Introducing Probability

Definitions, axioms, and examples

Welcome to Unit II: Probability

In an enormously entertaining paper written about a decade ago, the economist Peter Backus estimated his chance of finding a girlfriend on any given night in London at about 1 in 285,000 or 0.0000034%. As he writes, this is either depressing or cheering news for a person, depending on what you had estimated your chance to be *before* reading the paper and doing a similar computation for yourself.[1] The interesting point in the paper was using a probabilistic argument (originally developed by the astronomer and astrophysicist Frank Drake to estimate the probability of extra-terrestrial civilizations) to think about his dating problems. Anyone can follow the arguments put forward by Backus, including his statements that use probability.

We all have some notion of chance or probability, and can ask questions like: - What is the chance you will get an A in Stat 20? (About 32%, based on last fall.)[2] - What is the chance the 49ers will win the Super Bowl this year? (They are the favorites, with an implied probability of about 54.5%, .)[3] - What is the chance you will roll a double on your next turn to get out of jail while playing Monopoly? (One in six.) - What is the chance that Donald Trump will win the Presidential election? (About 47%.) [4]

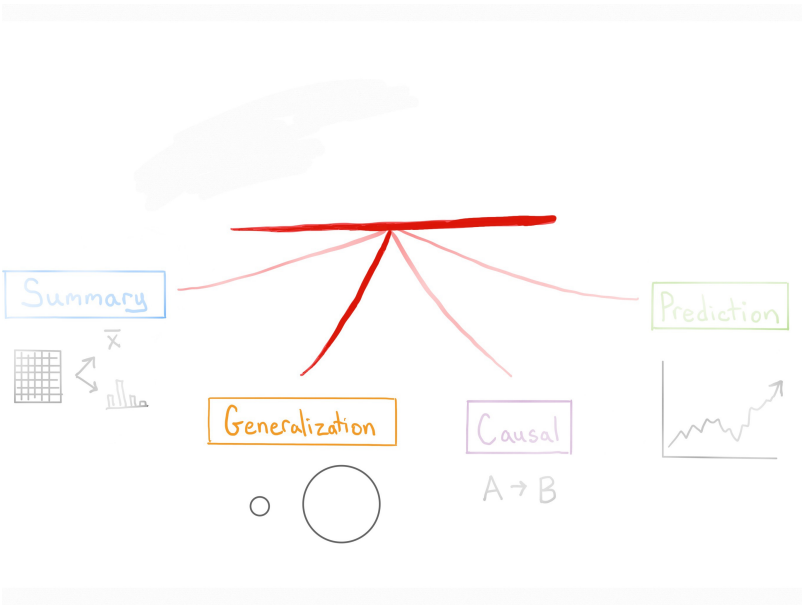
The second of our four types of claim we will investigate is a **generalization.** To do so, we first need to quantify uncertainty and randomness. This is the purpose of the **Probability** unit.

[1] Paper is at https://www.astro.sunysb.edu/fwalter/AST248/why_i_dont_have_a_girlfriend.pdf and a talk by Backus at https://www.youtube.com/watch?v=ClPPSry8bBw

[2] https://berkeleytime.com/grades/0-7077-all-all&1-7077-fall-2022-all

[3] https://www.freep.com/betting/sports/nfl-49ers-vs-chiefs-odds-moneylines-spreads-totals-best-nfl-odds-this-week

[4] https://www.thelines.com/odds/election/



So far, we have examined data sets and summarized them, both numerically and visually. We have looked at data distributions, and associations between variables. Can we extend the conclusions that we make about the data sets to larger populations? If we notice that bill length and flipper length have a strong linear relationship for the penguins in our data, can we say this is true about all penguins? How do we draw *valid* conclusions about the population our data was drawn from? These are the kinds of questions we we will study using tools from probability theory.

In order to be taken seriously, we need to be careful about how we collect data, and then how we generalize our findings. For example, you may have observed that some polling companies are more successful than others in their estimates and predictions, and consequently people pay more attention to them. Below is a snapshot of rankings of polling organizations from the well-known website FiveThirtyEight[5], and one can imagine that not many take heed of the polling done by the firms with C or worse grades. According to the website, the rankings are based on the polling organization's "historical accuracy and methodology".

[5]This website was begun as poll aggregation site, by the statistician Nate Silver.

Zogby Interactive/JZ Analytics	B-	+0.5	477
YouGov	B+	-0.3	455
Public Policy Polling	A-	-0.4	454
Mason-Dixon Polling & Strategy	A-	-0.4	445
American Research Group	C+	+0.7	277
SurveyMonkey	C	+1.0	268

In order to make estimates as these polling organizations are doing, or understand the results of a clinical trial, or other such questions in which we *generalize* from our data sample to a larger group, we have to understand the *variations* in data introduced by randomness in our sampling methods. Each time we poll a different group of voters, for example, we will get a different estimate of the proportion of voters that will vote for Joe Biden in the next election. To understand variation, we first have to understand how probability was used to collect the data.

Since classical probability came out of gambling games played with dice and coins, we can begin our study by thinking about those.

De Méré's Paradox

In 17th century France, gamblers would bet on anything. In particular, they would bet on a fair six-sided die landing 6 at least once in four rolls. Antoine Gombaud, aka the Chevalier de Méré, was a gambler who also considered himself something of a mathematician. He computed the chance of a getting at least one six in four rolls as $2/3$ ($4 \times (1/6) = 4/6$). He won quite often by betting on this event, and was convinced his computation was correct. Was it?

The next popular dice game was betting on at least one double six in twenty-four rolls of a pair of dice. De Méré knew that there were 36 possible outcomes when rolling a pair of dice, and therefore the chance of a double six was $1/36$. Using this he concluded that the chance of at least one double six in 24 rolls was the same as that of at least one six in four rolls, that is, $2/3$ ($24 \times 1/36$). He happily bet on this event (at least one double six in 24 rolls) but to his shock, lost more often than he won! What was going on?

We will see later how to compute this probability, but for now we can estimate the value by **simulating** the game many times (1000 times each) and looking at the proportion of times we see at least one six in 4 rolls of a fair die, and do the same with at least one double six in 24 rolls.

```
Number of simulations = 1000
```

```
  prop_wins_game_1
1            0.514
```

```
  prop_wins_game_2
1            0.487
```

You can see here that the poor Chevalier wasn't as good a mathematician as he imagined himself to be, and didn't compute the chances correctly. The simulated probabilities are nowhere close to 4/6 and 2/3, the probabilities that he computed for the first and second game, respectively.

By the end of this unit, you'll be able to conduct simulations like these yourself in R! For today, we are going to begin by introducing the conceptual building blocks behind probability.

Basics

First, let's establish some terminology:

Experiment An action, involving chance, that can result in a finite number of possible *outcomes* (results of the experiment). For example, a coin toss is an experiment, and the possible outcomes are the coin landing heads or tails.

Outcome space This is just a set. It is the collection of all the possible outcomes of an experiment is called an outcome space or *sample* space, and we denote it by the upper case Greek letter Ω ("Omega"). For example, if we toss a coin, then the corresponding outcome space is $\Omega = \{\text{Heads, Tails}\}$. If we roll a die, then the corresponding outcome space $\Omega = \{1, 2, 3, 4, 5, 6\}$. We will denote a set by enclosing the elements of the set in braces: $\{\}$.

Event A collection of outcomes as a result of the experiment being performed, perhaps more than once. For example, we could toss a coin twice, and consider the *event* of both tosses landing heads. We usually denote events by upper case letters from the beginning of the alphabet: $A, B, C, ...$ An event is a subset of the outcome space, and we denote this by writing $A \subset \Omega$.

P(A) For any event A, we write *the probability of* A as $P(A)$.

Equally likely outcomes When all the possible outcomes in a finite outcome space of size n happen with the same probability, which is $\frac{1}{n}$.

Let's say that there are n possible outcomes in the outcome space Ω, and an event A has k possible outcomes out of those n. If all the outcomes are equally likely to happen (as in a die roll or coin toss), then we say that the probability of A occuring is $\frac{k}{n}$.

$$P(A) = \frac{k}{n}$$

Example: Tossing a fair coin

Suppose we toss a fair coin, and I ask you what is the chance of the coin landing heads. Like most people, you reply 50%. Why? Well… (you reply) there are two possible things that can happen, and if the coin is fair, then they are both equally likely, so the probability of heads is 1/2 or 50%.

Here, we have thought about an event (the coin landing heads), seen that there is one outcome in that event, and two outcomes in the outcome space, so we say the probability of the event, $P(\text{Heads})$, is 1/2.

Example: Tossing a fair six-sided die[6]

Consider rolling a fair six-sided die: six outcomes are possible so $\Omega = \{1, 2, 3, 4, 5, 6\}$. Since the die is fair, each outcome is equally likely, with probability $= \frac{1}{6}$. We can list the outcomes and their probabilities in a table.

Outcome	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Let A be the event that an even number is rolled. Then the set A can be written $\{2, 4, 6\}$. Since all of these outcomes are equally likely:

$$P(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$$

[6]The singular is *die* and the plural is *dice*. If we use the word "die" without any qualifiers, we will mean a fair, six-sided die.

Axioms of probability

In order to compute the probabilities of events, we need to set some basic mathematical rules called *axioms* (which are intuitively clear if you think of the probability of an event as the proportion of the outcomes that are in it). There are three basic rules that will help us compute probabilities:

Axiom 1 The chance of any event is at least 0: $P(A) \geq 0$ for any event A.

Axiom 2 The chance of an outcome being in Ω is 1: $P(\Omega) = 1$. This is true because we can consider that the probability of Ω is the number of outcomes in Ω divided by n, which is $n/n = 1$.

Before we write the third rule, we need some more definitions and notation:

Impossible event An event with no outcomes in it. Denoted by either empty braces $\{\}$ or the symbol for the empty set $\emptyset$. The probability of the impossible event is 0.

Union of events Given events A, B, we can define a new event called A *or* B, which consists of all the outcomes that are *either* in A *or* in B or in both. This is also written as $A \cup B$, read as "A union B''.

Intersection of events Given events A, B, we can define a new event called A *and* B, which consists of all the outcomes that are *both* in A *and* in B. This is also written as $A \cap B$, read as "A intersect B''.

Now we consider events that *don't* intersect or overlap at all, that is, they are *disjoint* from each other, or mutually exclusive:

Mutually exclusive events If two events A and B do not overlap, that is, they have *no outcomes in common*, we say that the events are **mutually exclusive**.

If A and B are mutually exclusive, then we know that if one of them happens, the other one *cannot*. We denote this by writing $A \cap B = \emptyset$ and read this as A *intersect* B is empty. Therefore, we have that

$$P(A \cap B) = P(\emptyset) = 0$$

.

For example, if we are playing De Méré's second game, the event A that we roll a pair of sixes and the event B that we roll a pair of twos cannot happen on the same roll. These events A and B are mutually exclusive.

However, if we roll a die, the event C that we roll an *even* number and the event D that we roll a *prime* number are *not* mutually exclusive, since the number 2 is both even and prime.

Here's another example that might interest soccer fans: The event that Manchester City wins the English Premier League (EPL) in 2024, and the event that Liverpool wins the EPL in 2024 are mutually exclusive, but the events that Manchester City are EPL champions in 2024 and Manchester City are EPL champions in 2023 are *not* mutually exclusive.

Now for the third axiom:

Axiom 3 If A and B are mutually exclusive ($A \cap B = \emptyset$), then

$$P(A \cup B) = P(A) + P(B)$$

That is, for two mutually exclusive events, the probability that *either* of the two events might occur is the *sum* of their probabilities. This is called the **addition rule**.

For example, consider rolling a fair six-sided die, and the two events A and B, where A is the event of rolling a multiple of 5, and B is the event that we roll a multiple of 2.

The only outcome in A is $\{5\}$, while B consists of $\{2, 4, 6\}$. $P(A) = 1/6$, and $P(B) = 3/6$. Since $A \cap B = \emptyset$, that is, A and B have no outcomes in common, we have that

$$P(A \cup B) = P(A) + P(B) = \frac{1}{6} + \frac{3}{6} = \frac{4}{6}$$

The complement rule

Here is an important consequence of axiom 3. Let A be an event in Ω. The *complement* of A, written as A^C, consists of all those outcomes in Ω that are *not* in A. Then we have the following rule:

$$P(A) + P(A^C) = 1$$

This is because $A \cup A^C = \Omega$, and $A \cap A^C\emptyset$).

An example with the axioms: penguins

Consider the penguins dataset, which has 344 observations, of which 152 are Adelie penguins and 68 are Chinstrap penguins. Suppose we pick a penguin at random, what is the probability that we would pick an Adelie penguin? What about a Gentoo penguin?

Check your answer

Let A be the event of picking an Adelie penguin, C be the event of picking a Chinstrap penguin, and G be the event of picking a Gentoo penguin.

Assuming that all the penguins are equally likely to be picked, we see that then $P(A) = 152/344$, and $P(C) = 68/344$.

Since only one penguin is picked, we see that A, C, and G are *mutually exclusive*. This means that $P(A) + P(C) + P(G) = 1$, since A, C, and G together make up all of Ω.

Therefore the complement of G, G^C, which is a penguin that is not Gentoo, consists of Adelie and Chinstrop penguins, and by the *addition rule*,

$$P(G^C) = P(A \cup C) = P(A) + P(C) = (152 + 68)/344 = 220/344$$

Finally, the *complement rule* tells us that

$$P(G) = 1 - P(G^C) = 1 - 220/344 = 124/344$$

.

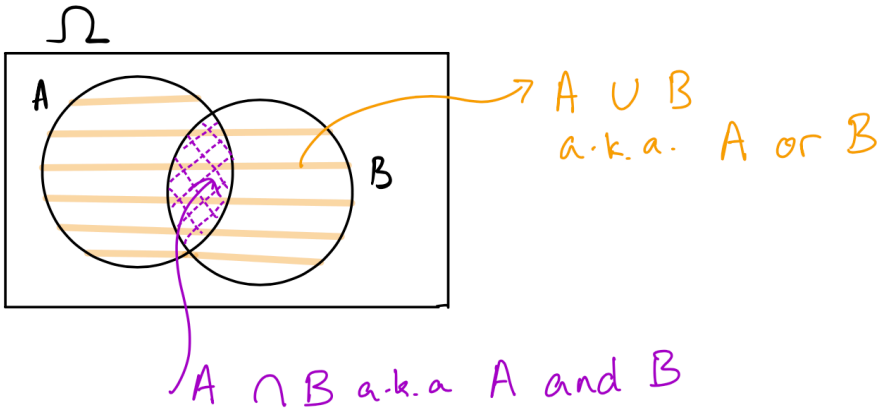
WARNING!!

We use A to denote an *event* or a set, while *P(A)* is a *number* - you can think of $P(A)$ as representing the relative size of A. This means that the following types of statements don't make sense as we haven't defined what it means to add sets or union numbers etc.:

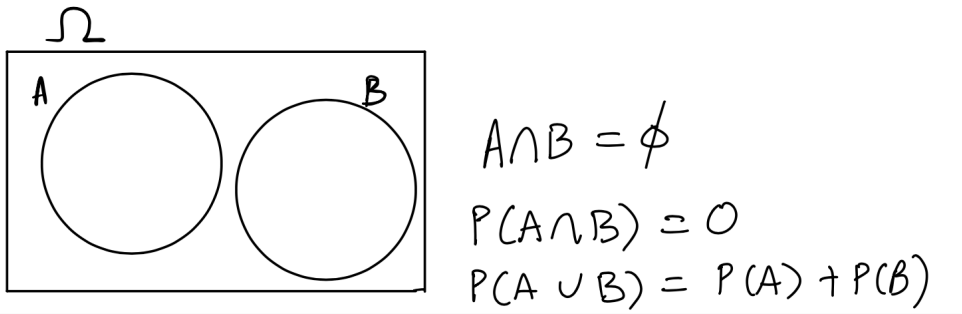
- $P(A) \cup P(B)$ or $P(A) \cap P(B)$
- $A + B$, or $A - B$, or $A \times B$ etc

Venn Diagrams

We often represent events using *Venn diagrams*. The outcome space Ω is usually represented as a rectangle, and events are represented as circles inside Ω. Here is a Venn diagram showing two events A and B, their intersection, and their union:



Here is a Venn diagram showing two mutually exclusive events (no overlap):



Further examples

1. Tossing a fair coin

Suppose we toss a coin twice and record the *equally likely* outcomes. What is Ω? What is the chance of at least one head?

Solution: $\Omega = \{HH, HT, TH, TT\}$, where H represents the coin landing heads, and T represents the coin landing tails. Note that since we can get exactly one head and one tail in **two** ways, we have to write out both ways so that all the outcomes are equally likely.

Now, let A be the event of getting at least one head in two tosses. We can do this by listing the outcomes in A: $A = \{HH, HT, TH\}$ and so $P(A) = 3/4$.

Alternatively, we can consider A^C which is the event of *no* heads, so $A^C = \{TT\}$ and $P(A^C) = 1/4$.

In this case, $P(A) = 1 - P(A^C) = 1 - 1/4 = 3/4$.

Now you try: Let Ω be the outcome space of tossing a coin *three* times. What is the probability of at least one head? What about exactly one head?

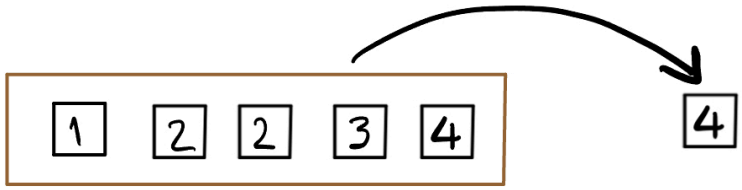
Check your answer

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.

Let A be the event of at least one head. Then A^C is the event of no heads, so $A^C = \{TTT\}$, and $P(A^C) = 1/8$. Therefore $P(A) = 1 - 1/8 = 7/8$. Note that this is much quicker than listing and counting the outcomes in A.

If B is the event of *exactly* one head, then $B = \{HTT, THT, TTH\}$ and $P(B) = 3/8$.

2. A box of tickets



Consider the box above which has five almost identical tickets. The only difference is the value written on them. Imagine that we shake the box to mix the tickets up, and then draw *one* ticket without looking so that all the tickets are *equally likely* to be drawn[7].

What is the chance of drawing an even number?

Check your answer

Solution:

Let A be the event of drawing an even number, then $A = \{2, 2, 4\}$: we list 2 twice because there are two tickets marked 2, making it twice as likely as any other number. $P(A) = 3/5$

[7] We call the tickets equally likely when each ticket has the same chance of being drawn. That is, if there are n tickets in the box, each has a chance of $1/n$ to be drawn. We also refer to this as drawing a ticket *uniformly at random*, because the chance of drawing the tickets are the same, or *uniform*.

3. Tossing a biased coin

Suppose I have a coin that is twice as likely to land heads as it is to land tails. This means that I cannot represent Ω as $\{H, T\}$ since heads and tails are not equally likely. How should I write Ω so that the outcomes are equally likely?

Check your answer

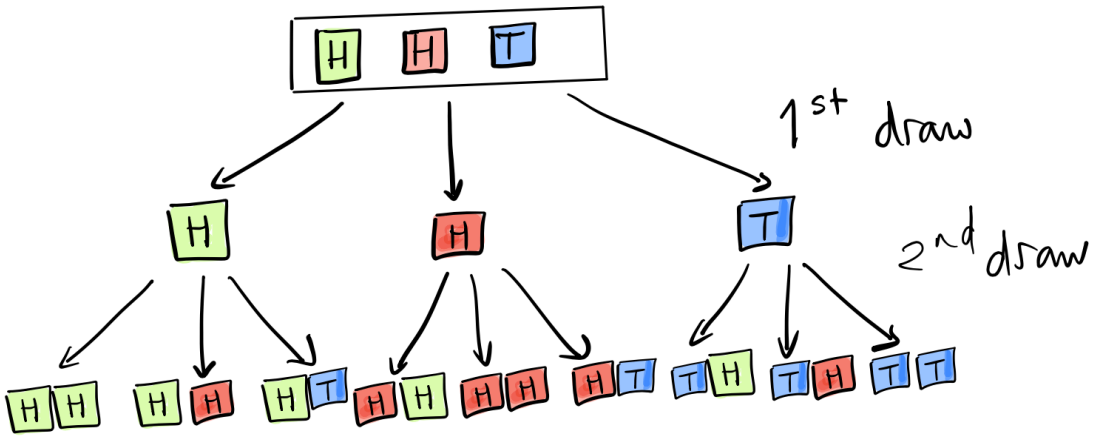
Solution:

In this case, we want to represent equally likely outcomes, and want H to be twice as likely as T. We can therefore represent Ω as $\{H, H, T\}$. Now the chance of the coin landing H can be written as $P(A)$ where A is the event the coin lands H is given by $2/3$.

Suppose we toss the coin twice. How would we list the outcomes so that they are equally likely? Now we have to be careful, and think about all the things that can happen on the second toss if we have H on the first toss.

This is much easier to imagine if we imagine drawing *twice* from a box of tickets, but putting the first ticket back before drawing the second (to represent the fact that the probabilities of landing H or T stay the same on the second toss.)

Now, imagine the box of tickets that represents Ω to be $\boxed{\text{H, H, T}}$. We draw one ticket at first, which could be one of three tickets (there are two tickets that could be H, and one T). We can represent it using the following picture:



From this picture, where we use color to distinguish the two different outcomes of heads and one outcome of tails, we can see that there are 9 possible outcomes that are equally likely, and

we get the following probabilities (where HT, for example, represents the event that the first toss is heads, followed by the second toss being tails.)

$P(HH) = 4/9,\ P(HT) = P(TH) = 2/9, P(TT) = 1/9$ (Check that the probabilities sum to 1!)

> Ask yourself
>
> What box would we use if the coin is not a fair coin, but lands heads 5 out of 6 times?

4. Betting on red in roulette

An American roulette wheel has 38 pockets[8], of which 18 are red, 18 black, and 2 are green. The wheel is spun, and a small ball is thrown on the wheel so that it is equally likely to land in any of the 38 pockets. Players bet on which colored or numbered pocket the ball will come to rest in. If you bet one dollar that the ball will land on red, and it does, you get your dollar back, and you win one more dollar, so your *net gain* is $1. If it doesn't, and lands on a black or green number, you lose your dollar, and your *net "gain"* is -$1.

What is the chance that we will win one dollar on a single spin of the wheel?

Hint Write out the chance of the ball landing in a red pocket, and not landing in a red pocket.

The Ideas in Code

Our first step toward simulating experiments is introducing randomness in R. The following three functions are a good start.

[8]Photo via unsplash.com

Three useful functions

1. `sample()`: randomly picks out elements (items) from a vector

Drawing from a box of tickets is easily simulated in R, since there is a convenient function `sample()` that does exactly what we need: draw tickets from a "box" (which needs to be a vector).

- **Arguments**
  - `x`: the vector to be sampled from, this *must* be specified
  - `size`: the number of items to be sampled, the default value is the length of `x`
  - `replace`: whether we replace a drawn item before we draw again, the default value is `FALSE`, indicating that we would draw *without* replacement.

Example: one sample of size 2 from a box with tickets from 1 to 6

```
die <- c(1, 2, 3, 4, 5, 6)
sample(die, size = 2, replace = FALSE)
```

```
[1] 6 3
```

What would happen if we don't specify values for `size` and `replace`?

```
die <- c(1, 2, 3, 4, 5, 6)
sample(die)
```

```
[1] 2 6 3 4 1 5
```

What would we do differently if we wanted to simulate two rolls of a die?

Check your answer

We would sample twice from the vector `die` *with* replacement:

```
die <- c(1, 2, 3, 4, 5, 6)
sample(die, size = 2, replace = TRUE)
```

```
[1] 6 4
```

2. `set.seed()`: returns the random number generator to the point given by the seed number

The random number generator in R is called a "Pseudo Random Number Generator", because the process can be controlled by a "seed number". These are algorithmic random number generators, which means that if you provide the same seed (a starting number), R will generate the same sequence of random numbers. This makes it easier to debug your code, and reproduce your results if needed.

- **Arguments**
  - `n`: the seed number to use. You can use any number you like, for example 1, or 31415 etc You might have noticed that each time you run sample in the code chunk above, it gives you a different sample. Sometimes we want it to give the same sample so that we can check how the code is working without the sample changing each time. We will use the `set.seed` function for this, which ensures that we will get the same random sample each time we run the code.

Example: one sample of size 2 from a box with tickets from 1 to 6

```
set.seed(1)
sample(die, size = 2, replace = TRUE)
```

```
[1] 1 4
```

Example: another sample of size 2 from a box with tickets from 1 to 6

```
set.seed(1)
sample(die, size = 2, replace = TRUE)
```

```
[1] 1 4
```

Notice that we get the same sample. You can try to run `sample(die)` without using `set.seed()` and see what happens.

Though we used `set.seed()` twice here to demonstrate its purpose, generally, you will only need to run `set.seed()` once time per document. This is a line of code that fits perfectly at the beginning of your work, when you are also loading libraries and packages.

3. `seq()`: creates a sequence of numbers

Above, we created the vector `die` using `die <- c(1, 2, 3, 4, 5, 6)`, which is fine, but this method would be tedious if we wanted to simulate a 20-sided die, for instance. The function `seq()` allows us to create any sequence we like by specifying the starting number, how we want to increment the numbers, and either the ending number or the length of the sequence we want.

- **Arguments**
  - `from`: where to start
  - `by`: size of jump from number to number (the increment)

You can end a sequence in one of two ways: - `to`: at what number should the sequence end - `length`: how long should the sequence be

Example: sequence with the to argument

```
odds_1 <- seq(from = 1, by = 2, to = 9)
odds_1
```

```
[1] 1 3 5 7 9
```

Example: sequence with the length argument

```
odds_2 <- seq(from = 1, by = 2, length = 5)
odds_2
```

```
[1] 1 3 5 7 9
```

Summary

- In this lecture, we introduced equally likely outcomes,and defined the outcome space of an experiment.
- Then, using equally likely outcomes, we defined the probability of an event as the ratio of the number of outcomes in the event to the number of total outcomes in the outcome space.
- We wrote down the axioms (fundamental rules) of probability, after defining unions, intersections, and mutually exclusive events and Venn diagrams.
- In the "Ideas in Code" section, explored how to simulate probabilities using `sample()` and `replicate()`, and learned another useful function `seq()`