

## 3.1: Introduction to Probability

### Learning Objectives

- Learn key initial terminology about probability
- Determine the sample space of a given situation
- Recognize and use the three basic methods of determining probability measures
- Explain the importance of the *Law of Large Numbers*
- Describe the complement of an event
- Use the Complement Probability Rule for determining the probability of an event

### Review and Preview

Inferential statistics seeks to make educated guesses about populations using statistics from randomly chosen samples. The usefulness of a sample statistic depends on the sample from which it was taken. We cannot guarantee that samples are representative of the population, but we can ensure that any bias present is due to random chance. What is the likelihood that our randomly chosen sample is representative? In other words, what is the probability that our sample statistic is close to the population parameter? These are fundamental questions to examine through the science of probability.

We hear probability claims daily. The weather forecast states a 30% chance of rain. The probability of a faulty product coming off a manufacturing line is 6.5%. One has a 20% probability of purely guessing an answer correctly on a single multiple-choice question. A cancer research group believes that 40% of women and 45% of men will have a diagnosis of some type of cancer during their lifetimes. (Note: This means that if we randomly select a man, we have a 45% chance of choosing someone who has had or will have a cancer diagnosis in his life. This does not mean that any specific man has a 45% chance of being diagnosed with cancer in his lifetime.)

Each of the above scenarios involves a situation in which something will happen, and an outcome will occur, but we are uncertain which outcome it will be. Will it rain, or will it not rain? Is the next product produced of high quality or not? In statistics, we refer to such situations as random experiments. We have a clear context and an idea of possible outcomes, one of which will happen, and then use probabilities to measure the likelihood of any particular outcome.

As we move through the course, we will develop an understanding of random sampling to build probabilities. We will develop the ability to measure how unusual a sample statistic is within a given situation. If the probability of a calculated sample statistic is "small," then we conclude that the outcome is unusual. The use of probability goes well beyond this application; therefore, we develop probability within a larger context.

### Basic Concept of Probability

We begin by examining the meaning of the term probability. Generally, **probability** is a numerical value **inclusively** between 0 and 1 measuring the likelihood that a specific event will occur within a given situation. The representation of that numerical value might be in decimal form, fractional form, or percentage form, such as  $0.75 = \frac{3}{4} = 75\%$ . An **outcome** of a random experiment is a potential result of the experiment. The term **event** is a set of outcomes one might expect from a given random experiment. For example, if one rolls a pair of standard game dice as shown in Figure 3.1.1, a possible event could be both dice landing with one facing up.



Figure 3.1.1: Game Dice

Before measuring probability accurately, we must clearly describe the given situation (such as rolling a pair of standard game dice) and the event of interest (such as both dice landing with one facing up). For a given clearly described situation, the collection of all

possible outcomes is called the **sample space** or **event space**. For reasonably simple situations, one can fully describe the sample space.

### ? Text Exercise 3.1.1

What is the sample space if one rolls a pair of standard game dice, as shown in Figure 3.1.1?

#### Answer

The graphic below shows the sample space, where one die outcome is red and the other in white. Notice there are thirty-six possible outcomes. We should also notice that, in general, we must be concerned with the order of the two dice. For example, we consider rolling a three on the red die, and a one on the white die a different outcome than rolling a one on the red die and a three on the white die. Why is this? Because these are two outcomes that we can distinguish. Since probability measures the degree of uncertainty, one must consider all available information. We produce a complete sample space by carefully reflecting on all possible outcomes.

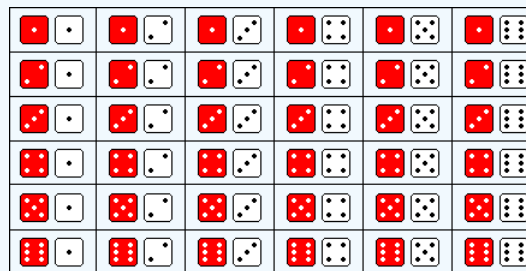


Figure 3.1.2 Sample space of rolling a pair of standard game dice

When rolling a pair of dice, the event of rolling two ones is an outcome. However, the event in which both dice land with a sum of five describes several outcomes in the sample space: 1 and 4; 2 and 3; 3 and 2; and 4 and 1.

Once we understand the sample space, we can examine the probability measures of various events within that situation. A probability measure near 1 indicates that the specific event is more likely to occur, and a probability measure near 0 indicates that the specific event is less likely to occur. We use the symbols  $P(A)$  to indicate the probability of a specific event  $A$ . For example,  $P(\text{rain}) = .30 = 30\%$  indicates "the probability of rain is 30%". If for some event  $A$ , we find that  $P(A) = 1.00 = 100\%$ , then we say this event is a **certain event**. Similarly, if for some event  $A$ , we find that  $P(A) = 0.00 = 0\%$ , then we say this event is an **impossible event**. For finite sample spaces, impossible events cannot occur, and certain events must occur. For infinite sample spaces, the situation is more complicated; we will discuss this further in chapter 4. For now, we are usually interested in events that are possible but not likely to occur.

#### 📌 Note: Unusual Events

In inferential statistics, we are often interested in outcomes that are not likely to happen; that is, they are not very probable; they are unusual. When improbable outcomes occur, either something rare happens, or we have reason to think our understanding of the situation needs to be updated, but what is the cutoff between improbable and probable, between unusual and usual? There is no fixed, agreed-upon value that will work for every situation. For this reason, we will often need to set a probability measure for considering an event improbable. Different individuals can generally feel the same event usual/likely or unusual/unlikely, depending on their personal life experiences. For example, if the probability of rain is 15%, one person may consider rain unlikely, and another may think it reasonably likely. A commonly chosen boundary measure for many statistical analysis areas is 5%. We will initially require a probability measure of 5% or less to label an event as unusual. Later, we will relax on this requirement, even leaving the choice to the consumer of our statistical measures to apply their chosen probability measure for unusual.

### ? Text Exercise 3.1.2

1. Explain why 150% cannot be the probability of some event.

#### Answer

The value 150% is larger than 100%. Any valid probability value will be between  $0.00 = 0\%$  and  $1.00 = 100\%$ , inclusively. Knowing such will help us recognize improper probabilities.

2. Explain why 1.21 cannot be the probability of some event.

#### Answer

The value 1.21 is larger than 1.00. Again, any valid probability value will be between  $0.00 = 0\%$  and  $1.00 = 100\%$ , inclusively.

3. Explain why  $-0.22$  cannot be the probability of some event.

#### Answer

The value  $-0.22$  is negative. Any valid probability value will be non-negative since the value must be between  $0.00 = 0\%$  and  $1.00 = 100\%$ .

4. Give an example of a situation and an impossible event for that situation.

#### Answer

Answers to this can vary greatly. An example tied to the situation about rolling pairs of dice would be the event of rolling two standard game dice with a sum of 13 up. Since the largest each die can be is 6, the sum cannot exceed 12; the event is an impossible event. That is,  $P(\text{rolling a sum of 13}) = 0.00\%$ .

There is one final fundamental property of probability for the events and sample space within any given situation. The **sum of the probabilities** of all possible outcomes within the sample space of a given situation must always total  $1.00 = 100\%$ .

## Basic Methods for Computing Simple Probabilities

With our basic terminology established, we focus on how to compute a given situation's probability. We initially examine three basic and commonly used methods. Our first method is called the **subjective** or **intuitive method**, where we produce a numerical estimate of the probability based on personal judgment, past experiences, or even personal opinion. For example, having purchased a few non-winning lottery tickets in the past, yet hearing of a winner on the news, we might estimate the probability of winning as a low value of 1%. We do not depend on subjectively determined probability values in quality statistical work. Instead, we form more robust and accurate methods for determining these values.

This leads us to our second method, commonly called the **classical method**. In this method, if a given situation has  $n$  different equally likely outcomes in the sample space and if event  $A$  can occur in  $x$  ways, then

$$P(A) = \frac{\text{number of ways event } A \text{ can occur}}{\text{number of outcomes in the sample space}} = \frac{x}{n}.$$

As an example, if a standard die is fair (so each face of the die has equal chance of landing up when the die is rolled), then

$$P(\text{ROLLING A THREE}) = \frac{\text{number of ways rolling a three can occur}}{\text{number of outcomes in the sample space of rolling a die}} = \frac{1}{6} \approx 0.1667 = 16.67\%$$

Notice that this outcome would not be considered unusual since the probability measure is over 5%.

### ? Text Exercise 3.1.3

Consider the situation in which a pair of fair dice are rolled. Find the probability of each event given. Write results in probability notation and determine if the outcome is considered unusual.

1. ROLLING A THREE ON THE FIRST DIE AND A FOUR ON THE SECOND DIE.

Answer


Figure 3.1.3 Sample space of rolling a pair of standard game dice

Since the dice are said to be fair dice, each of the thirty-six outcomes shown in our sample space above is equally likely. We will use the classical approach to determining the probability, using  $A$  to represent our event of ROLLING A THREE ON THE FIRST DIE AND A FOUR ON THE SECOND DIE. We notice only one outcome in the sample space matches the event description. Therefore,  $P(A) = \frac{1}{36} \approx 0.0278 = 2.78\%$ . Since this probability measure is less than 5%, then for our course, we will consider this outcome as unusual.

2. ROLLING A THREE ON ONE DIE AND A FOUR ON THE OTHER DIE.

Answer


Figure 3.1.4 Sample space of rolling a pair of standard game dice

Using similar reasoning as above and using  $B$  to represent our event of ROLLING A THREE ON ONE DIE AND A FOUR ON THE OTHER DIE, notice that there are two outcomes in the sample space that match the event description. Therefore,  $P(B) = \frac{2}{36} = \frac{1}{18} \approx 0.0556 = 5.56\%$ . Since this probability measure is more than 5%, then for our course, we will not consider this outcome as unusual.

3. ROLLING TWO DICE IN WHICH THE SUM OF THE NUMBER IS SIX.

Answer


Figure 3.1.5 Sample space of rolling a pair of standard game dice

Let  $C$  represent our event of ROLLING TWO DICE IN WHICH THE SUM OF THE NUMBER IS SIX, we notice that there are five outcomes in the sample space that match the event description...can you find all of these? Therefore,  $P(C) = \frac{5}{36} \approx 0.1389 = 13.89\%$ . Since this probability measure is more than 5%, then for our course, we will not consider this outcome as unusual.

Our third method of computing probabilities is the **empirical, experimental, or relative frequency method**. In this method, we repeatedly conduct an experiment, noting the outcomes of the trials, to establish an estimate of probability measures. In repeating a given experiment  $n$  times, and noting event  $A$  occurred  $f$  times, then

$$P(A) = \frac{\text{number of times event } A \text{ occurred}}{\text{number of times situation/experiment was repeated}} = \frac{f}{n}.$$

Notice how this method relates to our previous work producing [relative frequency distributions](#) when summarizing data sets. In a sense, we were producing probability measures with our relative frequency values.

The quality of the probability estimate is dependent on the number of repeated trials used. For example, a researcher finds that 25 of 150 randomly selected Kansas teens texted while driving during the last week, empirically indicating that the proportion of Kansas teens that drove while texting last week is  $\frac{25}{150} \approx 16.67\%$ . Equivalently, there is approximately a 16.67% probability that a randomly selected Kansas teen drove while texting last week. This measure based on only 150 teens does not give us high confidence in this estimated probability value; if the researcher could collect data from 1,500 Kansas teens, we could be more confident in the estimation.

In general, we require the number of carefully designed repeated trials to be as large as possible to produce an estimate of a probability value. Two underlying assumptions must be used with the experimental frequency approach. First, if an event occurred with a certain probability in past trials, this same event will occur about the same percentage of times in future trials. Second, the relative frequency probability of an event will tend to approach the true probability value as more and more trials are measured (this is commonly referred to as the **Law of Large Numbers**).

There are times when simulations (especially computer simulations) produce a large number of trials and a reasonably accurate measure of the true probability of specific events, especially when the outcomes in a situation are not equally likely. Much weather forecasting is based on computer simulation of outcomes based on regional weather conditions. Similarly, the spread of infectious diseases is modeled by computer simulations to predict outcomes while avoiding extensive medical research costs or without impacting living organisms. For example, suppose we wonder if our game die is fair--that each face has an equal likelihood of occurring when the die is rolled. We begin a simulation with the assumption that the die is fair, but upon rolling the die five times, we roll a one three of the five times--indicating an empirical probability of  $P(\text{ROLL OF ONE}) = \frac{3}{5} = 60\%$ . Of course, only rolling five times is insufficient for us to have high confidence in the accuracy of the empirical probability measure. We continue rolling the die three hundred times, in which we roll a one 188 times. This new empirical probability measure of  $P(\text{ROLL OF ONE}) = \frac{188}{300} \approx 62.67\%$  indicates that our die is not likely to be fair; the experimental probability for the one is not the expected fair value of  $P(\text{ROLL OF ONE}) = \frac{1}{6} \approx 16.67\%$ .

#### ? Text Exercise 3.1.4

In 1856, [Gregor Mendel](#) began to study different inherited features, such as the color of pea plants. According to one [source](#), in a second-generation group of pea plants, 6002 peas produced by the plants were yellow, and 2001 were green in color. What was the empirical probability that a randomly selected second-generation pea would be green? Is this close to the hypothesized value that Mendel claimed of 25%?

#### Answer

We compute that  $P(\text{GREEN COLORED PEA}) = \frac{2001}{6002+2001} \approx 25.0031\%$ . Mendel's experimental probability is extremely close to his hypothesized probability claim of 25%.

### Complement and Complement Probabilities

There are times we will be interested in finding the probability that an event  $A$  does not occur. The collection of all outcomes in a sample space in which a given event  $A$  does not occur is called the **complement** of event  $A$  and is denoted by  $\bar{A}$ . Other sources may use different notations to denote the complement; common ones include  $A^c$  or  $\sim A$ . The idea of complementary events allows us to divide the sample space into two mutually exclusive groups (no outcome can be found in both of the two groups  $A$  and  $\bar{A}$ ) and also exhaustive (every outcome of the sample space must be included in one of our two groups). For example, suppose  $A$  is the event of getting two different numbers on each die when two dice are rolled; there are 30 such outcomes in our sample space that

meet this event description.  $\bar{A}$  is the event of getting the same numbers on each die when two dice are rolled; there are 6 such outcomes in our sample space. Between  $A$  and  $\bar{A}$ , all 36 outcomes appear exactly once.

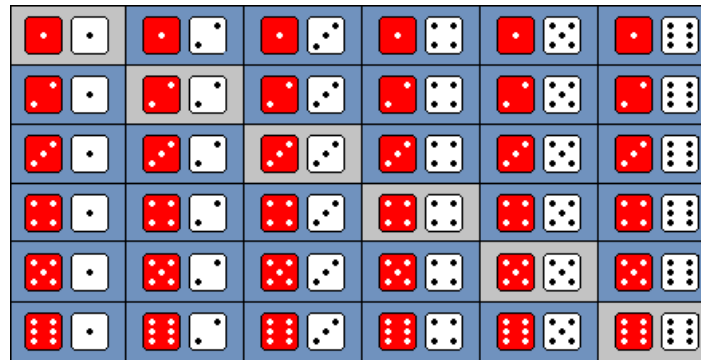


Figure 3.1.2: Events  $A$  (blue background) and  $\bar{A}$  (grey background)

Since all outcomes of a situation must be either in the event or the complement of the event, we have the following three key consequences:

1.  $P(A) + P(\bar{A}) = 1 = 100\%$
2.  $1 - P(A) = P(\bar{A})$
3.  $1 - P(\bar{A}) = P(A)$

Notice that the first consequence leads naturally to the other two with basic algebra. Suppose we wish to determine the probability of event  $A$  from above ( $P(\text{GETTING TWO DIFFERENT NUMBERS OF PIPS ON EACH DIE WHEN TWO DICE ARE ROLLED})$ ). We can go back to our sample space to count all such outcomes or use the complement concept to produce our results quickly. We have

$$\begin{aligned} P(A) &= 1 - P(\bar{A}) \\ &= 1 - \frac{6}{36} \\ &= \frac{30}{36} = \frac{5}{6} \approx 0.8333 \text{ or } 83.33\% \end{aligned}$$

This example illustrates that it is sometimes easier to determine the probability of a complement event instead of the given event. Once we know the probability of the complement event, we can easily determine the probability of the event.

### ? Text Exercise 3.1.5

Use the concepts of complement events to answer the questions below:

1. Suppose there are sixty tiles in a bag of which 13 are green, 6 are yellow, 4 are pink, 9 are red, 8 are purple, and 20 are black. The tiles are well-mixed. We will randomly draw one tile from the bag without looking into the bag. We want to determine the probability that we draw a tile that is not a **primary color**; that is, we are to find  $P(\text{NOT A PRIMARY COLORED TILE})$ .

#### Answer

As a reminder, there are three primary colors: red, yellow, and blue. Although the complement is not necessary to answer this question, the use of the complement makes the problem easier to compute. Since "a primary colored tile" is the complement event of "not a primary colored tile" in this situation, we notice that  $P(\text{NOT A PRIMARY COLORED TILE}) = 1 - P(\text{A PRIMARY COLORED TILE}) = 1 - \frac{6+9}{60} = \frac{45}{60} = 75\%$ .

2. A number is chosen randomly from the set of integers between 1 and 99, inclusively. What is the probability of randomly selecting a number that is not a perfect square?

#### Answer

We notice that there are only a few integers inclusively between 1 and 99 that are perfect squares and many that are not. Specifically the perfect square integers in this set are  $\{1, 4, 9, 16, 25, 36, 49, 64, 81\}$ . Thus  $P(\text{NOT A PERFECT SQUARE}) = 1 - P(\text{PERFECT SQUARE}) = 1 - \frac{9}{99} \approx 0.909 = 90.9\%$ .

3. Suppose we want to know the probability of randomly selecting a group of 35 people in which at least two people will have the same birth day in a year (such as September 1<sup>st</sup> or May 28<sup>th</sup> -- we will ignore leap years for simplicity). Suppose we also know the probability of randomly selecting a group of 35 people, in which no two people have the same birth day in the year, which is about 18.56%. (We will discuss how this value of 18.56% can be found later in the chapter). From this information, can we determine the probability of randomly selecting a group of 35 people in which at least two will have the same birth day in a year?

#### Answer

We see that the complement of "at least two people will have the same birth day in a year" is the description "less than two people (that is none) will have the same birth day in a year." We can use our complement rule to note that  $P(\text{AT LEAST TWO WILL HAVE THE SAME BIRTHDAY}) = 1 - P(\text{NO TWO WILL HAVE THE SAME BIRTHDAY}) = 1 - 0.1856 = 0.8144 = 81.44\%$

There are times when clearly describing the complement of an event can be simple. For example, in the roll of a single standard game die, it is common to quickly describe the complement of AN EVEN NUMBER ON A ROLL to be AN ODD NUMBER ON A ROLL. There are also times when clearly describing the complement of an event can be challenging. We must ensure the complement description covers all possible outcomes for a situation. For a different example, in random selection of a number from the real number line, it is common to quickly describe the complement of SELECTION OF A NEGATIVE NUMBER as SELECTION OF A POSITIVE NUMBER. However, there is the number, 0, that is neither positive nor negative which means we have an incorrect complement description. The true complement to SELECTION OF A NEGATIVE NUMBER is SELECTION OF A NON-NEGATIVE NUMBER. Now all possible outcomes have been accounted for in the descriptions.

We must think carefully about our sample space and correctly identify all outcomes which are not in our event. As another example, we might have the situation of randomly selecting an FHSU student with a focus on the outcome of STUDENT HAS HAD AT LEAST THREE COVID VACCINE SHOTS. The complement is the outcome of STUDENT HAS HAD TWO OR FEWER COVID VACCINE SHOTS or, equivalently, STUDENT HAS HAD FEWER THAN THREE COVID VACCINE SHOTS. Notice how the use of mathematical notation can help us here. We might represent the outcome of STUDENT HAS HAD AT LEAST THREE COVID VACCINE SHOTS more briefly in notation as  $x \geq 3$ . Then, in the context of counting number of shots, the complement is  $x < 3$ , which leads to a proper complement description of STUDENT HAS HAD FEWER THAN THREE COVID VACCINE SHOTS.

This leads to special cases that commonly cause problems in complement descriptions; specifically, events involving "at least," "at most," "all" or "none." The complement of "all are" is not "none are," but is instead "at least one is not." For example, the complement of STUDENT HAS HAD ALL AVAILABLE COVID VACCINE SHOTS would be STUDENT HAS MISSED AT LEAST ONE COVID VACCINE SHOT. The complement of STUDENT HAS HAD NONE OF THE AVAILABLE COVID VACCINE SHOTS would be STUDENT HAS HAD AT LEAST ONE OF THE AVAILABLE COVID VACCINE SHOTS. Note that the complement of "none are" is "at least one is." We must think carefully when dealing with complements of event claims involving such keywords.

#### ? Text Exercise 3.1.6

Give written descriptions of the complements of each event described below

1. The event: IT SNOWS ON CHRISTMAS DAY

#### Answer

The complement event description would be IT DOES NOT SNOW ON CHRISTMAS DAY. We note that IT RAINS ON CHRISTMAS DAY is an outcome that fits in the complement description, but is itself not the actual full complement since other outcomes exist in the complement besides rain.

2. The event: NATASHA IS LESS THAN 10 MINUTES LATE

### Answer

Notice the given event description of less than 10 minutes can be represented mathematically as  $x < 10$ , informing us that the complement must be related to  $x \geq 10$ . Therefore, the complement event description would be NATASHA IS AT LEAST 10 MINUTES LATE, or equivalently, NATASHA IS LATE BY AT LEAST 10 MINUTES. The phrasing we might give can vary, but the meaning of the phrasing must be tied to the inequality  $x \geq 10$ .

3. The event: ALL CARDS IN A POKER HAND ARE FACE CARDS ([Click here](#) for a full description and visualization of a standard deck of playing cards)

### Answer

The given event description of "all cards" is complemented by "at least one is not." So the complement event description would be AT LEAST ONE CARD IS NOT A FACE CARD. We must think clearly about this complement and also note why the description NONE OF THE CARDS ARE FACE CARDS is not the complement description. We could have one, two, three, or four cards that are not face cards as possible events in the complement.

4. The event: NONE OF THE STUDENTS FAILED THE LAST EXAM

### Answer

The given event description of "none failed" is complemented by "at least one did fail." So the complement event description would be AT LEAST ONE STUDENT FAILED THE LAST EXAM.

## Summary

To review this section, we list several important facts to remember when working with probabilities:

1. An outcome of a random experiment is any potential result of the experiment. An event is a set of outcomes one might expect from a given random experiment. The sample space is the collection of all possible outcomes in a given situation.
2. The probability of an event  $A$  is denoted by  $P(A)$  with the condition that  $0 \leq P(A) \leq 1$ . If  $P(A) \leq 5\%$ , we will currently consider the event as unusual.
3. The sum of the probabilities for all the possible outcomes in the sample space will always total to  $1 = 100\%$ .
4. Multiple methods may be used for computing probability values. We have discussed the methods of "subjective/intuition," "classical for equally likely simple events," and "empirical/experimental/relative frequency." Some methods only produce estimations of the actual probability value.
5. The complement of an event  $A$  is denoted by  $\bar{A}$  and must contain all outcomes of the sample space that are not part of the given event. As such, we have the probability benefit of  $P(A) + P(\bar{A}) = 1$  that can be used to find the probability of an event if we know the probability of the complement.

3.1: Introduction to Probability is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- [3.3: Measures of Central Tendency](#) by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.
- [2.1: Graphing Qualitative Variables](#) by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.