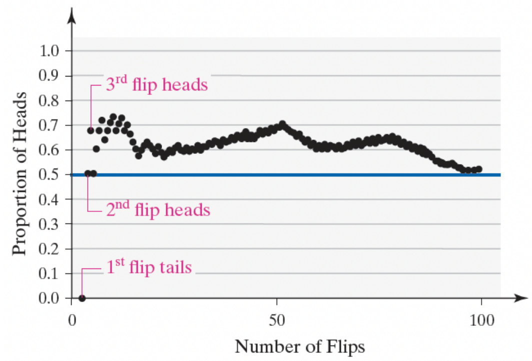
CHAPTER 5: PROBABILITY

[This chapter is based on Chapter 5 from the textbook]

The reason to study statistics is that we cannot predict the exact outcome of an experiment. For example, if we toss a coin, we cannot predict whether the outcome will be heads or tails. However, as we toss the same coin repeteadly, we know that the proportion of heads will be approximately $\frac{1}{2}$. In this example, analyzing the number of heads after doing the experiment 100 times would be statistics, and is related to Chapters 1-4. In this chapter, we will focus on analyzing the possible results before running any experiments or observing any data.

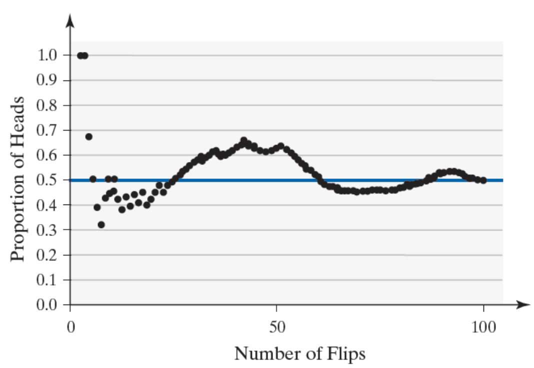
5.1 Probability rules

Let's consider a simple example. In the next figure, we show the proportion of heads with respect to the number of flips of a coin.



The first flip shows tails, so the proportion of heads is $0/1 = 0$; the second is heads, so the proportion becomes $1/2$; the third is heads again, so the proportion of heads up to the third toss is $2/3$, etc. Observe that, as we toss more coins, the proportion of heads gets closer to $1/2$.

If we repeat the experiment from scratch, we will get a different graph:



Observe that the biggest differences are in the first 10 flips (or so). In both cases, when the number of flips is small, the proportion rapidly changes. However, as the number of flips increases, the graphs become more similar.

Definition 5.1. *Probability is the measure of the likelihood of a random phenomenon or chance behavior occurring. It deals with experiments that yield random short-term results or outcomes.*

The definition above establishes that, as we repeat an experiment, we start seeing the proportion of time an outcome can be obtained, that is, its probability. In other words, as we repeat an experiment, we start observing the natural or inherent proportion of times that we will observe each outcome.

Theorem 5.1. *The Law of Large Numbers establishes that, as the number of repetitions of a probabilistic experiment increases, the proportion with which a certain outcome is observed gets closer to the probability of the outcome.*

We just saw an example of how the law of large numbers explains the behavior of the proportion of heads if we flip a coin repeteadly.

Definition 5.2. *An experiment is any process with uncertain results that can be repeated.*

For example, tossing a coin, rolling a die, counting the number of students in class on a given day, etc. Each of these experiments, have a set of possible values that make sense.

Definition 5.3.

(i) The sample space of a probability experiment, denoted by S, is the collection of all possible outcomes.

(ii) An event is any collection of outcomes from a probability experiment, that is, a subset of the sample space S. In general, we use E to denote events.

(iii) If an event only has one outcome, we call it simple event and we typically denote it as e_i, where i is a number.

In the example of flipping a coin, the experiment is flipping a coin, the sample space is $S = \{\text{Tails, Heads}\}$ and a simple event is obtaining Heads. Let's see another example.

Example 5.1. *A probability experiment consists of rolling a single fair die.*

(a) Identify the outcomes of the probability experiment

(b) Determine the sample space

(c) Define the event $E =$ "Roll an even number"

Solution.

(a) The outcomes from rolling a die can be 1, 2, 3, 4, 5, 6.

(b) The sample space is the set of possible outcomes, that is, $S = \{1, 2, 3, 4, 5, 6\}$

(c) The event "Roll an even number" occurs if we obtain a 2, 4, or 6. Then, $E = \{2, 4, 6\}$.

□

We defined probability of an outcome as the proportion of times we see the desired outcome when the number of experiments increases. Since probabilities are proportions, they must satisfy the following rules:

Theorem 5.2. *Let $P(E)$ denote the probability that event E occurs. Then,*

1. *The probability of any event E, satisfies:*

$$0 \leq P(E) \leq 1$$

2. *The sum of the probabilities of all outcomes must be 1. That is, if $S = \{e_1, e_2, \ldots, e_n\}$, then*

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

The first rule establishes that proportions always are between 0 and 1, and the second rule establishes that the total proportion of each outcome must add up to 1.

Now, not all the events will have a probability that is strictly greater than zero or strictly smaller than 1.

Definition 5.4.

(i) An event is impossible if its probability is 0

(ii) An event is certainty if its probability equals 1

For example, if we are rolling a die, the event "Roll a 7" is impossible, and the event "Roll a positive integer" is certainty.

Example 5.2. *Determine if the following two tables represent a probability model.*

Model 1:

Color	*Probability*
Red	*0*
Green	*0.1*
Blue	*0.1*
Brown	*0.3*
Yellow	*0.15*
Orange	*0.35*

Model 2:

Color	*Probability*
Red	*0.3*
Green	*0.1*
Blue	*0.1*
Brown	*0.3*
Yellow	*0.15*
Orange	*0.35*

Solution. In each case, we check that all the probabilities are between 0 and 1, and that they add up to 1.

- In model 1, all the probabilities are nonnegative and smaller than 1. Further, if we add them up we obtain

$$0 + 0.1 + 0.1 + 0.3 + 0.15 + 0.35 = 1$$

Hence, it is a probability model.

- In this case all the probabilities are nonnegative and smaller than 1 too. However, if we add them up, we obtain

$$0.3 + 0.1 + 0.1 + 0.3 + 0.15 + 0.35 = 1.3 \neq 1$$

Hence, model 2 is not a probability model.

□

The closer a probability is to 1, the more likely an event will occur. Similarly, the closer a probability is to zero, the less likely it will occur.

Now, how do we compute probabilities? We show the general steps in the following box.

Theorem 5.3 (Computing probability)**.** *If an experiment has sample space S, and for any set A, $N(A)$ denotes the number of outcomes in A, we have that for any event E,*

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

That is, the probability of an event E is the proportion of outcomes in which E occurs with respect to the sample size. Let's do some examples.

Example 5.3. *Consider the experiment of rolling a single fair die.*

(a) What is the probability of rolling a 3?

(b) What is the probability of rolling an even number?

Solution. In this problem, the sample size is $S = \{1, 2, 3, 4, 5, 6\}$ and has 6 elements, that is, $N(S) = 6$.

(a) In this case the event E of rolling a 3 only has one element, that is, $E = \{3\}$. Then,

$$P(E) = \frac{N(E)}{N(S)} = \frac{1}{6}$$

(b) In this case, the event E of rolling an even number is $E = \{2, 4, 6\}$. Then, $N(E) = 3$ and we obtain

$$P(E) = \frac{3}{6} = \frac{1}{2}$$

□

Example 5.4. *A pair of fair dice is rolled.*

(a) Compute the probability of rolling a 7

(b) Compute the probability of rolling "snake eyes," that is, compute the probability of rolling a two.

(c) Comment on the likelihood of rolling a 7 vs. rolling a 2.

Solution. In this case we are rolling 2 dice together. Then, the sample space involves the outcome from both dice. Let's use (i, j) to denote that the first die rolled an i, and the second, j. Then, the sample space is:

$$S = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

Then, $N(S) = 36$.

(a) The event of rolling a 7 is $E = \{(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1)\}$. Then,

$$P(E) = \frac{N(E)}{N(S)} = \frac{6}{36} = \frac{1}{6}$$

(b) In this case, the event can be described as $F = \{(1,1)$. Then,

$$P(F) = \frac{N(F)}{N(S)} = \frac{1}{36}$$

(c) Rolling a 7 has probability $P(E) = 6/36$ and rolling a two has probability $P(F) = 1/36$. Since $P(E) > P(F)$, rolling a 7 is more likely than rolling a 2. Further, rolling a 7 has 6 times the probability of rolling a 2. If we repeat this experiment a large number of times, we expect seeing a 7 six times more than seeing a 2.

□

Example 5.5. *Sophia has three tickets to a concert, but Alice, Bob, Charlie and Denisse all want to go with her. To be fair, Sophia randomly selects the two people that can go with her.*

(a) Determine the sample space of the experiment

(b) Compute the probability of the event $E =$"Bob and Charlie attend the concert"

(c) Compute the probability of the event $F =$"Denisse attends the concert"

(d) Interpret the probability in part (c)

Solution.

(a) We use the initial of each person to denote the state space. We obtain

$$S = \{AB, AC, AD, BC, BD, CD\}$$

Observe that $N(S) = 6$.

(b) The event $E = \{BC\}$. Then,

$$P(E) = \frac{1}{6}$$

(c) The event F contains all pairs that include Denisse, that is, $F = \{AD, BD, CD\}$. Then,

$$P(F) = \frac{3}{6} = \frac{1}{2}$$

(d) If we repeat the experiment many times, 1/2 of the times Denisse goes to the concert.

□

5.2 The Addition Rule and Complements

To learn how to compute probabilities, we need the following definition.

Definition 5.5. *Two events are disjoint if they have no outcomes in common. Another name for disjoint events is mutually exclusive.*

In other words, two events are disjoint if they cannot happen at the same time.

Theorem 5.4 (Computing probability of disjoint events)**.** *If E and F are two disjoint events, then $P(E \text{ or } F) = P(E) + P(F)$.*

We can think of computing the probability of two events as computing the area used by these events in a Venn's diagram. We illustrate this idea in the following example.

Example 5.6. *Suppose we randomly select chips from a bag. Each chip is labeled as 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and these labels are equally likely.*

What is the probability that we choose a number less than or equal to 2, or a number greater than or equal to 8?

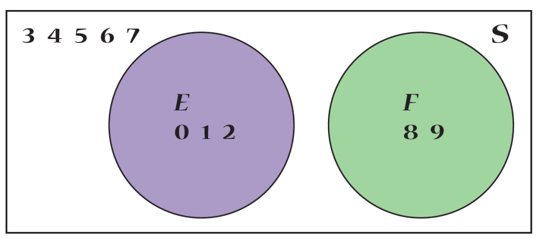
Solution. We first define our events:

- Let E be the event "choose a number less than or equal to 2"
- Let F be the event "choose a number greater than or equal to 8"

Then, we have:

$$E = \{0, 1, 2\} \qquad \text{and} \qquad F = \{8, 9\}$$

Since there is no label repeated in both, the events E and F are disjoint. The following picture represents the sample space and the events:



where the square represents the sample space S and part of the sample space is used by the events E and F. Since the events E anf F do not overlap, the probability of E or F is

$$\begin{aligned} P(E \text{ or } F) &= P(E) + P(F) \\ &= \frac{N(E)}{N(S)} + \frac{N(F)}{N(S)} \\ &= \frac{3}{10} + \frac{2}{10} \\ &= \frac{5}{10} = 0.5 \end{aligned}$$

□

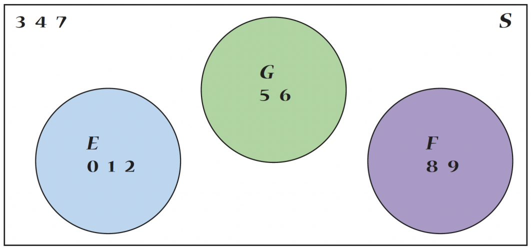
If we have more than two events that are disjoint, we can generalize the rule above. Specifically, we have the following rule.

Theorem 5.5 (Computing probability of more than two disjoint events)**.** *If $E_1, E_2, E_3, \ldots$ have no outcomes in common, then*

$$P(E_1 \text{ or } E_2 \text{ or } E_3 \text{ or } \ldots) = P(E_1) + P(E_2) + P(E_3) + \cdots$$

Example 5.7. *Consider the situation from Example 5.6 and now define the event G as "choose a 5 or a 6". What is $P(E \cup F \cup G)$?*

Solution. In this case, the event $G = \{5, 6\}$ and is disjoint with E and with F. Indeed, a Venn diagram of the three events is



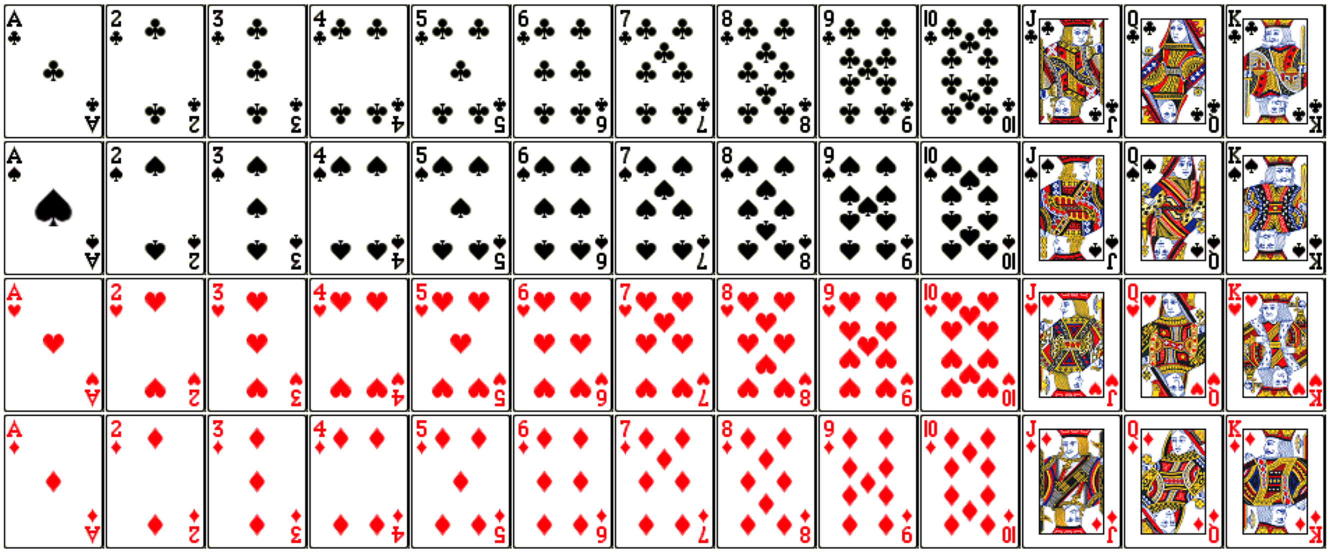
Then,

$$\begin{aligned} P(E \cup F \cup G) &= P(E) + P(F) + P(G) \\ &= \frac{N(E)}{N(S)} + \frac{N(F)}{N(S)} + \frac{N(G)}{N(S)} \\ &= \frac{3}{10} + \frac{2}{10} + \frac{2}{10} \\ &= \frac{7}{10} = 0.7 \end{aligned}$$

□

Let's do another example.

Example 5.8. *Consider a deck of 52 cards as shown in the figure:*



There are 4 suits (clubs, spades, hearts and diamonds in descending order) and each suit has 13 cards: A (representing 1), numbers from 2 to 10, jacks (representing 11), queen (representing 12) and king (representing 13).

(a) Compute the probability of drawing a king

(b) Compute the probability of drawing a king, a queen or a jack

Solution. We first define the following events:

- E represents "drawing a king"
- F represents "drawing a queen"

- G represents "drawing a jack"

Then, part (a) is:

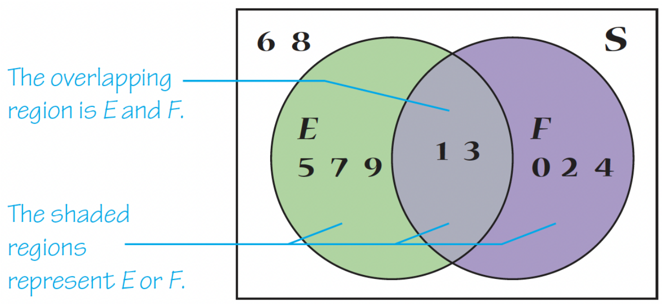
$$P(E) = \frac{N(E)}{N(S)} = \frac{4}{52} = \frac{1}{13}$$

To compute part (b), we use that the events E, F and G do not have any outcomes in common. Then,

$$\begin{aligned} P(E \text{ or } F \text{ or } G) &= P(E) + P(F) + P(G) \\ &= \frac{N(E)}{N(S)} + \frac{N(F)}{N(S)} + \frac{N(G)}{N(S)} \\ &= \frac{4}{52} + \frac{4}{52} + \frac{4}{52} \\ &= \frac{12}{52} = \frac{3}{13} \end{aligned}$$

□

In general, we may have events that overlap in some outcomes. For example, if we recall Example 5.6 and now consider the events $E =$"choose an odd number" and $F =$"choose a number less than or equal to 4, we cannot use the previous formula. In a Venn's diagram, this situation can be drawn as follows:



In this case, $P(E \text{ or } F)$ should be the probability of choosing 0,1,2,3,4,5,7,9, which is

$$P(E \text{ or } F) = \frac{8}{10} = 0.8$$

However, if we use the formula for disjoint events, we obtain

$$P(E \text{ or } F) = P(E) + P(F) = \frac{5}{10} + \frac{5}{10} = 1$$

that is, we are overestimating $P(E \text{ or } F)$. Further, we obtained probability 1 of an event that does not include the outcomes 6 and 8 (which are not impossible). So, what is the problem? The problem is that when we use the disjoint sets rule, we are counting the outcomes 1 and 3 twice: once in event E and once in event F. Then, we only need to subtract it. Hence, we obtain the following rule:

Theorem 5.6 (The general addition rule). *For any two events E and F,*

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F)$$

Indeed, if we use this rule for the example above, we have that E and F have 1 and 3, then,

$$\begin{aligned} P(E \text{ or } F) &= P(E) + P(F) - P(E \text{ and } F) \\ &= \frac{N(E)}{N(S)} + \frac{N(F)}{N(S)} - \frac{N(E \text{ and } F)}{N(S)} \\ &= \frac{5}{10} + \frac{5}{10} - \frac{2}{10} \\ &= \frac{8}{10} = \frac{4}{5} \end{aligned}$$

Let's apply this rule in the card-deck example.

Example 5.9. *Suppose that we have a card deck and we select one card. What is the probability that we draw a king or a diamond?*

Solution. We first introduce the notation. Suppose that E representa the event "draw a king" and F represents the event "draw a diamond." Since there is exactly one king of diamonds, the events E and F are not disjoint. Indeed,

$$\begin{aligned} P(E \text{ or } F) &= P(E) + P(F) - P(E \text{ and } F) \\ &= \frac{N(E)}{N(S)} + \frac{N(F)}{N(S)} - \frac{N(E \text{ and } F)}{N(S)} \\ &= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} \\ &= \frac{16}{52} = \frac{4}{13} \end{aligned}$$

□

The deck of cards is an example of situations where the possible outcomes have two features; in this case, suit and number. That is, we can find every number in each of the four suits, and each suit has all the numbers. A more general case is presented in the following example, where the two features are specified in a table. In this case, we look at marital status and gender.

Example 5.10. *The following table represents the marital status of males and females 15 years old or older in the United States in 2013*[1]

Marital status	*Males (millions)*	*Females (millions)*
Never married	*41.6*	*36.9*
Married	*64.4*	*63.1*
Widowed	*3.1*	*11.2*
Divorced	*11.0*	*14.4*
Separated	*2.4*	*3.2*

Determine the probability that a randomly selected U.S. resident 15 years old or older:

(a) is male

(b) is widowed

(c) is widowed or divorced

(d) is male or widowed

Solution. In this case, the state space is the U.S. population, that is, the list of adults 15 years old or older. Then, the number of possible outcomes in the state space is

$$N(S) = 41.6 + 64.4 + 3.1 + 11.0 + 2.4 + 36.9 + 63.1 + 11.2 + 14.4 + 3.2 = 251.3$$

(a) We are interested in the probability of the event $E =$"male." Then,

$$P(E) = \frac{N(E)}{N(S)} = \frac{41.6 + 64.4 + 3.1 + 11.0 + 2.4}{251.3} = \frac{122.5}{251.3} = 0.4875$$

(b) Now we are interested in the probability of the event $F =$"widowed." Then,

$$P(F) = \frac{N(F)}{N(S)} = \frac{3.1 + 11.2}{251.3} = \frac{14.3}{251.3} = 0.0569$$

(c) Now we are interested in the event $G =$"widowed or divorced." Observe that a person cannot be widowed and divorced at the same time. Then, the events $F =$"widowed" and $D =$"divorced" are disjoint. Then,

$$P(G) = P(F) + P(D)$$

[1] Source: U.S. Census Bureau, Current Population Reports

$$
\begin{aligned}
&= 0.0569 + \frac{N(D)}{N(S)} \\
&= 0.0569 + \frac{11.0 + 14.4}{253.1} \\
&= 0.0569 + 0.1003 \\
&= 0.1579
\end{aligned}
$$

where we used the result from part (b) to compute $P(F)$.

(d) Now we are interested in the event $H =$"male or widowed." Using the events that we have already introduced, we have $H = E$ or F. These events are not disjoint because a person can be male and widowed at the same time. Hence, we use the general formula from Theorem 5.6. We obtain:

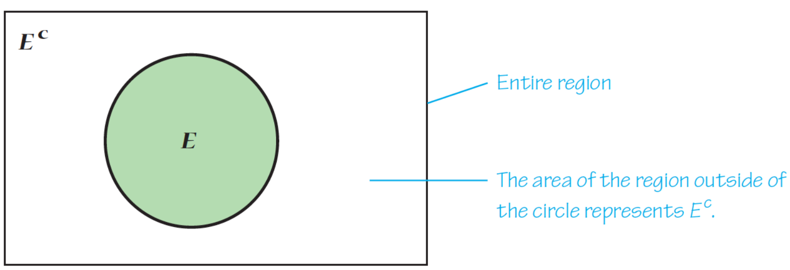
$$
\begin{aligned}
P(H) &= P(E \text{ or } F) \\
&= P(E) + P(F) - P(E \text{ and } F) \\
&= 0.4875 + 0.0569 - \frac{N(E \text{ and } F)}{N(S)} \\
&= 0.4875 + 0.0569 - \frac{3.1}{251.3} \\
&= 0.4875 + 0.0569 - 0.0123 \\
&= 0.5321
\end{aligned}
$$

□

The last topic of this section is computing the probability of the complement of an event. We start defining the complement.

Definition 5.6. *The complement of an event E is denoted by E^c, and is all outcomes in the sample space S that are not outcomes in the event E.*

The following figure illustrates the complement in a Venn diagram.



We have the following rule to compute the probability of the complement of an event.

Theorem 5.7. *If E represents and event and E^c represents the complement of E, then*

$$P(E^c) = 1 - P(E)$$

Let's see some examples.

Example 5.11. *The data in the following table*[2] *represent the income distribution of households in the United States in 2013.*

Annual income	*Number (in thousands)*
Less than \$10,000	*8,940*
Between \$10,000 and \$14,999	*6,693*
Between \$15,000 and \$24,999	*13,898*
Between \$25,000 and \$34,999	*12,756*
Between \$35,000 and \$49,999	*16,678*
Between \$50,000 and \$74,999	*21,659*
Between \$75,000 and \$99,999	*14,687*
Between \$100,000 and \$149,999	*15,266*
Between \$150,000 and \$199,999	*6,463*
\$200,000 or more	*5,913*

Compute the probability that a randomly selected household earned the following incomes in 2013:

(a) \$200,000 or more

(b) Less than \$200,000

(c) At least \$10,000

Solution. The state space are all the households. Then,

$$N(S) = 8,940 + 6,693 + 13,898 + \cdots + 5,913 = 122,953$$

(a) We compute the probability of the event $E =$"\$200,000 or more." We obtain:

$$P(E) = \frac{N(E)}{N(S)} = \frac{5,913}{122,953} = 0.048$$

(b) In this case, we are interested in the probability of the event $E^c =$"less than \$200,000." We use the complement rule from Theorem 5.7 and obtain:

$$P(E^c) = 1 - P(E) = 1 - 0.048 = 0.952$$

(c) We use a similar approach. In this case, we define the event $F =$"Less than \$10,000" and observe that we need the probability of F^c. We obtain

$$P(F^c) = 1 - P(F) = 1 - \frac{N(F)}{N(S)} = 1 - \frac{8,940}{122,953} = 1 - 0.0727 = 0.9273$$

□

[2] Source: U.S. Census Bureau

5.3 Independence and Multiplication Rule

In the previous section we focused on computing the probability that event E or F occurred. In this section, we focus on computing the probability that events E and F occur. We start with a definition.

Definition 5.7.

- *Two events E and F are independent if the occurrence of event E does not affect the probability of occurrence of event F.*
- *Two events E and F are dependent if he occurrence of event E affects the probability of event F.*

Let's see some examples.

Example 5.12. *Are the events E and F independent or dependent in the following situations?*

(a) You flip a coin and roll a die. Suppose $E =$obtain a head, and $F =$roll a 5

(b) $E =$earned a bachelor's degree, and $F =$earn more than \$100,000/year

(c) The events E and F are disjoint.

Solution.

(a) The outcome of flipping a coin does not affect the outcome of rolling a die, so E and F are independent events. In other words, obtaining a head does not affect the probability of rolling a 5.

(b) The education level of a person does affect their income, so the events E and F are not independent. In other words, earning a bachelor's degree increments the probability that a person earns more than \$100,000/year. Hence, E and F are not independent.

(c) Disjoint events are dependent because they do not overlap; that is, if we know that E occurred, it is impossible that F also occurred. For example, if the experiment is rolling a die and observing the number that faces up, the event E=roll an odd number and F =roll an even number, then E and F are disjoint. However, if we know that E occurred (that is, if we know that we rolled an odd number), the probability of F is 0. Hence, the occurrence of E makes the occurrence of F impossible in the same experiment.

□

Now that we know what independent events are, we can learn how to use this concept to compute probabilities.

Theorem 5.8 (Multiplication rule for independent events)**.** *If E and F are independent events, then*

$$P(E \text{ and } F) = P(E) \cdot P(F)$$

More generally, if $E_1, E_2, \ldots, E_n$ are n independent events, then

$$P(E_1 \text{ and } E_2 \text{ and } E_3 \text{ and } \cdots \text{ and } E_n) = P(E_1) \cdot P(E_2) \cdot P(E_3) \cdots P(E_n)$$

Let's see some examples.

Example 5.13. *Suppose that you roll a die twice.*

(a) What is the probability that you obtain a 1 twice?

(b) What is the probability that the sum of the outcomes is 3?

(c) What is the probability that the sum of the two outcomes is at least 4?

Solution. The outcomes from rolling a die twice are independent. Then, we use the theorem above to solve in each case.

(a) Let $E =$ roll a 1 the first time, and $F =$ roll a 1 the second time. Then,

$$\begin{aligned} P(E \text{ and } F) &= P(E) \cdot P(F) \qquad \text{(because } E \text{ and } F \text{ are independent)} \\ &= \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \end{aligned}$$

(b) In this case, the event of interest is "the sum of the outcomes is 3." The first task is to describe this event as "and" and "or" statements to be able to compute the probability. We have

$$F = \text{"roll a 1 in the first roll and a 2 in the second" or "roll a 2 in the first roll and a 1 in the second"}$$

Let's use E_1 to denote "roll a 1 in the first roll," E_2 to denote "roll a 2 in the first roll," D_1 to denote "roll a 1 in the second roll," and D_2 to denote "roll a 2 in the second roll." Then,

$$P(F) = P\big([E_1 \text{ and } D_2] \text{ or } [E_2 \text{ and } D_1]\big)$$

Observe that the events $F_1 = [E_1 \text{ and } D_2]$ and $F_2 = [E_2 \text{ and } D_1]$ are disjoint. Then,

$$\begin{aligned}
P(F) &= P(F_1 \text{ or } F_2) \\
&= P(F_1) + P(F_2) \\
&= P(E_1 \text{ and } D_2) + P(E_2 \text{ and } D_1) \\
&= P(E_1) \cdot P(D_2) + P(E_2) \cdot P(D_1) \\
&= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} \\
&= \frac{2}{36} \approx 0.0556
\end{aligned}$$

(c) Let's use G to denote the event that the sum of the two outcomes is at least 4, and A_i to denote that the sum of the two outcomes is i. Then,

$$P(G) = P(A_4 \text{ or } A_5 \text{ or } \cdots \text{ or } A_{36})$$

These are too many events, and will be hard to do. Then, we will use the complement. That is,

$$\begin{aligned}
P(G) &= 1 - P(G^c) \\
&= 1 - P(\text{ sum is at most } 3) \\
&= 1 - P(A_2 \text{ or } A_3) \\
&= 1 - \Big(P(A_2) + P(A_3)\Big) \qquad \text{(because } A_2 \text{ and } A_3 \text{ are disjoint events)}
\end{aligned}$$

Now we compute $P(A_2)$ and $P(A_3)$ separately. Notice that A_3 is exactly the same event as F from part (c). Then,

$$P(A_3) = \frac{2}{36}$$

Now we compute $P(A_2)$, and we use the notation from part (c), that is, E_i denotes that the outcome of the first roll is i, and D_i denotes that the outcome of the second roll is i. Then,

$$\begin{aligned}
P(A_2) &= P(E_1 \text{ and } D_1) \\
&= P(E_1) \cdot P(D_1) \qquad \text{(because the events are independent)} \\
&= \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}
\end{aligned}$$

Therefore,

$$\begin{aligned}
P(G) &= 1 - \left(\frac{1}{36} + \frac{2}{36}\right) \\
&= 1 - \frac{3}{36} \\
&= \frac{33}{36} \approx 0.9167
\end{aligned}$$

□

5.4 Conditional Probability and the General Multiplication Rule

In the last section we learned how to compute the probability of two events happening together (E and F) when they are independent. However, independent events are not that common in real life. In this section, we will learn how to compute probabilities when the two events are dependent. We start with a definition.

Definition 5.8. *The notation $P(E|F)$ is read conditional probability of the event E given event F, or probability of E given F, and represents the probability that event E occurs when we already know that event F occured.*

In other words, when we condition on the event F, the state space is reduced from S to F. Let's see some examples.

Example 5.14. *Suppose that we roll a die once.*

(a) What is the probability that the die comes up 3? What is the probability that the die comes up 4?

(b) Now suppose that, after rolling the die, I see the number that comes up before you and I give you the hint "the number is odd." What is the probability that the die came up 3? What is the probability that the die comes up 4?

Solution. The difference between both cases is the amount of information we have. Indeed, we obtain different probabilities.

(a) There is nothing new in this question. Let's introduce the following notation for our events:

$$D = \text{"roll a 3"}$$
$$E = \text{"roll a 4"}$$

Since all the outcomes of the die are equally likely, we obtain

$$P(D) = \frac{1}{6} \qquad \text{and} \qquad P(E) = \frac{1}{6}$$

(b) In this case we have additional information: the number is odd. There are three odd numbers and they are equally likely. Hence, the probability of rolling a 3 given that the number is odd is $\frac{1}{3}$; and the probability of rolling a 4 given that the number is odd is 0 because 4 is an even number.

To write these results using the definition above, we introduce the following event:

$$F = \text{"roll an odd number"}$$

Then,

$$P(D|F) = \frac{1}{3} \qquad \text{and} \qquad P(E|F) = 0$$

□

Example 5.15. *The data in the following table represent the marital status of males and females 15 years old or older in the United States in 2013. Suppose that we randomly select a woman. What is the probability that her marital status is widowed?*

Marital status	*Males (millions)*	*Females (millions)*	**Total (millions)**
Never married	*41.6*	*36.9*	**78.5**
Married	*64.4*	*63.1*	**127.5**
Widowed	*3.1*	*11.2*	**14.3**
Divorced	*11.0*	*14.4*	**25.4**
Separated	*2.4*	*3.2*	**5.6**
Total (millions)	**122.5**	**128.8**	**251.3**

Solution. In this case, we want the probability of widowed given that the individual is a woman. Then, defining the events:

$$E = \text{“widowed”}$$
$$F = \text{“woman”}$$

We want to compute $P(E|F)$. Since the state space now is F, then we obtain the following probability:

$$P(E|F) = \frac{11.2}{128.8} = 0.087$$

□

The last example gives an intuition to a general rule to compute conditional probabilities. We establish the rule below.

Theorem 5.9. *If E and F are two events on the same state space, then*

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)} = \frac{N(E \text{ and } F)}{N(F)}$$

In the example above, 11.2 represents the number of individuals that are widowed and female, and 128.8 represents the number of female individuals.

The theorem above yields the general multiplication rule. We formally state it below.

Theorem 5.10 (General multiplication rule)**.** *The probability that two events E and F both occur is*

$$P(E \text{ and } F) = P(E|F) \cdot P(F)$$

Notice that if E and F are independent, $P(E|F) = P(E)$ because the occurrence of F does not affect the probability of E. In such case, we recover the multiplication rule for independent events. Let's see some examples.

Example 5.16. *The probability that a driver who is speeding gets pulled over is 0.8. The probability that a driver gets a ticket, given that they are pulled over, is 0.9. What is the probability that a randomly selected driver who is speeding gets pulled over and gets a ticket?*

Solution. We first define our events. Let

$$E = \text{“get a ticket”}$$
$$F = \text{“being pulled over while speeding”}$$

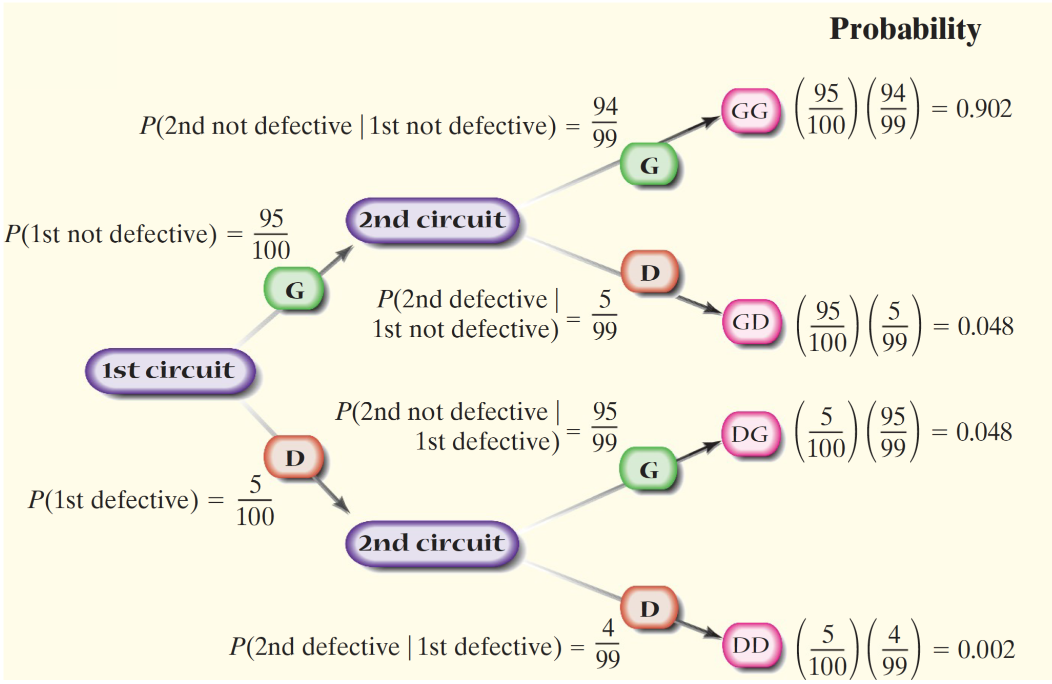
Then, we are interested in $P(E$ and $F)$ and the events E and F are not independent. Then, we use the general rule and obtain:

$$P(E \text{ and } F) = P(E|F) \cdot P(F) = 0.9 \cdot 0.8 = 0.72$$

□

Example 5.17. *Suppose that of 100 circuits sent to a manufacturing plant, 5 are defective. The plant manager receiving the circuits randomly selects 2 and tests them. If both circuits work, she will accept the shipment. Otherwise, the shipment is rejected. What is the probability that the plant manager discovers at least 1 defective circuit and rejects the shipment?*

Solution. Observe that in this case, we are using conditional probability to deal with a sequential experiment. We first select 1 of the 100 parts, and then we select 1 of the remaining 99 parts. The probability that the first part selected is defective is $\frac{5}{100}$, but the probability that the second part is defective depends on whether we selected a good or a defective part the first time. The possibilities of this experiment are represented in the following tree, where D represents defective and G represents good:



Then, the probability that at least one part is defective is:

$$\begin{aligned} P(\text{at least 1 defective}) &= P(GD \text{ or } DG \text{ or } DD) \\ &= P(GD) + P(DG) + P(DD) \qquad \text{(since the events are disjoint)} \\ &= 0.048 + 0.048 + 0.002 = 0.098 \end{aligned}$$

Observe that the tree was constructed using the definition of conditional probability and the general multiplication rule. □

Example 5.18 (Monty Hall problem)**.** *You are on TV and the presenter gives you the chance to win a car. The presenter shows you 3 doors and tells you that the car is behing one of the doors (he knows which one). You don't have any information about where the car is, so you choose a door uniformly at random. Let's say that you choose door 1.*

After you choose door number 1, the presenter reveals a door with a goat (of course, not door 1) and asks if you want to change your choice to the other closed door.

What is the probability that you win the car if you stay at door 1? What is the probability that you win the car if you change your choice?

Solution. This experiment is (slightly) similar to the manufacturer and parts because there are two steps, and in the second step you have more information than in the first step.

Let's define the following events:

$$\begin{aligned} C_1 &= \text{“car behind door 1”} \\ C_2 &= \text{“car behind door 2”} \\ C_3 &= \text{“car behind door 3”} \\ P_2 &= \text{“presenter opens door 2”} \\ P_3 &= \text{“presenter opens door 3”} \end{aligned}$$

Observe that the presenter will never open door 1 because that's the door you chose. Let's first compute the probability that the car is behind door 1. In this case, the presenter can open door number 2 or number 3 with the same probability. Then,

$$P(C_1 \text{ and } P_2) = P(C_1) \cdot P(P_2|C_1) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

$$P(C_1 \text{ and } P_3) = P(C_1) \cdot P(P_3|C_1) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

Then,

$$\begin{aligned} P(C_1) &= P\big([C_1 \text{ and } P_2] \text{ or } [C_1 \text{ and } P_3]\big) \\ &= P(C_1 \text{ and } P_2) + P(C_1 \text{ and } P_3) \\ &= \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \end{aligned}$$

Now let's compute the probability that the car is in door 2. In this case, the presenter **must** open door 3 because you chose door 1, and he won't reveal the door where the car is. Then,

$$P(C_2) = P(C_2 \text{ and } P_3) = P(C_2)P(P_3|C_2) = \frac{1}{3} \cdot 1 = \frac{1}{3}$$

Similarly,

$$P(C_3) = P(C_3 \text{ and } P_2) = P(C_3)P(P_2|C_3) = \frac{1}{3} \cdot 1 = \frac{1}{3}$$

If you decide to stay with your original choice (door 1), then the probability that you win is:

$$P(\text{win by staying}) = P(C_1) = \frac{1}{3}$$

Now, the probability of winning by changing doors uses the strategy of the presenter, that is, the presenter will never open the door with the car or the door that you chose. Then,

$$\begin{aligned} P(\text{win by changing}) &= P(\text{win and car in door 2}) + P(\text{win and car in door 3}) \\ &= P(C_2 \text{ and } P_3) + P(C_3 \text{ and } P_2) \\ &= P(C_2) \cdot P(C_2|P_3) + P(C_3) \cdot P(P_2|C_3) \\ &= \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 1 \\ &= \frac{2}{3} \end{aligned}$$

□