

[See this page in the course material.](#)

Learning OUTCOMES

- Interpret (in context) a probability as a long-run relative frequency of an event.

What Is the Relationship between Theoretical and Empirical Probability?

We investigate this question in the following two activities. We use coin flipping as a first step in understanding the connection between these two ways of determining the probability of an event.

A single flip of a coin has an uncertain outcome. We do not know if we will get heads or tails. If we flip the coin 10 times, we are not guaranteed to get 5 heads and 5 tails. So what exactly does it mean when we say $P(\text{heads}) = 0.5$? To answer this question, we use a simulation to simulate flipping a coin.

Our goal is to understand how the empirical probability $P(\text{head})$ relates to the theoretical probability of 0.5.

Activity 1: Fair Coin

The purpose of this activity is to experiment with a simulation that simulates flipping a **fair** coin, and to see if the $P(H) = 0.5$.



[See this interactive in the course material.](#)

Source: [GeoGebra](#), license: [CC BY SA](#)

Note that by clicking the GeoGebra link above you can launch a new window with this simulation in it if you would like to position it closer to the questions you'll be answering below to avoid scrolling so much.

Part (1)

1. Make sure **Coins = 1** and **P(heads) = 0.5**.
2. Press the “**1 Flip**” button 3 times.
3. Notice that for each flip, you will see either heads (1) or tails (0) appear in the histogram count.

Part (2)

1. Press the **Reset** button so that the count is cleared.
2. Make sure **Coins = 1** and **P(heads) = 0.5**.
3. This time press the “**10 Flips**” button 3 times so that you have 30 coin flips.

Try It

Now that you have flipped the coin 30 times, what did you get for $P(H)$?

To calculate $P(H)$ take the total number of heads (the count of the “1’s”) and divide by the total number of flips (30).

Write your essay resp

Check Answer

[See this interactive in the course material.](#)

Part (3)

1. Press the **Reset** button so that the count is cleared.
2. Make sure **Coins = 1** and **P(heads) = 0.5**.
3. Press the **Auto** button and watch the count of heads and tails change.
4. Click the **Pause (II)** button once **Total Flips** is over 1,000.

Try It

Describe what happened to the histogram as you flipped at least 1,000 times.

Write your essay resp

Check Answer

[See this interactive in the course material.](#)

Try It

What is the point of the previous activity?

☐ There is no connection between empirical and theoretical probability.

[See this interactive in the course material.](#)

In the preceding activity, the simulation simulates flipping a fair coin. $P(\text{heads}) = 0.5$ with a fair coin. How can we tell if a coin is not fair? Theoretical probability methods cannot answer this question. The only way we can answer this question is to collect data as we flip the coin.

Activity 2: Unfair Coin

The purpose of this activity is to experiment with an activity that simulates flipping an unfair coin.



[See this interactive in the course material.](#)

Source: [GeoGebra](#), license: [CC BY SA](#)

Note that by clicking the GeoGebra link above you can launch a new window with this simulation in it if you would like to position it closer to the questions you'll be answering below to avoid scrolling so much.

1. Make sure **Coins = 1** and **P(heads) = 0.2**.
2. Click the **Auto** button and watch the count of heads and tails change.
3. Click the **Pause (II)** button once **Total Flips** is over 100 or so.
4. Record the total number of Heads (1's) and the total number of flips.
5. Calculate $P(H)$ (Number of heads / Total Flips) when Total Flips is about 100.
6. Click the **Auto** button again to continue the flips.
7. Click the **Pause (II)** once **Total Flips** is over 1,000 or so.

8. Record the total number of Heads (1's) and the total number of flips.
9. Calculate $P(H)$ (Number of heads / Total Flips) when Total Flips is about 1,000.

Try It

What happened to $P(H)$ as the number of flips increased from 100 to 1,000?

Write your essay response

Check Answer

[See this interactive in the course material.](#)

Try It

$P(H)$ is [Select] , therefore the coin is [Select] .

Check Answer

[See this interactive in the course material.](#)

Let's summarize what we have learned from these activities:

- The empirical probability will approach the theoretical probability after a large number of

repetitions. In some situations, such as in flipping an unfair coin, we cannot calculate the theoretical probability. In these cases, we have to depend on data.

- There is less variability in a large number of repetitions. This means that in the long run, we will see a pattern, so we are more confident about estimating the probability of an event using empirical probability with a large number of repetitions.

What Do We Mean When We Say an Event Is Random or Due to Chance?

In the discussion of the role of probability in the Big Picture of Statistics, we said that probability is the machinery that allows us to draw conclusions about a population on the basis of a random sample. To understand why we can trust random selection in an observational study and random assignment in an experiment, we need to look more closely at what we mean by random or chance behavior.

When we say that an event is random or due to chance, we mean that the event is unpredictable in the short run but has a regular and predictable behavior in the long run. This is obviously true for the coin-tossing activity. We cannot predict whether an individual toss will be heads, but in the long run, the outcomes have a predictable pattern. The relative frequency of heads is very close to 0.5 for a fair coin.

We can make probability statements only about random events.

What Is the Connection between the Coin-Flipping Activities and the Discussion of Probability in the Previous Module?

Let's look at two probability questions that we might answer using the familiar data set from *Relationships in Categorical Data with Intro to Probability*. Recall that 6,198 of the 12,000 students at a West Coast community college are female. Previously, we calculated $P(\text{female}) = 6,198 / 12,000 = 0.5165$. What is the random event in this case? Let's be very specific about the question this calculation is meant to answer.

What is the probability that a student at the West Coast community college is a female?

- In this case, the relative frequency $6,198 / 12,000$ is the actual proportion of females at the college. This is like the fair coin situation. Because we know the gender distribution at the college, we can think of 0.5165 as the theoretical probability that a *randomly selected student at this particular college* is a female. Tossing the fair coin in the simulation is like randomly selecting a student from the spreadsheet of data. We do not know if a randomly selected student will be female. But if we repeat this process many, many times, in the

long run, the relative frequency of females will have a predictable pattern. The relative frequency will be very close to the proportion of females in the data set.

What is the probability that a community college student in the United States is female?

- In this case, we are using the data from the 12,000 West Coast community college students to represent students at all community colleges in the United States. The relative frequency is an estimate for the chance that a *randomly selected U.S. student* is female. This is like tossing the unfair coin 12,000 times and using the relative frequency of heads as an estimate of $P(\text{head})$. We do not know $P(\text{female})$ for all community colleges, just as we did not know the $P(\text{heads})$ with an unfair coin. But if the sample is random, we can use the relative frequency of females in the sample as an estimate of $P(\text{female})$ in all community colleges.

The main points are these:

- We can make probability statements only about random events.
- Probability of an event A is the relative frequency with which that event occurs in a long series of repetitions.

Licenses and Attributions

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** [CC BY: Attribution](#)

</div