Introduction to probability

Michelle Arnetta

Summary
Probability theory is an important branch of mathematics, which is foundational to statistics. This guide covers sample space, independent events, dependent events, and tree diagrams.

What is probability?

Uncertain situations are everywhere: from rolling a dice and flipping a coin, to weather forecasting and financial markets. How exactly can you approach decision making in the face of uncertainty? For example, should you bring an umbrella tomorrow? Or, should you invest in a certain stock?

When you are facing these situations, you could turn to **probability theory** to predict the likelihood of certain outcomes, especially when you have to make important decisions.

This guide introduces you to probability. First, a distinction will be made between theoretical probability and experimental probability, and the concept of sample space will be explained. Then, you will learn about independent and dependent events.

Theoretical Probability vs. Experimental Probability

There are two main types of probability: theoretical probability and experimental probability.

> **i** Definition of theoretical probability
>
> **Theoretical probability** is probability based on what you expect to happen.

Here is how you can find theoretical probability:

$$Theoretical\ Probability = \frac{number\ of\ desired\ outcomes}{total\ number\ of\ possible\ outcomes}$$

> **i** Definition of experimental probability
>
> **Experimental probability** is probability based on what actually happens.

Here is how you can find experimental probability:

$$Experimental\ Probability = \frac{number\ of\ times\ that\ a\ desired\ event\ occurs}{total\ number\ of\ trials\ in\ an\ experiment}$$

You can try this out for yourself!

Example 1

Try to "flip this coin".

SLIDES ONE

Example 2

Try to "roll a six-sided die".

SLIDES TWO

Tip

There are three common ways of representing probability:

- As a fraction, such as $\frac{1}{2}, \frac{1}{5}, \frac{3}{4}$
- As a decimal, such as 0.5, 0.2, 0.75
- As a percentage, such as 50%, 20%, 75%

Tip

The more trials you do, the closer your experimental probability will get to theoretical probability.

POSSIBLE FIGURE: graph showing the more coins you flip, the closer it gets to 50/50 for heads or tails

Tip

Did you realize that the probabilities of all outcomes in an event add up to 1? When flipping a coin, the probability of getting heads is $\frac{1}{2}$, and the probability of getting tails is $\frac{1}{2}$. Add these up together to get:

$$\frac{1}{2} + \frac{1}{2} = 1$$

Notice that, in this case, the probability of 'heads' is equal to the probability of 'tails' **not** occurring, and the probability of 'tails' is equal to the probability of 'heads' **not** occurring. Therefore, 'heads' and 'tails' can be called **complementary events**. You could then, for example, subtract the probability of 'heads' to get the probability of 'tails'.

$$1 - \frac{1}{2} = \frac{1}{2}$$

This is known as the **complement rule**.

Sample Space

When thinking about probability, it is important to consider what all the possible outcomes are. This is known as **sample space**.

Definition of sample space

Sample space is the set of all possible outcomes in an experiment.

The **sample space** can be represented in various different ways, but one common example is to represent them as a **list**. To clarify this, you can look back to the previous two examples and try representing their sample space as a list.

Example 3

There are two possible outcomes when you flip a coin: heads (H) or tails (T). Therefore, the sample space of flipping a coin can be represented as {H, T}.

Example 4

There are six possible outcomes when you roll a die: 1, 2, 3, 4, 5, and 6. Therefore, the sample space of rolling a die can be represented as {1, 2, 3, 4, 5, 6}.

But what if the coin was flipped, then the die was rolled? When you are representing the sample space or all possible outcomes of two events, it would be helpful to use a **table**. When you are representing the sample space of two or more events, it could also be helpful to use a **tree diagram**.

Example 5

You flip a coin and roll a die. There are twelve possible outcomes when you do this: heads and 1, heads and 2, heads and 3, heads and 4, heads and 5, heads and 6, tails and 1, tails and 2, tails and 3, tails and 4, tails and 5, tails and 6. When representing this as a **list**, you could write this as {H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6}.

Since you are working with two events, you could also represent this as a **table**:

	1	2	3	4	5	6
H	H1	H2	H3	H4	H5	H6
T	T1	T2	T3	T4	T5	T6

Alternatively, you could represent it as a **tree diagram** because you are working with two or more events:

POSSIBLE FIGURE: TREE DIAGRAM

Notice that since the coin flipping occurred first, the tree diagram begins with this event. It then branches out to the possible outcomes for when the die is rolled.

Outcomes that Vary in Probability

The examples you have examined so far involve events with equally probable outcomes. For instance, you have a 50/50 chance of getting either heads or tails after flipping a coin. Similarly, it is equally likely for you to get any number between 1 and 6 after rolling a die, a $\frac{1}{6}$ chance. But what if the outcomes varied in probability? It becomes all the more useful to represent these outcomes with tree diagrams.

For example, suppose that you have a bag containing 5 marbles in total, with 2 red marbles and 3 blue marbles. If you draw one marble from the bag, what is the probability of it being a red marble, and what is the probability of it being a blue one? This can be represented as a tree diagram:

POSSIBLE FIGURE: tree diagram 1 for candy jar

As shown in the diagram, the probability of the drawn marble being red is $\frac{2}{5}$, and the probability of it being blue is $\frac{3}{5}$.

You can also use a tree diagram to represent events with more than two outcomes. Imagine a jar containing 10 candies. It has 1 yellow candy, 4 green candies, and 5 purple candies. If you take one candy from the jar, what is the probability of each color being taken?

POSSIBLE FIGURE: tree diagram 2 for marble bag

Since $\frac{4}{10} = \frac{2}{5}$ and $\frac{5}{10} = \frac{1}{2}$, the diagram can be simplified as follows:

POSSIBLE FIGURE: tree diagram 1 for candy jar (simplified)

As shown in the diagram, the probability of the taken candy being yellow is $\frac{1}{10}$, the probability of it being green is $\frac{2}{5}$, and the probability of it being purple is $\frac{1}{2}$.

Example 6

The **complement rule** also applies here. Consider the candy jar example. What if you knew the probability of taking a yellow or purple candy, but not the probability of taking a green candy?

POSSIBLE FIGURE: tree diagram with ? for candy jar

To apply the complement rule, you should first find the probability of **not** taking the green candy. You can do this by adding together the probabilities of taking a yellow or purple candy.

$$\frac{1}{10} + \frac{1}{2} = \frac{1}{10} + \frac{5}{10} = \frac{6}{10}$$

The probability of taking the green candy is **complementary** to the probability of not taking the green candy. From this, you can calculate the mystery probability:

$$1 - \frac{6}{10} = \frac{4}{10} = \frac{2}{5}$$

Independent Events

When you are representing the probabilities of two or more events, it can be important to consider the relationships between the events. Events can either be **dependent** on or **independent** of each other.

Definition of independent events

Two events are **independent** if the occurrence of one event does not affect the likelihood of the other event happening.

To illustrate this, you can return to the marble bag example. What if two events occurred?

- Event 1: One marble is drawn, and the color of the marble is recorded.
- Event 2: After the marble from event 1 is put back into the bag, one marble is drawn, and the color of the marble is recorded.

These events can be outlined by a tree diagram:

POSSIBLE FIGURE: tree diagram 2 for marble bag

No matter what the color of the marble from event 1 is, the probabilities of each marble color will remain the same for event 2. This is because the first marble was drawn and replaced by the same color. Hence, this is an example of **probability with replacement**. Considering this, you can conclude that 1 and 2 are **independent events**, as the outcome of event 1 does not influence the outcome of event 2.

You can also use the tree diagram as a guide to calculating the probabilities of two particular events occurring:

- Probability of drawing a red marble, then a red marble: $(\frac{2}{5})(\frac{2}{5}) = \frac{4}{25}$
- Probability of drawing a red marble, then a blue marble: $(\frac{2}{5})(\frac{3}{5}) = \frac{6}{25}$
- Probability of drawing a blue marble, then a red marble: $(\frac{3}{5})(\frac{2}{5}) = \frac{6}{25}$
- Probability of drawing a blue marble, then a blue marble: $(\frac{3}{5})(\frac{3}{5}) = \frac{9}{25}$

Dependent Events

Definition of dependent events

Two events are **dependent** when the occurrence of one event affects the likelihood of another event happening.

The previous marble bag example can be adjusted to demonstrate **dependent events**. What if the two events occurred in this way instead?

- Event 1: One marble is drawn, and the color of the marble is recorded.
- Event 2: The marble from event 1 is not replaced. One marble is drawn, and the color of the marble is recorded.

This is how the events would look on a tree diagram:

POSSIBLE FIGURE: tree diagram 3 for marble bag

You can observe that the **denominators** of the probabilities change from 5 to 4. This is because there were initially 5 marbles in the bag, and one marble was drawn without replacement, leaving 4 marbles in the bag for event 2.

The **numerators** only change for the marble color already drawn. If a red marble is drawn, then 1 red marble will be left among the 4 remaining marbles in the bag, so the probability of drawing a red marble in event 2 is $\frac{1}{4}$. On the other hand, if a blue marble is drawn, then 2 blue marbles will be left among the 4 remaining marbles in the bag, so the probability of drawing a blue marble in event 2 is $\frac{2}{4}$.

Therefore, the probabilities of outcomes in event 2 are **dependent** on the outcome of event 1.

Because $\frac{2}{4} = \frac{1}{2}$, the diagram can be simplified:

POSSIBLE FIGURE: tree diagram 3 for marble bag (simplified)

As with independent events, you can use the tree diagram as a guide to calculating the probabilities of two particular events occurring:

- Probability of drawing a red marble, then a red marble: $(\frac{2}{5})(\frac{1}{4}) = \frac{2}{20} = \frac{1}{10}$
- Probability of drawing a red marble, then a blue marble: $(\frac{2}{5})(\frac{3}{4}) = \frac{6}{20} = \frac{3}{10}$
- Probability of drawing a blue marble, then a red marble: $(\frac{3}{5})(\frac{1}{2}) = \frac{3}{10}$
- Probability of drawing a blue marble, then a blue marble: $(\frac{3}{5})(\frac{1}{2}) = \frac{3}{10}$

Quick check problems

1. Today, it can either rain or not rain. Suppose that the probability of it raining is 0.7. What is the probability of it not raining?
2. A researcher flips a coin 10 times, and it lands on heads 7 times. Therefore, the researcher concludes that the probability of 'heads' is $\frac{7}{10}$. What type of probability is this?
3. You are given three statements below. Decide whether they are true or false.

(a) The sum of the probabilities of complementary events is 1.

(b) Only tables can be used to represent the sample space of two events.

(c) Tree diagrams can be used to represent both dependent and independent events.

Further reading

For more questions on the subject, please go to Questions: Introduction to probability.

Version history

v1.0: initial version created