MAT220 Lecture 5: Introduction to Probability

Contents

In Lecture 4, we introduced the set theory and counting methods that are needed for probability theory. Today, we will use those ideas to describe probability mathematically. The central idea is that an event is represented by a subset of possible outcomes, and the probability of that event is a number between 0 and 1. When working with classical probability theory, questions about uncertainty can be turned into questions about counting. Throughout this lecture we will explore this idea in depth.

1 What is Probability?

1.1 Uncertainty

In the first few lectures of this course, we introduced the basic language of data and statistics. We saw that data is usually collected from messy real-world situations, and then converted into something that can be analysed mathematically. Today we will deal with another kind of mess: uncertainty.

Consider the following two statements:

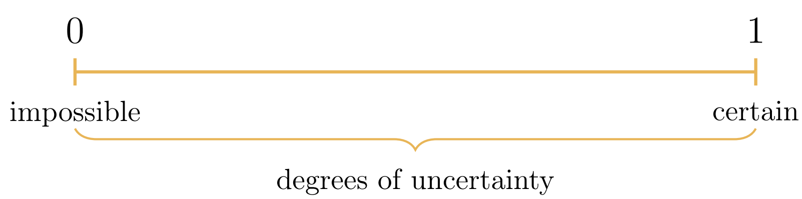
$$2 + 2 = 4 \quad \text{and} \quad 2 + 2 = 5.$$

As you probably know, the first statement is true and the second statement is false. Not only is this obvious, it is *unequivocal*: there is no sense in which the statements could be otherwise. In contrast, consider the two statements:

"Dr. O'Connell's middle name is John", and "Dr. O'Connell's middle name is Jennifer".

As you can see, these non-mathematical statements aren't so clear anymore: if you happened to not know Dr. O'Connell, then you would be uncertain of his middle name. However, even though you don't know my middle name, you can still imagine that having a middle name of Jennifer would be unusual for an Englishman, and having a middle name of John is more plausible. So, although you cannot definitely point to the exact truth of these statements, you can still compare their plausibility.

Probability theory tries to capture this uncertainty using numbers. In reality, the world is full of uncertain things that are not purely logical, and probability theory gives us a way to quantify these beliefs so that they can be compared. If we interpret the number 0 to mean "definitely impossible" and the number 1 to mean "absolutely certain", then all of our uncertain statements would exist somewhere in between:



Basic idea

Probability is the mathematics of uncertainty. It gives us a way to assign numbers to events, so that we can compare how likely those events are.

It should be noted that probability is **not** a spectrum of truth.[1] Instead, it is a spectrum of certainty. So, a probability value of 0.5 does not mean that something is "half true", instead it means that we assign a 50% chance to the statement being true.

[1]For the curious: there is actually a logical theory that does this, called "fuzzy logic". But this is **not** what probability is.

1.2 Defining probability

Mathematically, probability is a function that assigns a number to an event. If the letter A represents something that may or may not happen, then we write $P(A)$ for the probability that A happens. In this sense, we can view the letter P as a little machine that eats events in the world, and spits out a numerical measure of their certainty:

$$\underset{\text{event}}{A} \longrightarrow \boxed{P} \longrightarrow \underset{\text{chance of } A}{P(A)}$$

Generally, we refer to these "things that can possibly happen" as *events*, and we denote them with capital letters such as A, B, C and so on. Formally, the probability assigned to an event can be summarised as follows.

Probability values

For any event A, its probability satisfies

$$0 \leq P(A) \leq 1.$$

Here $P(A) = 0$ means that A is impossible, while $P(A) = 1$ means that A is certain.

Since probabilities are mapped somewhere onto the line between 0 and 1, we can directly compare degrees of uncertainty of different events. Suppose that we have two events A and B, and we happened to know their probabilities $P(A)$ and $P(B)$. If $P(A) < P(B)$, then event B is more likely than event A. For example, if

$$P(A) = 0.2 \qquad \text{and} \qquad P(B) = 0.7,$$

then B is more likely to happen than A, because $P(B)$ is closer to the absolute certainty of 1 than $P(A)$ is.

1.2.1 Expressing probabilities

Since any probability is a number between 0 and 1, we can write it down in several equivalent ways. For example, if the probability of an event A is 50%, we could write it in any of the following ways:

$$P(A) = \frac{1}{2} = 0.5 = 50\% = \frac{250}{500}.$$

Here $\frac{1}{2}$ is a reduced fraction, 0.5 is a decimal, 50% is a percentage, and $\frac{250}{500}$ is an unreduced fraction. In the last case, perhaps we are working with a system that has 500 possibilities, and 250 of the possible outcomes are favourable. The unreduced fraction $\frac{250}{500}$ tells us this information, but the reduced fraction $\frac{1}{2}$ does not.

1.3 Classical probability

In this lecture, we mostly focus on *classical probability*. This is the setting in which there are finitely many possible outcomes and each one is equally likely.

Definition: Classical probability

A probability problem is a problem in **classical probability** when there are finitely many possible outcomes and each outcome is equally likely.

Examples include:

- rolling a fair dice,
- flipping a fair coin,
- dealing cards from a well-shuffled deck,
- selecting answers to true/false questions.

In this setting, probability can usually be calculated by counting.

1.3.1 A note on randomness

The question of true randomness is deeply philosophical. Some people believe the world is totally deterministic, meaning everything is determined from past events and the laws of physics. Other people believe there are genuinely random things, for example observations in quantum mechanics, but even there, things are debated. In any case, it does not matter for our purposes whether true randomness exists or not. In this course we use probabilities to quantify uncertainty. So, even if things are not truly random in reality, *they may still appear random to us*.

Randomness in this course

For our purposes, an outcome is random if we cannot predict it with certainty before the experiment is performed.

This practical interpretation is enough for the kind of probability calculations we will do in this course.

1.4 Assigning probabilities by counting

When working with classical probability, we can calculate the probabilities of individual events by taking a ratio of the "positive cases" compared to the collection of all possibilities. For any event, we can assign a probability using the formula:

$$\text{Probability of an event} = \frac{\text{number of outcomes favourable to the event}}{\text{total number of outcomes}}.$$

Put differently, we count the number of outcomes that would cause the event to be true, and then divide by the total number of possible outcomes. In this fraction, the numerator will always be less than or equal to the denominator, so when we take the ratio we will always obtain a value between 0 and 1.

The Golden Formula for Classical Probabilities

If all outcomes are equally likely, then

$$P(A) = \frac{\text{number of outcomes favourable to } A}{\text{total number of outcomes}}.$$

We will now demonstrate how to use this formula to calculate some basic probabilities.

1.4.1 The Probability of an Individual Outcome

The probability of any individual outcome can be determined with this formula. For example:

- for a fair coin, the probability of the coin landing on heads is $\frac{1}{2}$, because there are only two possibilities, one of which is heads.
- When rolling a fair dice, the probability of landing on a 5 will be $\frac{1}{6}$, because there is only one 5 written on the dice, and there are 6 numbers in total.
- When randomly drawing a card from a shuffled deck, the probability of picking the ace of spades is $\frac{1}{52}$, because there are 52 cards in a deck, and only one of them is the ace of spades.

Generally, for any system in classical probability, the formula above tells us:

$$\text{Probability of any individual outcome} = \frac{1}{\text{total number of outcomes}}.$$

1.4.2 The probability of a more-complicated event

We can also use our Golden Formula to calculate more complicated probabilities. As previously mentioned, this can be achieved by looking through the space of possibilities and counting up all of the outcomes that make the event true. As an example, suppose that we are rolling a fair dice. Let the event A be:

$$A = \text{“we land on an even number”},$$

and let the event B be:

$$B = \text{“we land on a number less than 5”}.$$

To determine these probabilities, we need to calculate the number of “positive outcomes” for each, and then divide by 6, since there are 6 sides of the dice. For the event A we have:

$$P(A) = \frac{\text{how many numbers are even}}{\text{how many numbers there are in total}} = \frac{3}{6} = \frac{1}{2},$$

and similarly for B, the probability is:

$$P(B) = \frac{\text{how many numbers are less than 5}}{\text{how many numbers there are in total}} = \frac{4}{6} = \frac{2}{3}.$$

1.4.3 Example: the Toy Cabinet

Suppose that your overly enthusiastic cousin has a cabinet with 58 different toys inside, and he picks one at random to show you. Before he does this, you do not know which one he will pick, so from your perspective his choice is essentially random. Observe the cabinet below:

Let event A be: “the toy chosen is a blue car”. By examining the above image, we see that there are two blue cars in the cabinet. Therefore, according to our Golden Formula

$$P(A) = \frac{2}{58}.$$

Let event B be: “the toy chosen comes from the fourth row”. In the image above, we count and see that the fourth row contains 8 toys, so event B contains 8 outcomes. Therefore:

$$P(B) = \frac{8}{58}.$$

The calculation is the same in both cases: count the favourable toys, then divide by the total number of toys.

Exercise: True/false answer sequences

Suppose that three homework questions are true/false questions. To choose the correct answers randomly, Dr. O’Connell writes down all possible true/false sequences and chooses one from a hat. There are 8 possible sequences of 3 T/Fs in a row:

TTT, TTF, TFT, FTT, TFF, FTF, FFT, FFF

(a) What is the probability that all 3 answers are False?
(b) What is the probability that exactly two answers are False?
(c) If a student blindly guesses all three answers, what is the probability that they get all 3 correct?

Solution

These three questions can all be answered using the Golden Formula mentioned previously. Note first that there are 8 possible true/false sequences.

(a) Only one outcome is favourable, namely FFF. Therefore

$$P(\text{all 3 False}) = \frac{1}{8}.$$

(b) The favourable outcomes are TFF, FTF and FFT. Therefore

$$P(\text{exactly two False}) = \frac{3}{8}.$$

(c) There is exactly one correct sequence, so

$$P(\text{all 3 correct}) = \frac{1}{8}.$$

2 The Rules of Probability

2.1 Simple events, sample spaces and events

We now organise the vocabulary of probability more carefully. The most important objects are simple events, the sample space and events.

Important definitions

(1) A **simple event** consists of exactly one outcome of the experiment.

(2) The **sample space**, usually written S, is the set of all possible simple events.

(3) An **event** is any subset of the sample space, meaning it can contain one outcome or many outcomes.

(4) The probability of an event A is written $P(A)$. This is a number between 0 and 1.

(5) The sum of the probabilities of all simple events in a sample space is always 1.

To see these definitions in action, let us return to the toy cabinet example. In this case, the sample space will be the set of all 58 toys in the cabinet:

$$S = \{\text{all toys in the cabinet}\}.$$

In this example, a simple event would be the selection of an individual toy, something like "picking this toy in particular". Mathematically, this would be any subset of S that is of size one. A general event would be any subset of S at all, and these subsets would represent more general statements such as "picking a toy from the fourth row" or "picking a blue car".

Just to emphasise this: the event *is* the subset of the sample space. For example, in the toy cabinet example above, we considered the event A meaning "selecting a blue car", and the probability was determined to be $\frac{2}{58}$. The mathematical definition of the event A would be the set containing two objects: the two blue cars taken from the cabinet.

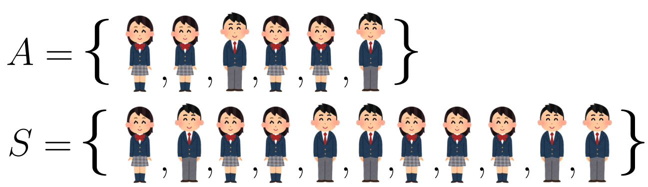
2.2 Probability is secretly set theory

Let's consider another example. Suppose that Dr. O'Connell has 11 students in his class. He lines them up in two rows: 6 students in the front row, and 5 students in the back row:

For argument's sake, what would be the probability that Dr. O'Connell randomly selects somebody from the first row? Well, assuming that each choice were equally likely, according to our golden formula:

$$P(A) = \frac{\text{number of students in the front row}}{\text{number of students in total}} = \frac{6}{11}.$$

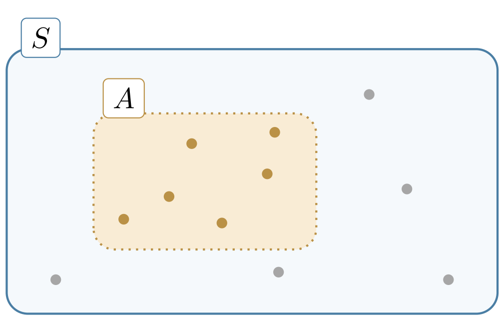
In terms of the terminology from before, the sample space S corresponds to the set of all students, and the event A would be the subset consisting of those students in the first row:



By a quick count, we see that $|A| = 6$ and $|S| = 11$. Taking the ratio of these returns the probability $P(A)$. So, in set-theoretic language, our golden formula now reads:

$$P(A) = \frac{|A|}{|S|}.$$

This says that to calculate the probability of an event A, we count the size of the subset A and divide by the size of the sample space S. More abstractly still, we are performing a count like the diagram below.



$$|A| = 6$$

$$|S| = 11$$

$$P(A) = \frac{6}{11}$$

For future reference, we will restate our formula.

The Golden Formula for Classical Probabilities Restated

If all outcomes are equally likely, then the probability of an event A can be determined by

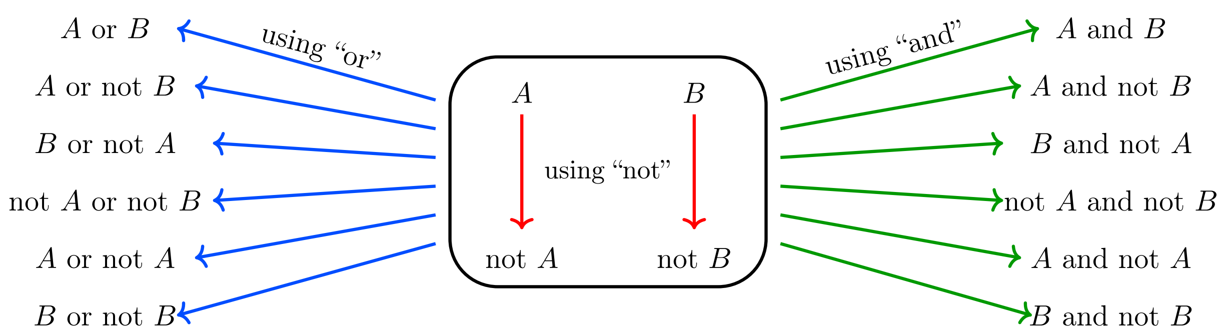
$$P(A) = \frac{|A|}{|S|},$$

that is, the cardinality of A divided by the cardinality of the sample space S.

3 Compound Probabilities

3.1 Building compound events

Suppose we have two events A and B. Based on our discussion so far, we know how to determine the probabilities $P(A)$ and $P(B)$: we simply count up the number of outcomes that correspond to A or B, and then divide that number by the size of the sample space. As a matter of fact, we can build many other events from A and B by applying three logical building blocks, namely: *not*, *and*, and *or*. For instance, applications of these three words to A and B allow us to create many other statements, which are all events in their own right:



This diagram looks suspiciously similar to a diagram we saw in Lecture 4 when discussing set theory. In fact, it is the *same diagram*. The three operations of *not*, *and*, and *or* can be directly translated into set theory as follows:

Language	**Set operation**	**Meaning**
not A	A^c	the complement of A
A and B	$A \cap B$	the intersection of A and B
A or B	$A \cup B$	the union of A and B

According to the previous section, we can calculate the probability of any event by counting up the size of the subset, and then dividing by the size of the sample space. In the case of the three statements above, this creates three new formulas:

$$P(\text{not } A) = \frac{|A^c|}{|S|},$$

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|},$$

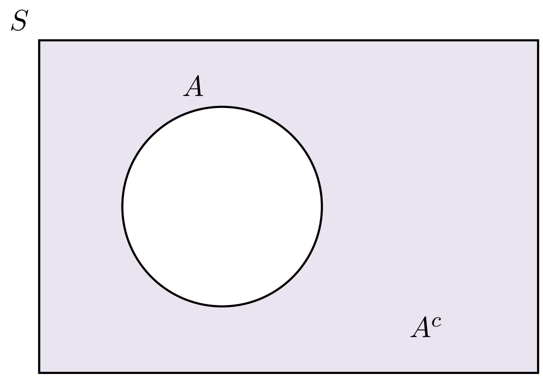
and

$$P(A \text{ or } B) = \frac{|A \cup B|}{|S|}.$$

Therefore, if we can figure out the sizes of these sets, then we can create formulas to calculate the corresponding compound probabilities. We will now explain these one-by-one.

3.2 The complement: not A

The complement A^c means everything in the sample space that is not in the subset A. In the picture below, the complement is the shaded area.



According to the classical probability formula, to calculate the probability of not-A, we should count the size of the complement A^c:

$$P(\text{not } A) = \frac{|A^c|}{|S|}.$$

Since A and A^c together cover the whole sample space, we have

$$|A| + |A^c| = |S|.$$

Therefore

$$|A^c| = |S| - |A|.$$

Substituting this into the probability formula gives

$$P(A^c) = \frac{|A^c|}{|S|} = \frac{|S| - |A|}{|S|} = \frac{|S|}{|S|} - \frac{|A|}{|S|} = 1 - P(A).$$

Complement rule

For any event A,

$$P(\text{not } A) = 1 - P(A).$$

Example. In the true/false sequence example, let E be the event "FFF is chosen". Since there are 8 equally likely sequences,

$$P(E) = \frac{1}{8}.$$

The complement E^c is the event "FFF is not chosen". There are 7 favourable outcomes, namely everything except FFF. Therefore,

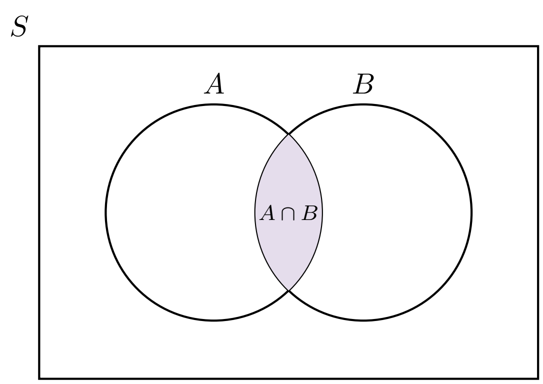
$$P(E^c) = \frac{7}{8}.$$

Equivalently, using the complement rule,

$$P(E^c) = 1 - P(E) = 1 - \frac{1}{8} = \frac{7}{8}.$$

3.3 The intersection: A and B

The intersection of A and B, written $A \cap B$, is the set of outcomes that are in both A and B at the same time. Visually, it is the overlap of the two sets:



According to our Golden Formula, the probability of the compound event A and B should involve the size of their intersection:

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|}.$$

Unfortunately, there is no general formula for the size of $A \cap B$ in terms of only $|A|$ and $|B|$. Two large sets might overlap a lot, overlap a little, or not overlap at all. So, to calculate $P(A \text{ and } B)$, we need information about how the two events are related to each other.

3.4 Conditional probability

Suppose that we try to take our formula for $P(A \text{ and } B)$ and perform some algebra. We will be clever and artificially insert a copy of 1, disguised in the form $\frac{|B|}{|B|}$. Performing some basic algebra, we obtain:

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|} = \frac{|A \cap B|}{|S|}\frac{|B|}{|B|} = \frac{|A \cap B|}{|B|} \cdot \frac{|B|}{|S|}.$$

Thus we have managed to express our probability as a product of two other ratios. The second of these fractions is familiar: it is the definition of $P(B)$. In contrast, the first ratio written above is new: it is a special probability known as a *conditional probability*.

Definition: Conditional probability

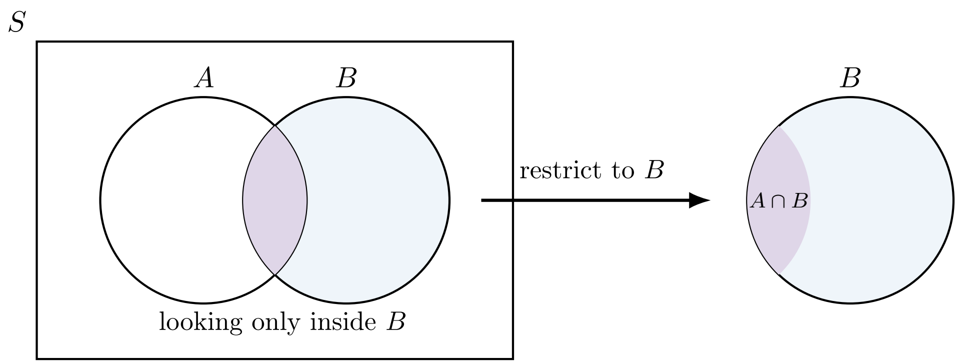
The conditional probability $P(A \mid B)$ is the probability that A happens, assuming that B has happened. In classical probability,

$$P(A \mid B) = \frac{|A \cap B|}{|B|},$$

provided that $|B| \neq 0$.

This notion of conditional probability seems confusing at first glance. So, we will now explain what is happening using set theory.

As previously mentioned, there is no perfect formula that we can use to calculate the size of the intersection $A \cap B$. One way to figure out this size is to look inside the set B and then to count up all of the instances of A that we find. This is like taking our entire sample space and selectively ignoring everything that is not already in B:



If we were to try and compute the probabilities based on this new set up, we notice that the sample space is no longer S, instead it is the subset B. In that sense, since the sample space is now B, if we were to try to count the probability of A happening in this restricted context, we would need to use B in the denominator in our golden formula:

$$P(A,\ \text{within } B) = \frac{\text{number of elements of } A \text{ that are also in } B}{\text{the size of } B}.$$

Any element of A that is also in B falls precisely in the intersection $A \cap B$. Therefore, the probability associated to the above allows us to recover the previous formula for conditional probability:

$$P(A \mid B) = P(A,\ \text{within } B) = \frac{\text{number of elements of } A \text{ that are also in } B}{\text{the size of } B} = \frac{|A \cap B|}{|B|}.$$

As an example, suppose that we roll a fair dice. Let

$$A = \{\text{we roll an even number}\} \quad \text{and} \quad B = \{\text{we roll a number less than } 4\}.$$

Written out as subsets, the events A and B read:

$$A = \{2, 4, 6\}, \quad \text{and} \quad B = \{1, 2, 3\}.$$

So, you throw the dice up into the air. Imagine that before the dice lands, God came down from the heavens and whispered into your ear: "the outcome will be less than 4". This means that, in that moment, you know for *certain* that it is not possible to roll a 4, 5 or 6. In this situation, if you were to evaluate your chances of event A happening and try to calculate a new probability, you would cut your space of possibilities from

$$\{1, 2, 3, 4, 5, 6\}$$

down to the smaller set

$$\{1, 2, 3\}.$$

Inside this smaller set, only one outcome is even, namely 2. Therefore

$$P(A \mid B) = \frac{1}{3}.$$

Using set notation,

$$A \cap B = \{2\},$$

so

$$P(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{1}{3}.$$

Putting this all together, we see that

$$P(B) = \frac{3}{6} = \frac{1}{2}, \qquad P(A \text{ and } B) = \frac{1}{6}, \qquad P(A \mid B) = \frac{1}{3}.$$

Funnily enough, it just so happens that:

$$\frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}.$$

But, this is no coincidence: it is a manifestation of the general formula we mentioned before.

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|} = \frac{|A \cap B|}{|B|} \cdot \frac{|B|}{|S|} = P(A \mid B)P(B).$$

General multiplication rule

When the relevant conditional probability is defined, for any events A and B,

$$P(A \text{ and } B) = P(B)P(A \mid B).$$

Equivalently,

$$P(A \text{ and } B) = P(A)P(B \mid A).$$

3.5 Independent and dependent events

In some cases, the formula for $P(A \text{ and } B)$ can be heavily simplified. In order to properly explain how, we need to first make the distinction between independent and dependent events.

Independent and dependent events

Let A and B be events.

- A and B are **independent** if learning that one happened does not change the probability of the other.
- A and B are **dependent** if learning that one happened *does* change the probability of the other.

As a reminder, probability is the science of uncertainty. So, the idea behind dependent events is that knowing the event B happened might affect the probability of A. This can potentially happen whenever the information about B is somehow relevant for the event A.

In the previous example of rolling the dice, the events A being "rolling an even number" and B being "rolling a number less than 4" are *dependent* events, because knowing that B would happen actually changed our probability from $\frac{1}{2}$ to $\frac{1}{3}$. In other words, information about one event affected our knowledge about the possibility of the other.

However, we can also imagine situations where knowing one piece of information gives us no reason to change our assessment of probability. For example, let the event C be "we roll a number greater than 2". Then, as a subset,

$$C = \{3, 4, 5, 6\}.$$

If, for some reason, we absolutely *knew* that the outcome of our dice roll would be larger than 2, then we could restrict our sample space from

$$\{1, 2, 3, 4, 5, 6\}$$

down to the subset

$$\{3, 4, 5, 6\}.$$

Given that this restricted set consists of two odd numbers and two even numbers, the probability of rolling an even number is still $\frac{1}{2}$. This is the same as $P(A)$, so knowing that C will happen does not actually change the amount of certainty that we have about the system. Therefore, the events A and C are independent.

In symbols, A and B are independent when

$$P(A \mid B) = P(A),$$

and equivalently,

$$P(B \mid A) = P(B).$$

3.6 Multiplication rule for independent events

If A and B are independent, then $P(A \mid B) = P(A)$ can be used to drastically simplify the formula for $P(A \text{ and } B)$. We may perform a substitution to give the following multiplication rule.

Multiplication rule for independent events

If A and B are independent events, then

$$P(A \text{ and } B) = P(A)P(B).$$

Example: Two coin flips. Suppose that we flip a fair coin twice.
Let A be the event "the first flip is tails", and let B be the event "the second flip is tails". These events are independent, because the result of the first flip does not affect the result of the second flip.
Therefore,

$$P(A \text{ and } B) = P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Equivalently, the sample space is

$$S = \{\text{HH, HT, TH, TT}\},$$

and only TT is favourable.

Example: Two separate decks. Suppose that we shuffle two separate decks of cards.
Let A be the event "the top card of deck 1 is an ace", and let B be the event "the top card of deck 2 is an ace". These events are independent, and each has probability

$$\frac{4}{52} = \frac{1}{13}.$$

So

$$P(A \text{ and } B) = \left(\frac{1}{13}\right)\left(\frac{1}{13}\right) = \frac{1}{169}.$$

Equivalently, there are $52 \times 52 = 2704$ possible ordered pairs of top cards, and $4 \times 4 = 16$ of these pairs consist of two aces. Hence

$$\frac{16}{2704} = \frac{1}{169}.$$

The following is an example of dependent events.

Example: Two aces from one deck. Suppose that we shuffle one deck of 52 cards. Let

$$A = \{\text{the first card is an ace}\},$$

and

$$B = \{\text{the second card is an ace}\}.$$

These events are dependent, because if the first card is an ace, then there are only 3 aces left among the remaining 51 cards when we try to take a second. We calculate:

$$P(A) = \frac{4}{52} = \frac{1}{13},$$

and

$$P(B \mid A) = \frac{3}{51}.$$

Therefore,

$$P(A \text{ and } B) = P(A)P(B \mid A) = \frac{1}{13} \cdot \frac{3}{51} = \frac{1}{221}.$$

How to calculate $P(A$ and $B)$

Step 1: Decide whether the events are independent. If $P(A) = P(A \mid B)$, then A and B are independent.

Step 2: If A and B are independent, use

$$P(A \text{ and } B) = P(A)P(B).$$

Step 3: For any events, independent or dependent, we may use

$$P(A \text{ and } B) = P(A)P(B \mid A),$$

or equivalently,

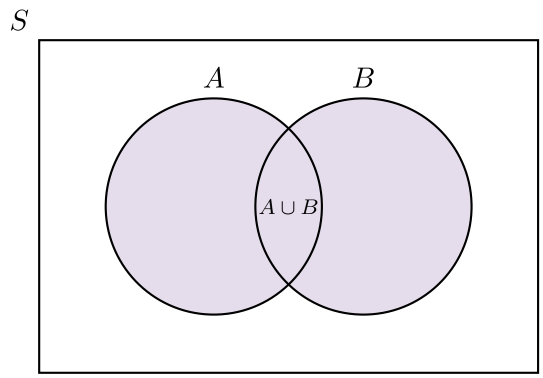
$$P(A \text{ and } B) = P(B)P(A \mid B).$$

3.7 The union: A or B

According to Section 3.1, the compound event "A or B" corresponds to the set-theoretic union $A \cup B$. Thus, in order to calculate the probability $P(A$ or $B)$, we can count up the size of this subset and divide it by the size of the sample space:

$$P(A \text{ or } B) = \frac{|A \cup B|}{|S|}.$$

The union of A and B, written $A \cup B$, is the set of outcomes that are in A, in B, or in both.



As with the intersection discussed earlier, generally we cannot deduce the size of the union $A \cup B$ based only on the sizes of A and B. In fact, we noticed last lecture that it is very tempting to naively write:

$$|A \cup B| = |A| + |B| .$$

However, this is wrong in general, because the overlap $A \cap B$ accidentally gets counted twice. For example, if

$$A = \{1, 2, 3\}, \qquad \text{and} \qquad B = \{2, 3, 4\},$$

then their union is the set

$$A \cup B = \{1, 2, 3, 4\},$$

which has cardinality equal to 4. However,

$$|A| + |B| = 3 + 3 = 6 \neq 4.$$

The issue is that there are two copies of the intersection $\{2, 3\}$ hiding away inside the sum $3 + 3 = 6$.

Generally speaking, to fix this issue we can subtract one copy of the intersection:

$$|A \cup B| = |A| + |B| - |A \cap B| .$$

We can use this observation to derive a general formula for $P(A \text{ or } B)$ in terms of $P(A)$, $P(B)$ and $P(A \text{ and } B)$:

$$\begin{aligned} P(A \text{ or } B) &= \frac{|A \cup B|}{|S|} \\ &= \frac{|A| + |B| - |A \cap B|}{|S|} \\ &= \frac{|A|}{|S|} + \frac{|B|}{|S|} - \frac{|A \cap B|}{|S|} \\ &= P(A) + P(B) - P(A \text{ and } B). \end{aligned}$$

This is known as the *addition rule*, and it deserves its own box.

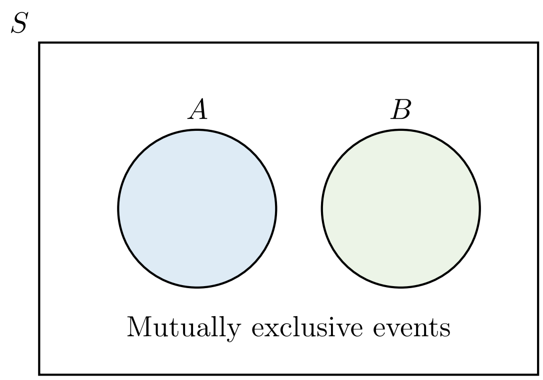
Addition rule

For any events A and B,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

3.8 Mutually exclusive events

Sometimes two events cannot happen at the same time. For example, a single coin flip cannot land on heads and tails at the same time. Such events are called *mutually exclusive*. Visually, mutually exclusive events have no overlap:



Definition: Mutually exclusive events

Two events A and B are **mutually exclusive** if they cannot happen at the same time. Equivalently, $A \cap B = \emptyset$, meaning that

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|} = \frac{0}{|S|} = 0.$$

Using this observation, we now notice that in our formula for the probability of A or B, the third term vanishes for mutually exclusive events:

$$\begin{aligned} P(A \text{ or } B) &= \frac{|A \cup B|}{|S|} \\ &= \frac{|A| + |B| - \cancel{|A \cap B|}}{|S|} \\ &= \frac{|A|}{|S|} + \frac{|B|}{|S|} - \cancel{\frac{|A \cap B|}{|S|}} \\ &= P(A) + P(B) - \cancel{P(A \text{ and } B)} \\ &= P(A) + P(B). \end{aligned}$$

This means that our formula will reduce to the simple addition of the probabilities of A and B. This is known as the addition rule for mutually exclusive events.

Addition rule for mutually exclusive events

If A and B are mutually exclusive, then

$$P(A \text{ or } B) = P(A) + P(B).$$

As an example, suppose that we shuffle a standard deck of 52 cards. Define the events:

$$A = \{\text{the top card is an ace}\},$$

$$B = \{\text{the top card is a spade}\},$$

and

$$C = \{\text{the top card is a king}\}.$$

The events A and B are not mutually exclusive, because the ace of spades is both an ace and a spade. However, A and C are mutually exclusive, because the top card cannot be both an ace and a king at the same time.

Exercise: Ace, spade and king

Compute the following.

(a) $P(A \text{ or } B)$.

(b) $P(A \text{ or } C)$.

Solution

(a) We have

$$P(A) = \frac{4}{52}, \qquad P(B) = \frac{13}{52}, \qquad P(A \text{ and } B) = \frac{1}{52}.$$

Therefore,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}.$$

(b) Since A and C are mutually exclusive,

$$P(A \text{ or } C) = P(A) + P(C).$$

Therefore,

$$P(A \text{ or } C) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}.$$

Summary: Compound probabilities

The main compound events in this lecture are built from the words *not*, *and*, and *or*. In set notation, these correspond to

$$\text{not } A \longleftrightarrow A^c, \qquad A \text{ and } B \longleftrightarrow A \cap B, \qquad A \text{ or } B \longleftrightarrow A \cup B.$$

For complements, use

$$P(\text{not } A) = 1 - P(A).$$

For intersections, use the multiplication rule

$$P(A \text{ and } B) = P(A)P(B \mid A),$$

or equivalently,

$$P(A \text{ and } B) = P(B)P(A \mid B).$$

If the events are independent, this simplifies to

$$P(A \text{ and } B) = P(A)P(B).$$

For unions, use the addition rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

If the events are mutually exclusive, then $P(A \text{ and } B) = 0$, so this simplifies to

$$P(A \text{ or } B) = P(A) + P(B).$$