MAT120: Lecture 10 Handout

Introduction to Probability

In the previous part of the course we studied modelling, meaning we used equations to describe relationships between variables. In this lecture we start a new topic: probability. Probability is the mathematics of uncertainty. It gives us a way to talk about how likely events are, and it will be the foundation for the statistics unit that comes next.

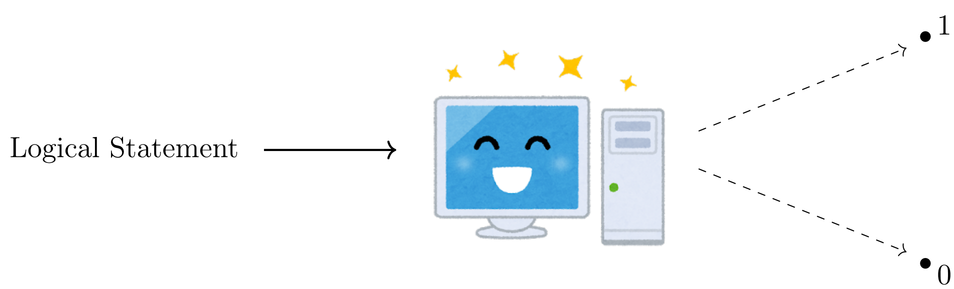
1 What is Probability?

1.1 Logic and Uncertainty

In the first few lectures of this course, we introduced Logic, which was a“formal language” that we can use to express the structure of arguments. We saw that some arguments were *valid*, meaning that they followed a correct logical structure, and some arguments were *sound*, meaning that their premises were actually true in the real world. A good argument was both valid and sound, i.e. one which starts from true premises and follows correct steps of logic to draw a conclusion that is also true.

Fundamental to this approach is the idea that we can assign a“truth value” to every statement – every logical statement is either true or false. In fact, back in Lecture 5 we saw that computers also operate in a similar way: they interpret electrical signals as either“on” or“off” (i.e. 1 and 0), and the computations that they do are simply operations built from logical combinations of these input signals. Simply put: a computer *thinks in logic*, and *counts in binary*.

We can imagine computers as a sort of“logic machine”, in that they are able to take any logical statements and spit out an evaluation of either True or False. If we replace T with 1 and F with 0, we can visualize the computers evaluation of statements as follows:

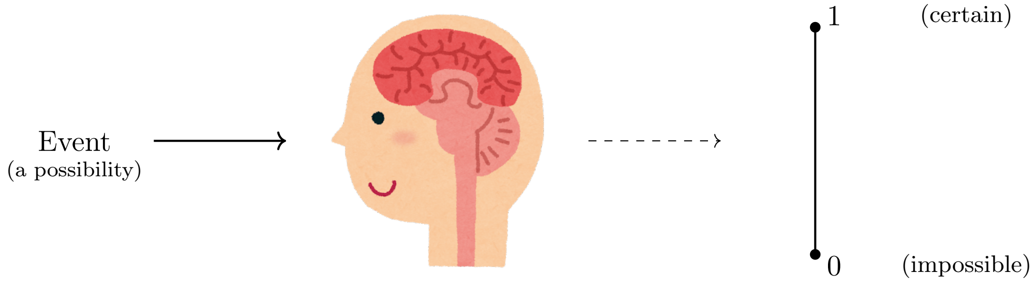


Computers are very useful for talking about logical statements that involve clear, precise things like mathematics or electrical signals. However, there is a sense in which we humans do not exactly think like this: often times, the world forces us to reason without fully knowing whether a statement is true or false. For example, consider the statement: “it is going to rain tomorrow morning”. Perhaps if we were a meteorologist we might know the answer, but we are not meteorologists, so the most that we can honestly say is “I don’t know.. maybe?” In this sense, our limitation is not that the statement lacks a truth value, but that we do not know which truth value is correct. In other words, the world can be uncertain to us, yet we still manage to reason under that uncertainty.

1.2 Probability Theory

Probability theory is the mathematical study of uncertainty. It is different from logic in the following sense: logic tells us what conclusions follow *with certainty* from given premises, while probability gives us a way to measure how likely different outcomes are when we do not have complete information. In a way, probability theory can be seen as the science of quantifying beliefs.

In this course, we will think of an *experiment* (like rolling a die or flipping a coin) as having several possible outcomes, but before the experiment is performed we do not know which outcome will actually happen. Probability assigns a number between 0 and 1 to each of these possible outcomes, and we interpret this number as a measure of how plausible that outcome is. The theory of probability takes this idea and extends it by developing clever rules that we can use to determine these numbers based on previous information or on the set-up of the experiment that we are working with. In distinction to the computer we mentioned previously, our uncertain brains might think of unpredictable outcomes as more of a spectrum:



It should be noted that probability is **not** a "spectrum of truth" between 0 and 1.[1] Instead, it is a spectrum of *certainty*: for each event A (a collection of outcomes), we can assign a number $P(A)$ between 0 and 1, where $P(A) = 0$ means the outcome is impossible, $P(A) = 1$ means the outcome is guaranteed to happen, and values in between represent different levels of uncertainty about which outcome will occur.

1.3 Randomness

We will assume that the actual outcomes in our systems are random, meaning that there is no rule (available to us) that decides what the outcome will be. Now we might ask: what is randomness, exactly? Is the real world actually random? For example, suppose that I randomly pick a number between 1 and 10, for example 3. Why did I pick this number? Did I really pick it randomly, or did I pick it based on the sum of my past experiences that lead me to select 3 as my choice?

Putting this another way: suppose that I flip a coin. Is the outcome truly random? Surely it is determined by the laws of physics somehow. If I knew everything about how much force I applied to the coin as I flicked it up into the air, could I calculate the outcome with certainty?

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 aren't truly random in reality, *they may still appear random to us.* Thus, probability theory is still useful: it can help us make decisions amongst the uncertainty.

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

2 Basic Properties of Probability

2.1 Probability as a concept

Mathematically, probability is a function that assigns a number to the possible outcomes of some random process. In probability theory, we tend to consider collections of outcomes called *events*, and we usually write these with capital letters $A, B, C, \ldots$. We then write $P(A)$ to mean the probability of the event A. An *outcome* is one specific result (like "roll a 4"), the set of all outcomes is called the *sample space* and is written S, and an event A is a subset of S (a collection of outcomes). There are some important ground rules that are used in probability:

(1) For any event A, we always have $0 \leq P(A) \leq 1$.

(2) $P(A) = 0$ means "impossible" and $P(A) = 1$ means "certain".

(3) If $P(A) < P(B)$, then event B is more likely than event A.

In this course we mostly focus on *classical probability*, meaning that we have a finite number of possible outcomes and we treat them as equally likely. Examples include rolling a die, flipping a coin, dealing cards, or guessing a true/false multiple choice question.

2.2 Assigning a Probability

Suppose that an experiment has a finite set of equally likely outcomes (meaning every outcome has the same probability of happening). Recall that an "event" is just some collection of outcomes. 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 favourable outcomes and 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.

Observe also that the above formula is consistent with our assumption that each outcome is equally likely: we can imagine an individual outcome as a collection of size 1, so that makes the formula above reduce to the special case:

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

2.2.1 Expressing probabilities

A probability is always a number between 0 and 1, but we can write it in several equivalent ways. Sometimes the way we write a probability gives extra information about the system. For example, if the probability of an event A is 50%, we could equivalently write it as:

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

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, and they all represent the same probability. In the last case, perhaps we are working with a system that has 500 possibilities, and 250 of the possible outcomes are positive. The unreduced fraction $\frac{250}{500}$ tells us this information, but the reduced fraction $\frac{1}{2}$ does not.

2.3 Example: The Toy Cabinet

Your overly enthusiastic cousin has a cabinet with 58 different toys, and he picks one at random to show you. Before he does this, you don't know which one he will pick, and you assume that he could pick any toy with the same likelihood.

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:

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

Let event B be: "the toy chosen comes from the 4th row". In the image above, we count and see that the 4th row contains 8 toys, so event B contains 8 outcomes. Therefore:

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

Let event C be: "the toy chosen is green". After some work scanning the cabinet for green toys, we see that there are 10 green toys. Therefore:

$$P(C) = \frac{10}{58}.$$

2.4 Exercise: true/false combinations

Dr. O'Connell writes a homework set with 3 true/false questions. He lists all possible answer patterns, then chooses one at random.

The sample space has 8 equally likely outcomes:

$$\{\text{TTT, TTF, TFT, FTT, TFF, FTF, FFT, FFF}\}.$$

Exercise 1

Answer the following.

(1) What is the probability that all 3 answers will be False?

(2) What is the probability that exactly two of the answers will be True?

(3) If a student blindly guesses, what is the probability they get all 3 correct?

Solution

(1) Only one outcome is favourable (FFF), so

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

(2) The favourable outcomes are TTF, TFT, and FTT (3 outcomes), so

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

(3) There is exactly one correct pattern, so

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

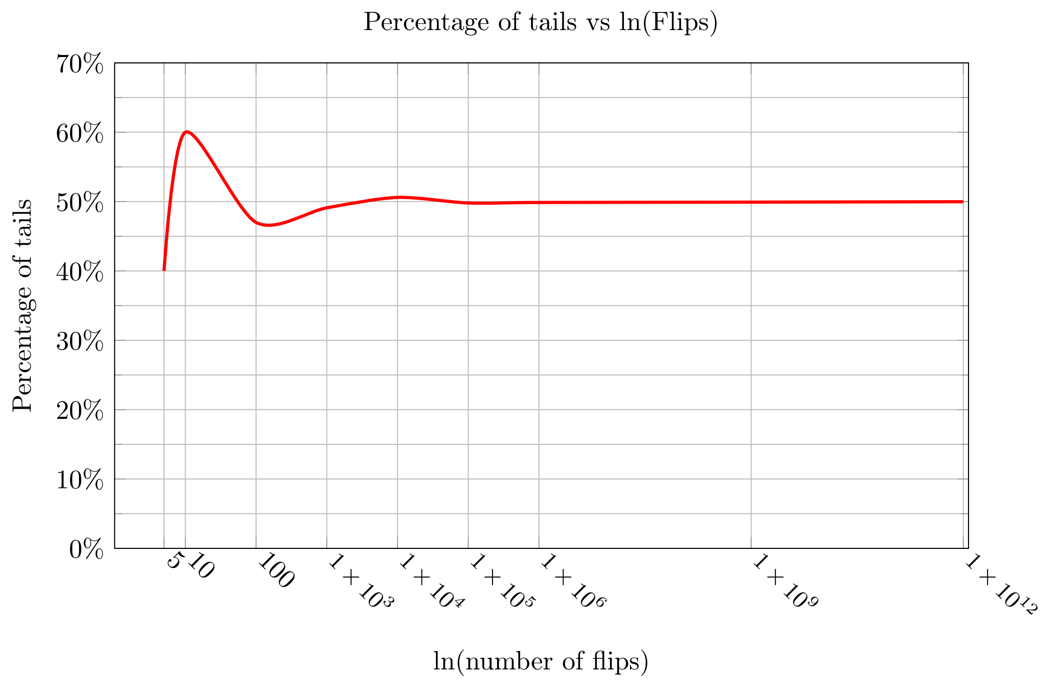
2.5 The Law of Large Numbers

The law of large numbers says that if you repeat an experiment many times, then the average result tends to get closer to the theoretical average. As an example of this, suppose that we have a fair coin. We expect that the probability of flipping the coin and landing on tails is 50%, i.e. $P(T) = \frac{1}{2}$. Our interpretation of probability says that we expect there to be about 50% tails in the long run.

To see this in action, suppose that we start flipping a coin over and over again. Using a random number generator, below is an example of the proportion of tails in the sample of coin flips:

- 5 flips: HHTTH. $T = 2$, $H = 3$, so tails = 40%.
- 10 flips: HHTTHTHTTT. $T = 6$, $H = 4$, so tails = 60%.
- 100 flips: $T = 47$, $H = 53$, so tails = 47%.
- 1000 flips: $T = 491$, $H = 509$, so tails = 49.1%.
- 10000 flips: $T = 5062$, $H = 4938$, so tails = 50.6%.
- 100000 flips: $T = 49872$, $H = 50128$, so tails = 49.8%.
- 1 million flips: tails = 49.87%.
- 1 billion flips: tails = 49.92%.
- 1 trillion flips: tails = 49.982%.

We can visualize this by plotting the graph (where here we take a logarithmic scale $\ln(\cdot)$ so that the coin flips fit on the same axis):



As you can see, in the beginning there is quite a lot of random fluctuation in the results. However, over time these random changes mostly cancel out, and the proportion of Tails converges to the theoretical average of 50%. This offers us one possible interpretation of probability:

Interpretation of probability

When we say $P(A) = x\%$, we mean that if we repeated the experiment a huge number of times, the proportion of outcomes corresponding to event A would get closer and closer to $x\%$.

2.6 Important definitions

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

(2) The *sample space* is the set of all possible outcomes (simple events).

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

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

(5) For equally likely outcomes, we compute probability by counting favourable outcomes and dividing by total outcomes.

To see these definitions in action, let's return to the toy cabinet of Example 2.3. In this case, the sample space will be

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

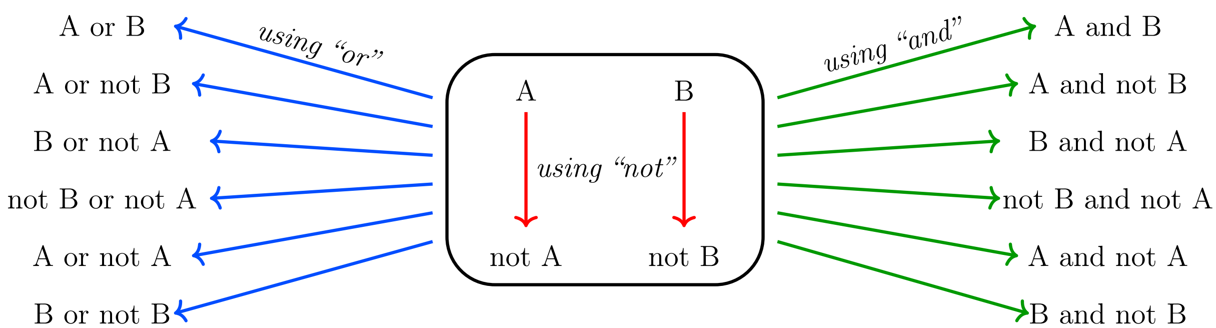
so the set S has size 58. In this example, an outcome would be the selection of an individual toy – something like "picking this toy *in particular*". Mathematically, this would be any subset of S that is of size one. A general event would be any subset of S at all, and they would represent more general statements such as"picking a toy from the fourth row" or"picking a blue car". Just to emphasize this: the event *is* the subset of the sample space. For example, in Section 2.3 we considered the event C

meaning "selecting a green toy", and the probability was determined to be $\frac{10}{58}$. The mathematical definition of the event C would be the set of all green toys in the cabinet.

3 Compound Probabilities

3.1 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 (which is the number of all possible outcomes). As a matter of fact, we can build many other events using three logical building blocks, namely and, or, and not:



These look very similar to the connectives of propositional logic that we saw in previous lectures. However, there are some differences: here we are *not* working with the precise world of 1's and 0's, instead we are trying to reason about the numbers we assign to uncertainty. Therefore, although the picture above looks very similar to the logical connectives of propositional logic, in probability theory these connectives do not behave in exactly the same way. As a matter of fact, in probability theory, it is more useful to translate these logical operations into set-theoretic operations.

3.2 Set-Theoretic Operations

The three operations of *not*, *and*, and *or* can be translated into set theory as follows.

(1) For any event A, the phrase "not A" means that the event in which A does not happen. In set language, the event "not A" is given by the *complement* of A, written A^c.

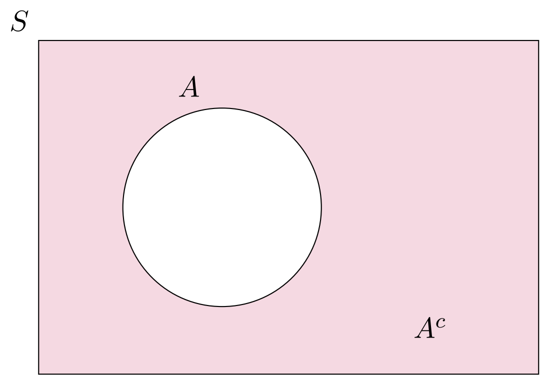
(2) For a pair of events A and B, the event "A and B" means the event in which both A and B happen. In set language, the event "A and B" is given by the *intersection* of A and B, written $A \cap B$.

(3) For a pair of events A and B, the event "A or B" means the event in which A happens, or B happens, or both. In set language, the event "A or B" is given by the *union* of A and B, written $A \cup B$.

We will now describe these set-theoretic operations one by one.

3.2.1 The Complement of a set

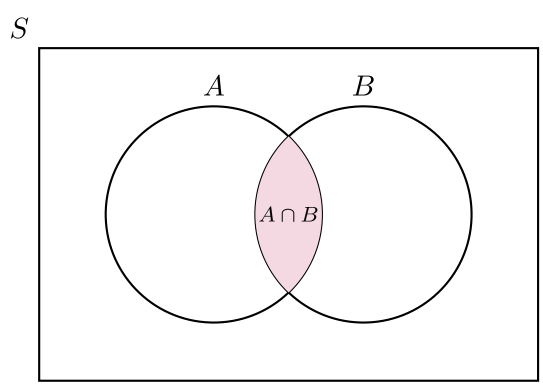
The complement A^c means everything in the sample space that is not in the subset A. In the picture below, the complement is the shaded area:



As an example, let $S = \{0, 1, 2, 3, \ldots\} = \mathbb{N}$ (the natural numbers). If we take the subset $A = \{\text{even numbers}\}$, then the complement of A will be all those numbers in S that are *not in* A. Put differently: the complement A^c consists of all those elements in S that are *missing* from A. Since A is all of the even numbers, the complement will be $A^c = \{\text{odd numbers}\}$.

3.2.2 The Intersection of two sets

For a pair of subsets A and B, their intersection $A \cap B$ is another subset whose members are those things that are in both A and B at the same time. The word "intersection" is commonly used for roads: an intersection is the place where two roads cross each other. In this sense, if we were to stand right on the intersection, then we would be standing on both roads at the same time. The set-theoretic intersection works the same way: the subset $A \cap B$ is like the *overlap* between two sets, shaded in the picture below:



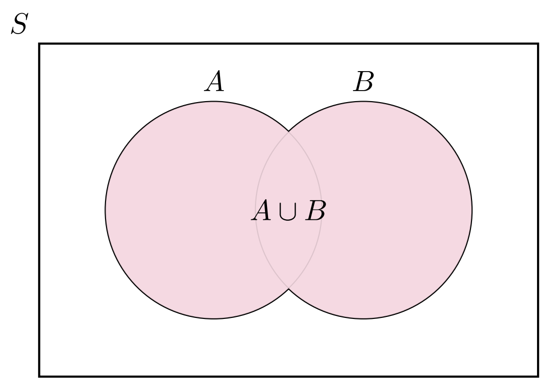
For example, let's again work with $S = \mathbb{N}$ and $A = \{\text{even numbers}\}$. Suppose now that B is another subset of S, namely $B = \{\text{numbers less than 10}\}$. Then the intersection of A and B will be the set of numbers that are in A and B *at the same time*. Since A is the set of all the even numbers and B

is the set of all numbers less than ten, the intersection $A \cap B$ will be the set of even numbers that are also less than 10. Written out:

$$A \cap B = \{0, 2, 4, 6, 8\}.$$

3.2.3 The Union of two Sets

The word "union" often means "things joined together", for instance the European Union is a political institution that consists of some European countries joined together. In set theory, the idea is similar: for a pair of subsets A and B, the union $A \cup B$ is a subset consisting of all those things that are either in A, or B, or both. Pictorially, we can imagine it as covering both sets A and B:



As an example of this, let's keep things simple and suppose that $A = \{1, 2, 3\}$ and $B = \{3, 4, 5, 6\}$. Then the union $A \cup B$ will be the set that is formed by taking all the members of both A and B and putting them into one big set:

$$A \cup B = \{1, 2, 3, 4, 5, 6\}.$$

Note that we don't write 3 twice – it only gets counted once.

3.2.4 Exercise: working with sets

Let $S = \mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}$, the natural numbers. Consider two subsets: $A = \{\text{even numbers}\}$ and $B = \{\text{numbers less than 10}\}$.

Exercise 2

Compute the following.

(1) B^c (the complement of B).

(2) $A \cap B$.

(3) $A^c \cap B$.

Solution

(1)

$$B^c = \{\text{numbers greater than or equal to } 10\}.$$

(2)

$$A \cap B = \{0, 2, 4, 6, 8\}.$$

(3) A^c is the odd numbers, so

$$A^c \cap B = \{1, 3, 5, 7, 9\}.$$

3.3 The Probability $P(\textbf{not } A)$

For any event A, the event A and the event A^c cover the whole sample space and do not overlap. In set language, $A \cup A^c = S$ and $A \cap A^c = \varnothing$. So their probabilities add to 1:

$$P(A) + P(A^c) = 1.$$

Therefore:

$$P(A^c) = 1 - P(A).$$

3.3.1 Exercise: the complement of FFF

From the earlier true/false experiment, $P(\text{FFF}) = \frac{1}{8}$.

Let E be the event "FFF happens". Then E^c means "not FFF", meaning any of the other 7 outcomes.

Exercise 3

Compute $P(E^c)$.

Solution

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

3.4 The Probability $P(A \textbf{ and } B)$

In our setting of "classical probability", the probability $P(A \text{ and } B)$ should be determined by counting the size of the intersection and dividing it by the size of the sample space, i.e.

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

In most simple situations, we can calculate the probability $P(A \text{ and } B)$ simply by counting. For example, if we had a fair die and we let event A be "we roll an even number" and we let event B be "we roll a number less than 4", then we can calculate $P(A \text{ and } B)$ by simply counting the favourable outcomes explicitly: there is only one number between 1 and 6 that is less than 4 and even at the same time: it is the number 2. Therefore, the subset $A \cap B = \{2\}$ has size equal to 1 and the associated probability will be $P(A \text{ and } B) = \frac{1}{6}$.

The example above works because we had a very small sample space of size 6. However, we can imagine that in more complicated situations it becomes very difficult to list out and keep track of all of the possibilities in order to count the set $A \cap B$. In practice, it is often helpful to use some formulas that relate the probability $P(A \text{ and } B)$ to the probabilities $P(A)$ and $P(B)$. Assuming that we knew the size of A and B, we can always calculate the probabilities of $P(A)$ and $P(B)$ using counting formulas similar to the above. However, the size of the subsets A and B cannot determine the size of their intersection. Therefore, calculating the probability $P(A \text{ and } B)$ is slightly nuanced, and depends heavily on how the two events are related to each other.

3.4.1 Conditional Probability

One method for figuring out the size of the intersection $A \cap B$ would be to look inside the set B and count up all of the instances of A that we find. In terms of probability, this would be like trying to consider the probability of A happening assuming that the sample space has been restricted to the subset B instead of the whole of S. In turn, this means that you would be trying to evaluate the possibility that A happens, given that you have learned that B will happen. In fact, this idea is known as a *conditional probability*, and it is written $P(A|B)$. We read this as "the probability of A, assuming B occurs".

Going back to the example of the die: the event A corresponds to rolling an even number, and the event B corresponds to rolling a number less than 4. Imagine that before you roll the die, God whispered in your ear "the outcome will be less than 4". This means that, in that moment, you would realise that it's not possible to roll a $4, 5$ or 6. In this situation, if you were to evaluate your chances of event A happening, you would cut your space of possibilities from $\{1, 2, 3, 4, 5, 6\}$ down to the smaller set $\{1, 2, 3\}$, which happens to equal B. So, when you evaluate your chances, you take the count

$$P(A|B) = \frac{\text{size of } A \cap B}{\text{size of } B}.$$

The subsets corresponding to A and B are $\{2, 4, 6\}$ and $\{1, 2, 3\}$, respectively. This means that the intersection $A \cap B = \{2\}$, which has size 1. Since B has size 3, our conditional probability becomes $P(A|B) = \frac{1}{3}$.

Now, we make a useful observation. In this example we have seen that $P(B) = \frac{3}{6} = \frac{1}{2}$, $P(A \text{ and } B) = \frac{1}{6}$ and $P(A|B) = \frac{1}{3}$. It just so happens that $\frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$. In symbols, we have just discovered an example of a general formula:

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

So, in order to compute the probability of the compound event "A and B", we can always take a product involving conditional probabilities. (Equivalently, $P(A \text{ and } B) = P(A) \cdot P(B \mid A)$.)

3.4.2 Probabilistic Independence

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

Independent vs Dependent Events

Let A and B be two events. We say that:

- A and B are **independent** if learning A does not give you any information about whether B happens, and
- A and B are **dependent** if learning A happened *does* give you some information about whether B happens.

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

In the previous example of rolling the die, the events of A being "rolling an even number" and B being "rolling a number less than four" are dependent events, because information about one affects knowledge about the possibility of the other. However, we can also imagine other situations where knowing one piece of information doesn't give us anything reason to change our assessment of probability. For example, let the event C be "we roll a multiple of 3". Then as a subset, $C = \{3, 6\}$. If, instead, God now whispered in your ear "the outcome will be a multiple of 3", then in that moment you know to restrict your sample space from $\{1, 2, 3, 4, 5, 6\}$ down to $\{3, 6\}$. Given this new restricted set consists of one odd and one even number, 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 doesn't actually change the amount of certainty that you have about the system. Therefore, we see that the events A and C are independent.

According to this line of reasoning, we see that a pair of events are independent if knowing the extra information doesn't actually change the uncertainty that you have. In other words, two events are independent if the conditional probability coincides with the unconditional probability, i.e. events A and B are independent if $P(A|B) = P(A)$ and $P(B|A) = P(B)$. This reduces the formula for $P(A \text{ and } B)$ to the following.

Multiplication Rule for Independent Events

If A and B are independent events, then:

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

Example: Let's consider a fair coin that is flipped two times. Let event A be "first flip is tails" and let event B be "second flip is tails". These are independent events: if you were to flip the first coin, you get no information about the outcome of the next coin flip. So, we may apply the independent events formula to conclude that:

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

Equivalently, the sample space for two flips is {HH, HT, TH, TT}, and only TT is a favourable outcome, so the probability is $\frac{1}{4}$, which agrees with the above calculation.

3.4.3 Example: independent events (two decks)

Suppose I take two separate decks of cards and shuffle each one. Let $A =$ "the top card of deck 1 is an ace". Let $B =$ "the top card of deck 2 is an ace".

These events are independent, and each has probability $\frac{4}{52} = \frac{1}{13}$. So

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

3.4.4 Example: dependent events (two aces from one deck)

Suppose I shuffle one deck of 52 cards. Let $A =$ "the first card is an ace". Let $B =$ "the second card is an ace".

These events are dependent because if the first card is an ace, there are fewer aces left in the deck.

Compute:

$$P(A) = \frac{4}{52} = \frac{1}{13}, \qquad P(B \mid A) = \frac{3}{51}.$$

So:

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

3.4.5 Summary of the multiplication rules

Step 1: Decide whether events are independent.
(Assuming B is not impossible.) If $P(A) = P(A \mid B)$, then they are independent.

Step 2: Use the correct multiplication rule.
If independent: $P(A \text{ and } B) = P(A) \cdot P(B)$.
If not independent: $P(A \text{ and } B) = P(A) \cdot P(B \mid A)$.

3.5 The Probability $P(A$ or $B)$

According to Section 3.2, the compound event "A or B" corresponds to the set-theoretic union $A \cup B$. Thus, in order to calculate the probability $P(A \text{ or } B)$, we can count up the size of this subset and divide it by the size of the sample space:

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

As with the intersection discussed earlier, generally we cannot deduce what the size of the union $A \cup B$ will be based on the sizes of A and B – we also need to have some information about how they overlap. Fortunately, we have just derived a way to count the overlap, so we can actually relate the size of the union $A \cup B$ to the sizes of A, B and the intersection $A \cap B$. The formula is:

$$\text{size of } A \cup B = (\text{size of } A) + (\text{size of } B) - (\text{size of } A \cap B).$$

The formula works because we are trying to add up everything in both sets A and B. Therefore it feels quite reasonable to simply add up the size of A and the size of B, and declare that this is the size of the union $A \cup B$. However, if we were to do that, then we risk double-counting the intersection $A \cap B$. To see this for yourself: consider the two sets $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$. These are both of size 3, and the sum of their sizes is $3 + 3 = 6$. However, the union $A \cup B$, being the collection of all members of both sets, is the set $\{1, 2, 3, 4\}$. This has size 4, not 6. When we make the calculation $3 + 3 = 6$, we are actually counting the elements 2 and 3 *twice* by accident. But, if we remove the extra copy of $\{2, 3\}$ we end up with $6 - 2 = 4$, which is the correct size of the union. We can use this observation to derive a general formula for $P(A\text{or } B)$ in terms of $P(A)$, $P(B)$ and $P(A \text{ and } B)$:

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

Sometimes, it may be the case that the events A and B have subsets that don't overlap at all. In this case, the intersection $A \cap B$ is empty (i.e. size zero) and therefore there is no risk of double-counting anything. Whenever this happens, we call the events A and B *mutually exclusive*. Since the subsets of A and B correspond to the set of outcomes which realise the events A and B, mutually exclusive events *cannot happen* at the same time, i.e. $P(A \text{ and } B) = 0$.

If A and B are mutually exclusive, then the formula for $P(A \text{ or } B)$ reduces to:

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

3.5.1 Example: ace, spade, king (top card)

Suppose I shuffle a deck of cards. Let A = "the top card is an ace". Let B = "the top card is a spade". Let C = "the top card is a king".

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

Exercise 4

Compute the following.

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

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

Solution

(1)

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

So

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

(2) A and C are mutually exclusive, so

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

4 Exercises

4.1 Rolling a die

Suppose we roll a fair six-sided die. Let $A =$ "we roll an even number". Let $B =$ "we roll a number ≥ 4".

Exercise 5

Compute the following.

(1) $P(A)$.

(2) $P(B)$.

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

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

Solution

(1) $A = \{2, 4, 6\}$, so

$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

(2) $B = \{4, 5, 6\}$, so

$$P(B) = \frac{3}{6} = \frac{1}{2}.$$

(3) A and B means "even and ≥ 4", so $\{4, 6\}$, hence

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

(4) A or B means $\{2, 4, 5, 6\}$, so

$$P(A \text{ or } B) = \frac{4}{6} = \frac{2}{3}.$$