125	205	330
Total	2835	2905	5740

Define the events

F: user is female

A: users uses more than 10 GB of data

Practice

Suppose a cellular provider collects the data below about a sample of users.

Amount of Data Used	Male Users	Female Users
Less than 1GB	425	300
1 GB up to, but not including 2GB	1505	1225
2 GB up to, but not including 5 GB	450	550
5 GB up to, but not including 10GB	330	625
10 GB or more	125	205

a. Find the likelihood that a male user will use less than 1 GB of data.

In terms of these events, we must find the joint probability that the user is female and uses more than 10 GB, $P(F \text{ and } A)$. From the table, we recognize that there are 205 female users who use more than 10 GB of data. Since the total number of users is 5740, the relative frequency may be calculated,

$$P(F \text{ and } A) = \frac{205}{5740} \approx 0.036$$

The likelihood that a user in the survey is female and used more than 10 GB of data is approximately 3.6%.

b. Find the probability that a user is female.

Solution To use calculate the relative frequency of the event, we must divide the number of female users by the total number of users. The total number female users is at the bottom of the third column. Dividing this by the total number of users in the bottom of the last column gives,

$$P(F) = \frac{2905}{5740} \approx 0.506$$

The marginal probability that a user in the survey is female is approximately 50.6%.

c. Find the probability that a user in the survey will use more than 10 GB of data.

Solution According to the survey, 330 users of the total 5740 users used more than 10 GB of data. This means the probability of using more than 10 GB of data is

$$P(A) = \frac{330}{5740} \approx 0.057$$

The likelihood of using more than 10 GB of data is approximately 5.7%.

b. Find the probability that a user is male.

c. Find the probability that a user in the survey will use less than 1 GB of data.

d. Find the probability that the user is female or more than 10 GB of data is used. **Solution** The event "the user is female or more than 10 GB of data is used" corresponds to the compound event F or A. We calculate the probability of this event by using the probabilities of the events in parts a through c, $P(F \text{ or } A) = P(F) + P(A) - P(F \text{ and } A)$ $= \frac{2905}{5740} + \frac{330}{5740} - \frac{205}{5740}$ $= \frac{3030}{5740}$ ≈ 0.528 The probability that a user is female or uses more than 10 GB of data is approximately 52.8%.	d. Find the probability that the user is male or less than 1 GB of data is used.

Section 7.3 Conditional Probability

Question 1 – What is conditional probability?

Question 2 – How is conditional probability computed?

Question 3 - What are independent events?

Question 4 – What is the product rule for probability?

Question 5 - How is Bayes' Rule used to compute conditional probability?

Question 1 – What is conditional probability?

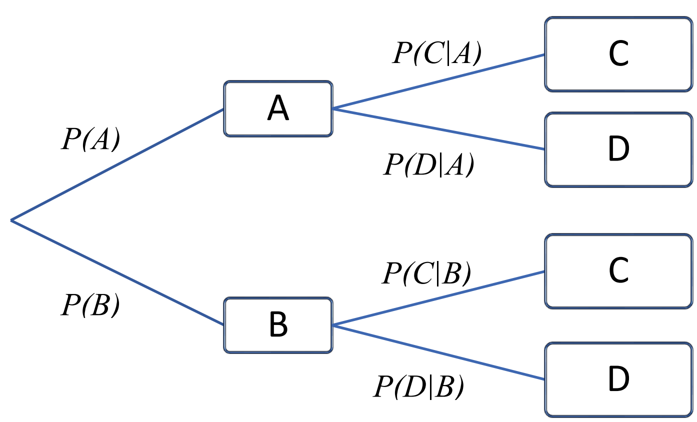
Key Terms

Conditional probability

Summary

Conditional probability is the likelihood of an event occurring given that another event has occurred. A vertical line is used to indicate the event whose probability is being computed and the event that has already occurred. For instance, the symbols $P(A \mid B)$ correspond to the probability of A occurring given that B has already occurred. The vertical bar separates the probability we are interested in calculating from the event that is assumed to have occurred.

A tree diagram is often used to represent conditional probabilities.



In this context, the events are depicted in the boxes and the corresponding probabilities are labeled on the branches connecting the boxes. If you follow the set of branches to A and then C, note that the first branch is labeled with $P(A)$ indicating the probability of A. Continuing to C, we see that the branch is labeled $P(C \mid A)$ indicating the probability of C given that A has occurred.

<u>Guided Example 1</u>

A survey is administered to a group of consumers who own a mobile phone. The results of the survey are shown below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

Explain what each of the probabilities below mean.

a. $P(SP)$

Solution Since *SP* is the event "consumer owns a smartphone", $P(SP)$ is the probability that a consumer owns a smartphone. To find this probability, we simply count the number of consumers in this event and divide by the total number of consumers who took the survey:

$$P(SP) = \frac{n(SP)}{n(S)} = \frac{2802}{3300}$$

b. $P(F)$

Solution The event F corresponds to the event "consumer is female". The probability that a consumer in the survey is female is $P(F)$. This probability is found by dividing the number of males in the survey by the total number of consumers in the survey,

$$P(F) = \frac{n(F)}{n(S)} = \frac{1852}{3300}$$

<u>Practice</u>

A survey is administered to a group of consumers who own a mobile phone. The results of the survey are shown below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

Explain what each of the probabilities below mean.

a. $P(M)$

b. $P(BP)$

c. $P(SP \text{ and } F)$ **Solution** The event SP and F is all of the outcomes in common between "consumer owns a smartphone" and consumer is female". So, $P(SP \text{ and } F)$ is the probability of a female consumer who owns a smartphone. To find this probability, divide the number of female consumers who own a smartphone by the total number of consumers in the survey, $$P(SP \text{ and } F) = \frac{n(SP \text{ and } F)}{n(S)} = \frac{1601}{3300}$$	c. $P(M \text{ and } BP)$
d. $P(SP \mid F)$ **Solution** The vertical bar tells us that we are describing an event with conditional probability. In this case, we are given the event that the "consumer is female" and we want to know the likelihood that the "consumer owns a smartphone". To find the probability of this event, we need to recognize that we are not interested in all consumers in the survey, only the female consumers. A total of 1852 female consumers took the survey. Of those female consumers, 1601 owned a smartphone. The probability of a consumer owning a smartphone given they are female is $$P(SP \mid F) = \frac{1601}{1852}$$	d. $P(M \mid BP)$

Guided Example 2

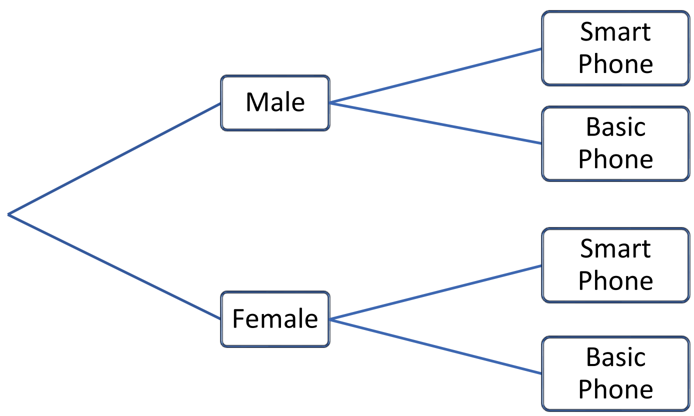
A survey was administered to a group of consumers who own a mobile phone. The results of the survey are below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

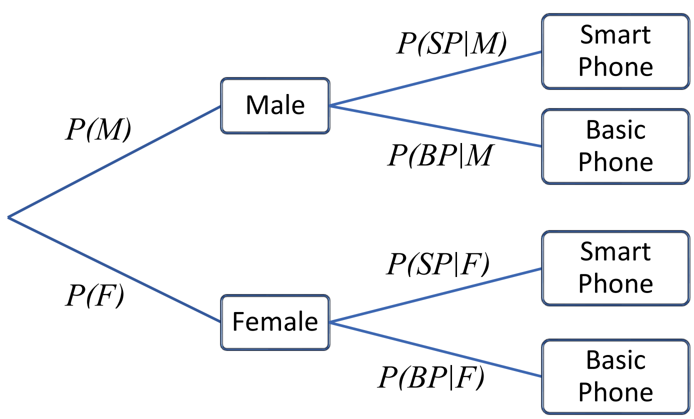
Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

Find the probabilities on each branch of the tree diagram below.



Solution In the tree diagram, we start on the left and work to the right. The branches are labeled with the probabilities below:



Examine the diagram carefully to note that the second level of the tree consists of conditional probabilities. In each case, the given part is where the branch originates and the probability we want is where the branch terminates.

To find the first set of probabilities, calculate the relative frequencies of males and females in the survey:

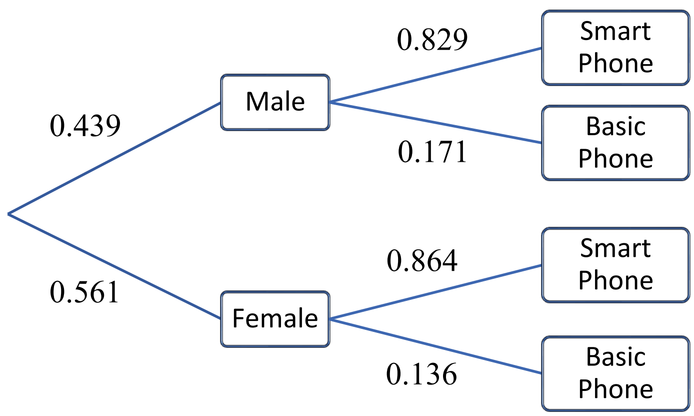
$$P(M) = \frac{1448}{3300} \approx 0.439, \quad P(F) = \frac{1852}{3300} \approx 0.561$$

The conditional probabilities are found by taking into account the given condition that the consumer is male or the consumer is female:

$$P(SP \mid M) = \frac{1201}{1448} \approx 0.829, \quad P(BP \mid M) = \frac{247}{1448} \approx 0.171$$

$$P(SP \mid F) = \frac{1601}{1852} \approx 0.864, \quad P(BP \mid F) = \frac{251}{1852} \approx 0.136$$

Label these probabilities on the tree diagram to give



<u>Practice</u>

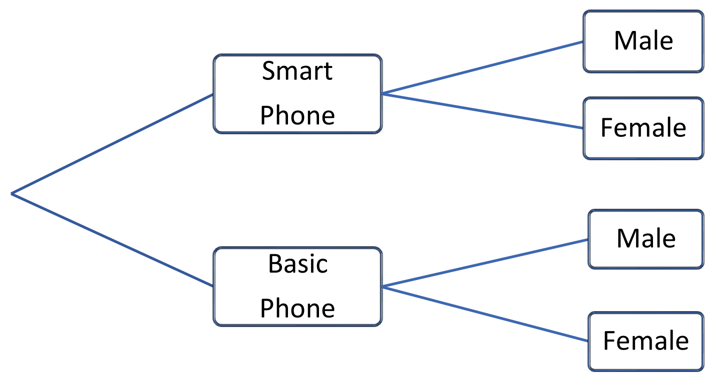
A survey was administered to a group of consumers who own a mobile phone. The results of the survey are below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

Find the probabilities on each branch of the tree diagram below.



Question 2 – How is conditional probability computed?

<u>Key Terms</u>

<u>Summary</u>

In Guided Example 1 of this section, we computed a conditional probability from the data below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

We defined the events follows:

M: Consumer is male

F: Consumer is female

BP: Consumer owns a basic phone

SP: Consumer owns a smart phone

To find the conditional probability $P(SP \mid F)$, we recognized that we are not interested in all consumers in the survey, only the female consumers. A total of 1852 female consumers took the survey. Of those female consumers, 1601 owned a smartphone. The probability of a consumer owning a smartphone given they are female is

$$P(SP \mid F) = \frac{1601}{1852}$$

Let's look at these numbers more closely. The denominator is the number of female consumers in the survey, $n(F)$. The numerator corresponds to female consumers who own a smartphone, $n(SP \text{ and } F)$. In words, these are consumers who are female and own a smartphone.

$$P(SP \mid F) = \frac{\overbrace{1601}^{n(SP \text{ and } F)}}{\underbrace{1852}_{n(F)}}$$

If we divide the top and bottom of this fraction by the number of people who took the survey,

$$P(SP \mid F) = \frac{\frac{1601}{3300}}{\frac{1852}{3300}}$$

we see that the top and the bottom are now relative frequencies and can be written as

$$P(SP \mid F) = \frac{P(SP \text{ and } F)}{P(F)}$$

This relationship allows us to write a conditional probability in terms of joint and marginal probabilities. In general, we can compute conditional probability with this relationship.

Conditional Probability

If *A* and *B* are events, the likelihood of *A* occurring given that *B* has occurred is

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

provided that $P(B) \neq 0$.

Notes

Guided Example 3	Practice
Suppose a batch of batteries is produced at a factory. A sample of batteries coming off the production line are sampled. From this sample, 32% of the batteries are mislabeled and 42% provide inadequate current. Twenty percent of the batteries are mislabeled and provide inadequate current. a. What is the probability that a battery provides inadequate current given that the battery is mislabeled? **Solution** Start by defining the events in the problem: M: Battery is mislabeled C: Battery provides inadequate current We can match the probabilities given in problem with events: $P(M) = 0.32$ $P(C) = 0.42$ $P(C \text{ and } M) = 0.20$ The question asks us to find the conditional probability $P(C \mid M)$. This may be found using the formula $P(C \mid M) = \dfrac{P(C \text{ and } M)}{P(M)}$ Substitute the probabilities given in the problem to get $P(C \mid M) = \dfrac{0.20}{0.32} = 0.625$ b. What is the probability the battery is mislabeled given that the battery provides inadequate current? **Solution** To find the probability $P(M \mid C)$, apply the formula for conditional probability and substitute the values given in the problem:	A mathematically inclined auto mechanic determines that 35% of repairs involve dead batteries and 15% of repairs involve bad alternators. Five percent of repairs involve dead batteries and bad alternators. a. What is the probability that a repair involves a dead battery given that the repair involves a bad alternator? b. What is the probability the repair involves a bad alternator given that the repair involves a dead battery?

$P(M \mid C) = \dfrac{P(M \text{ and C})}{P(C)}$ $= \dfrac{0.20}{0.42}$ ≈ 0.476	

Question 3 – What are independent events?

Key Terms

Independent events

Summary

When two events are **independent events**, one event occurring has no effect on the likelihood of the other event occurring.

Independent Events

If one event occurring does not change the likelihood of another event occurring, the two events are independent. This means that for events A and B,

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

In each case, the given events do not change the likelihood of the other events occurring. If one or both of the relationships are not equal, the events are said to be dependent.

Notes

Guided Example 4

A survey is administered to a group of consumers who own a mobile phone. The results of the survey are shown below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

Are the events SP and F independent events?

Solution In order for SP and F to be independent events,

$$P(SP \mid F) = P(SP) \text{ and } P(F \mid SP) = P(F)$$

We can calculate each of these four probabilities using the numbers in the table above.

$$P(SP) = \frac{n(SP)}{n(S)} = \frac{2802}{3300} \approx 0.849$$

$$P(SP \mid F) = \frac{n(SP \text{ and } F)}{n(F)} = \frac{1601}{1852} \approx 0.864$$

and

$$P(F) = \frac{n(F)}{n(S)} = \frac{1852}{3300} \approx 0.561$$

$$P(F \mid SP) = \frac{n(F \text{ and SP})}{n(SP)} = \frac{1601}{2802} \approx 0.571$$

In both cases, the probabilities are not equal, so the events are dependent. For the events to be independent, both equations would need to be true.

Practice

A survey is administered to a group of consumers who own a mobile phone. The results of the survey are shown below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

Are the events BP and M independent events?

Question 4 – What is the product rule for probability?

Key Terms

Product Rule for Probability

Summary

The rule for computing conditional property can be interpreted different. In Question 2, we defined the conditional probability $P(A \mid B)=\frac{P(A \text{ and } B)}{P(B)}$. If we multiply both side of this equation by $P(B)$, we get

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

We can also apply this strategy to the conditional probability $P(B \mid A)=\frac{P(B \text{ and } A)}{P(A)}$ to obtain a similar expression,

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

These expressions give the joint probability of *A* and *B* as a product of a conditional probability and a marginal probability.

Product Rule for Probability

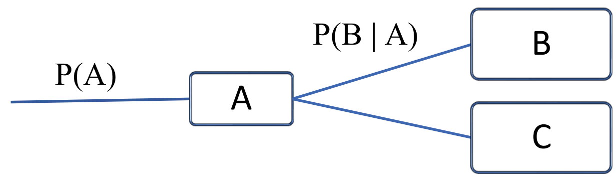
The probability of the event *A* and *B* is

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

or

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

We can utilize these relationships when we use a tree diagram. The probabilities on the right side of the second rule, $P(A \text{ and } B)=P(B \mid A)P(A)$, lie along the branch connecting to A followed by B. This means we can find the probability of A and B by multiplying the probabilities that connect to A followed by B.



Product Rule for Tree Diagrams

The product of all probabilities along a branch on a tree diagram is the likelihood of all events occurring that are on the branch.

<u>Notes</u>

Guided Example 5

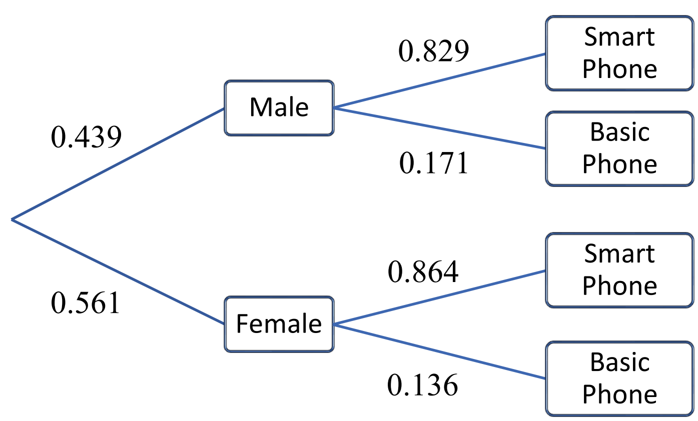
A survey is administered to a group of consumers who own a mobile phone. The results of the survey are shown below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

In an earlier example, we used these data and events to create the tree diagram below.



Compute the likelihood that a male consumer owns a smart phone.

Solution In terms of the events, we are being asked to find $P(M \text{ and } SP)$. Apply the formula for finding intersections of events to give

$$P(M \text{ and } SP) = P(SP \mid M)P(M)$$

The probabilities on the right side are found along the branch through Male and Smart Phone. Put these into the formula to yield

$$P(M \text{ and } SP) = (0.829)(0.439) \approx 0.364$$

Practice

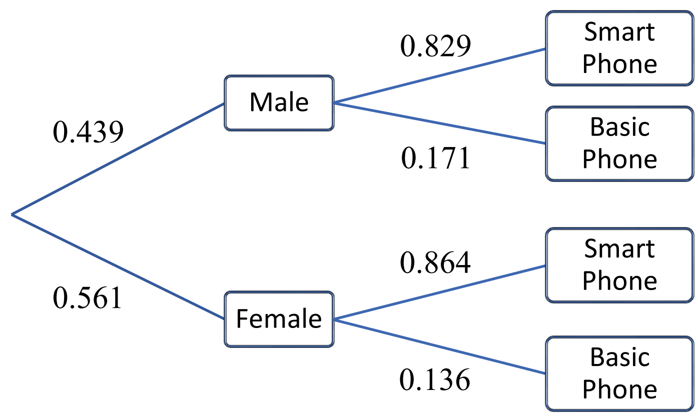
A survey is administered to a group of consumers who own a mobile phone. The results of the survey are shown below.

	Male	Female	Total
Basic Phone	247	251	498
Smart Phone	1201	1601	2802
Total	1448	1852	3300

Define the events below:

M: Consumer is male
F: Consumer is female
BP: Consumer owns a basic phone
SP: Consumer owns a smart phone

In an earlier example, we used these data and events to create the tree diagram below.



Compute the likelihood that a female consumer owns a basic phone.

Question 5 – How is Bayes' Rule used to compute conditional probability?

Key Terms

Bayes' Rule

Summary

In Question 2, we learned that the likelihood of an event A occurring given that an event B has already occurred is

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

We can also use the same basic expression to find the likelihood of an event B occurring given that an event A has already occurred,

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

Each of these expressions may be solved for the joint probability in the numerator to give

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

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

The joint event A and B is exactly the same event as the joint event B and A. This means their probabilities are also the same. Setting the left sides of these expressions equal gives

$$P(A \mid B)P(B) = P(B \mid A)P(A)$$

We can solve for either conditional probability, but if we solve for $P(B \mid A)$ we get the most basic form of **Bayes' Rule**.

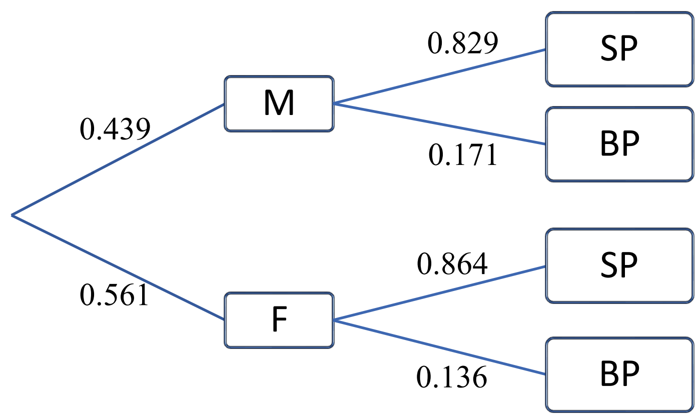
> **Bayes' Rule**
>
> If A and B are events, the conditional probability $P(B \mid A)$ may be computed in terms of $P(A \mid B)$ using
>
> $$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

This expression allows us to compute one conditional probability in terms of the "reverse" conditional probability. In practice, the most challenging part of using Bayes' Rule is identifying the events and computing the probabilities on the right side. We can simplify this task using a tree diagram.

Notes

<u>Guided Example 6</u>

Suppose you have the tree diagram below.



Use the tree diagram to compute $P(M \mid SP)$.

Solution Start by writing out the joint probabilities from the product rule for probabilities:

$$P(M \text{ and } SP) = P(M \mid SP)P(SP)$$
$$P(SP \text{ and } M) = P(SP \mid M)P(M)$$

Since the joint probabilities on the left side are equal, the expression on the left must be equal,

$$P(M \mid SP)P(SP) = P(SP \mid M)P(M)$$

Solving for $P(M \mid SP)$ gives the appropriate Bayes' Rule,

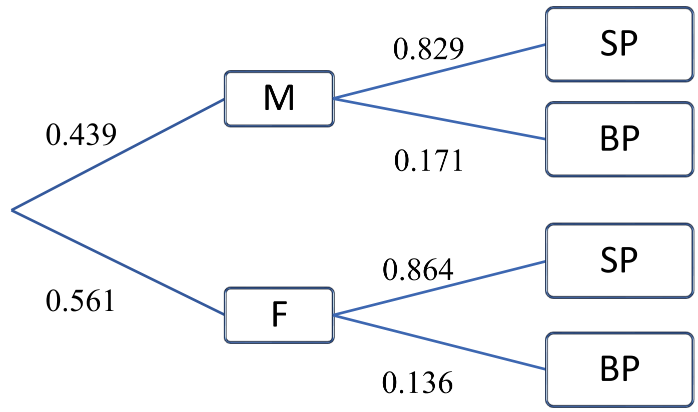
$$P(M \mid SP) = \frac{P(SP \mid M)P(M)}{P(SP)}$$

The probabilities for the numerator lie along the branch to M that continues to SP. To find the marginal probability in the denominator, locate all of the branches that terminate at SP. Multiply along these branches and then add the product to find the probability of SP. Putting this into Bayes' Rule yields

$$P(M \mid SP) = \frac{(0.829)(0.439)}{(0.829)(0.439)+(0.864)(0.561)} \approx 0.428$$

Practice

Suppose you have the tree diagram below.



Use the tree diagram to compute $P(F \mid BP)$.

Guided Example 7

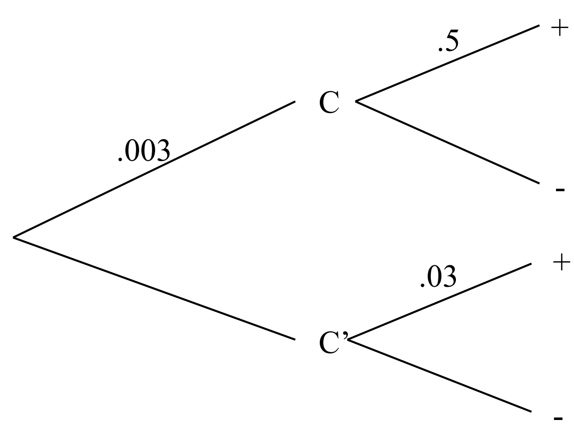
The probability of colorectal cancer can be given as .3%. If a person has colorectal cancer, the probability that the hemoccult test is positive is 50%. If a person does not have colorectal cancer, the probability that he still tests positive is 3%. What is the probability that a person who tests negative does not have colorectal cancer?

Solution To solve this problem, we'll draw and label an appropriate tree diagram. Then we'll apply Bayes' Rule to the problem.

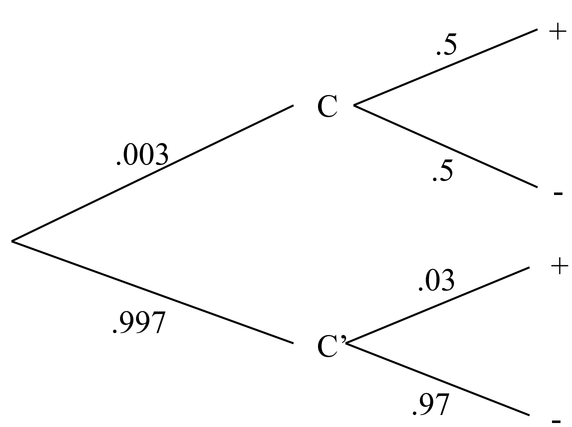
Look at the information given in the problem. If C is the event "person has colorectal cancer" and + is the event "the hemoccult test is positive", we know that

$$P(C) = .003 \quad P(+ \mid C) = .5 \quad P(+ \mid C') = .03$$

This suggests the following tree diagram:



Knowing that the sum of the probabilities from one point on the tree should add to 1, we can finish the tree diagram as follows:



The probability we are looking for is $P(C' \mid -)$. Notice that the tree diagram has $P(- \mid C')$, but not the reverse conditional probability that we are looking for. This is a sign we need to use Bayes' Rule. Let's find the appropriate form of Bayes' Rule. The definition of conditional probability applied to these events tells us

$$P(C' \mid -) = \frac{P(C' \cap -)}{P(-)}$$
$$P(- \mid C') = \frac{P(- \cap C')}{P(C')}$$

Solving each of these for the intersection yields

$$P(C' \mid -)P(-) = P(C' \cap -)$$
$$P(- \mid C')P(C') = P(- \cap C')$$

Since the intersection on the right-hand side is the same in each equation, we know the left-hand sides must be equal.

$$P(C' \mid -)P(-) = P(- \mid C')P(C')$$

Solving for $P(C' \mid -)$ gives

$$P(C' \mid -) = \frac{P(- \mid C')P(C')}{P(-)}$$

This is Bayes' Rule for this problem. Now we are ready to use the tree diagram. $P(- \mid C')$ and $P(C')$ are both labeled on the tree diagram. We can calculate $P(-)$ by following the branches on the tree diagram (multiply) that lead to a negative result, and then summing up the products from these branches.

$$P(-) = (.003)(.5) + (.997)(.97) = .96859$$

Putting these values into Bayes' Rule gives

$$P(C' \mid -) = \frac{(.97)(.997)}{.96859} \approx .9985$$

This means that is you test negative, the likelihood that you do not have colorectal cancer is 99.85%. The test is quite good at screening that you do not have the disease.

Practice

The probability of colorectal cancer can be given as .3%. If a person has colorectal cancer, the probability that the hemoccult test is positive is 50%. If a person does not have colorectal cancer, the probability that he still tests positive is 3%. What is the probability that a person who tests positive does have colorectal cancer?

Section 7.1

Problem 1 a) $S = \{0,1,2,3,4\}$,

b) $S = \{YYYY, YNNN, NYNN, NNYN, NNNY, NYYY, YNYY, YYNY, YYYN, YYNN, YNNY, NNYY, NYYN, YNYN, NYNY, NNNN\}$

c) $\{YYYY, YYYN, YYNY, YNYY, NYYY\}$

Problem 2 a) No, b) Yes, c) No

Problem 3 ¼

Problem 4 $\frac{76}{108}$

Section 7.2

Problem 1 $\frac{13}{52}$

Problem 2 $\frac{13}{20}$

Problem 3 a) $\frac{38659}{159172} \approx 0.243$, b) $\frac{4023}{159172} \approx 0.025$

Problem 4 ¾

Problem 5 a) $\frac{7}{20}$, b) $\frac{18}{20}$

Problem 6 $\frac{120513}{159172} \approx 0.757$

Problem 7 a) B and A, b) Needs a new battery or needs a new alternator, c) 0.55

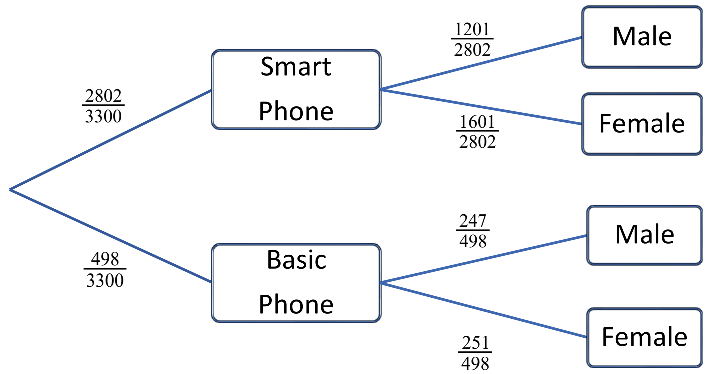
Problem 8 $\frac{10}{36} \approx 0.278$

Problem 9 a) $\frac{425}{5740} \approx 0.074$, b) $\frac{2835}{5740} \approx 0.494$, c) $\frac{725}{5740} \approx 0.126$, d) $\frac{3135}{5740} \approx 0.546$

Section 7.3

Problem 1 a) $\frac{1448}{3300}$, b) $\frac{498}{3300}$, c) $\frac{247}{3300}$, d) $\frac{247}{498}$

Problem 2



Problem 3 a) approximately 0.333, b) approximately 0.143

Problem 4 No

Problem 5 approximately 0.076

Problem 6 approximately 0.504

Problem 7 approximately 0.048