Supporting Australian Mathematics Project

A guide for teachers – Years 11 and 12

Probability and statistics: Module 18

Probability

Education Services Australia

AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Probability – A guide for teachers (Years 11–12)

Professor Ian Gordon, University of Melbourne
Associate Professor Ray Watson, University of Melbourne

Editor: Dr Jane Pitkethly, La Trobe University

Illustrations and web design: Catherine Tan, Michael Shaw

Full bibliographic details are available from Education Services Australia.

Published by Education Services Australia
PO Box 177
Carlton South Vic 3053
Australia

Tel: (03) 9207 9600
Fax: (03) 9910 9800
Email: info@esa.edu.au
Website: www.esa.edu.au

This publication is funded by the Australian Government Department of Education, Employment and Workplace Relations.

Supporting Australian Mathematics Project

Australian Mathematical Sciences Institute
Building 161
The University of Melbourne
VIC 3010
Email: enquiries@amsi.org.au
Website: www.amsi.org.au

Probability

Assumed knowledge

- An informal understanding of probability, as covered by the series of TIMES modules *Chance* (Years 1–10).
- Some familiarity with sets, set notation and basic operations on sets, as covered by the TIMES module *Sets and Venn diagrams* (Years 7–8).

Motivation

Chance in everyday life

Ideas of chance are pervasive in everyday life, and the use of chance and risk models makes an important impact on many human activities and concerns.

- When tossing a fair coin, what is the chance that the coin lands heads?
- What is the probability that I win the lottery?
- When playing Tetris, what are the chances of three long sticks (piece `I`) in a row?
- What chance do I have of getting into my preferred tertiary course?
- What chance do I have of getting into my preferred tertiary course if I get a score of at least 87 at the end of Year 12?
- What is the probability that the Reserve Bank will cut interest rates by half a percent?
- What are the chances of getting a sexually transmitted disease from unprotected sex?
- If I leave my school bag at the bus stop briefly to run home and get something I've forgotten, what are the chances of it still being there when I get back?
- If I leave my Crumpler bag at the bus stop briefly to run home and get something I've forgotten, what are the chances of it still being there when I get back?
- What is the probability of my bus being more than 15 minutes late today?
- What is the chance that my football team will win the premiership this year?
- What is the chance that Tom/Laura will like me?

There are important distinctions between some of these questions. As we shall see in this module, there are several different approaches to probability.

In this module we primarily deal with *idealised models* of probability based on concepts of symmetry and random mixing. These are simplified representations of concrete, physical reality. When we use a coin toss as an example, we are assuming that the coin is perfectly symmetric and the toss is sufficiently chaotic to guarantee equal probabilities for both outcomes (head or tail). From time to time in the module, we will hint at how such idealised models are applied in scientific enquiry. But that is really a topic for later discussion, in the modules on sampling and inference (*Random sampling*, *Inference for proportions*, *Inference for means*).

A second way of obtaining probabilities is as an extrapolation from a relative frequency. This really involves inference, but it is one way that probability and chance ideas are used in everyday life. For example, when the Bureau of Meteorology predicts that the chance of rain tomorrow is 20%, there is no clear, simple procedure involving random mixing as in the coin toss. Rather, there is a long history of conditions like those leading up to tomorrow. So the reasoning is, approximately: 'Among all of the previous instances of weather conditions similar to those we are experiencing now, there was some rain on the following day in 20% of instances.' Hence, by a form of inductive reasoning, we say that the probability of rain tomorrow is 20%.

These first two approaches can converge in some cases. Even if we are unconvinced by propositions about the randomness of the coin toss and the symmetry of the coin, we may have observed that the proportion of heads obtained from many similar tosses is close to 0.5, and therefore assert that the probability of obtaining a head is one half.

A third method of obtaining probabilities is subjective. This use is generally more casual. I may say that there is a 90% chance that Laura will like me, but you may disagree. Betting sometimes involves subjective probabilities on the part of the person making the bet. We do not consider subjective probability further in this module.

Chance in the curriculum

Ideas about chance or probability receive extensive coverage in all of the previous years of the Australian Curriculum. In this module we develop these ideas more formally; this is the stage in the curriculum where a structured framework for probability is presented.

Even in the very early years of the curriculum, there is coverage of basic ideas such as the words we use to describe degrees of certainty — words such as 'might', 'is likely' and 'definitely'. Students learn how to associate these words with events in their own experience, such as 'Our teacher *might* be sick today' or 'I will *definitely* sleep tonight'.

As these ideas are developed further, students will gain an understanding of chance or probability in relation to events: phenomena that we can observe. They learn that probability is measured on a scale from zero to one, and that impossible events have a probability of zero, while events that are certain have a probability of one.

The treatment in this module follows the path of conventional probability theory. The standard axioms are introduced, the important properties of probability are derived and illustrated, and the key ideas of independence, mutually exclusive events and conditional probabilities are covered. These are all foundational ideas for the other modules on probability and statistics.

Content

How to think about probability

The approach to the study of chance in earlier years has been quite intuitive. This is a reasonable approach which is sufficient for many purposes; many people who have not studied probability in detail manage to think about chance events quite correctly. For example, some professional gamblers know about odds and can manage risk in a sensible way, even making an income, without necessarily having a strong mathematical background.

However, this intuitive approach only goes so far. And this is the point in the curriculum at which the consideration of probability is formalised. In order to do this, we start by thinking in a fundamental way about probability statements.

The Bureau of Meteorology may say, for example, that the probability of rain tomorrow is 20%, or 0.20 on a scale of 0 to 1. We are immediately reminded that a probability is a number between 0 and 1 (inclusive), or 0% and 100% (inclusive) if the percentage scale is used.

These are hardly different scales, they simply provide two different ways to express the same number. Nonetheless, because both usages are common, it is important to recognise in a given context whether probabilities are being expressed as numbers between 0 and 1, or percentages between 0% and 100%. There is potential for confusion, especially where very small probabilities are concerned.

So if a probability is a number, how is it obtained, or defined? A probability does not come from an ordinary mathematical operation in the usual sense of functions, such as logarithms and the trigonometric functions. When we speak in a natural way about chance and probability, we refer to something that may happen (an **event**), and we assign a number to the probability of the event.

This sounds a bit like a functional statement, which is an important insight. When we write $f(x) = 2$, we say 'f *of* x equals 2'. By analogy, we say 'the probability *of* rain tomorrow equals 0.2'.

From this perspective, we can see that the domain of probability statements consists of events: things that may happen. But — to state the obvious — there are many things that may happen, and not only that, there is a plethora of types of events in a wide variety of contexts. Ordinary mathematical functions — at any rate, the ones students have met so far — tend to have simple numerical domains such as $\mathbb{R}$ or $\mathbb{R}^+$.

We need a different mathematical structure to deal with probability, which we discuss in this module.

Random procedures

We want to talk about the probabilities of events: things that may happen. This sounds impossibly broad and unwieldy, and it is. We therefore usually limit ourselves to a particular context in which possible outcomes are well defined and can be specified, at least in principle, beforehand. We also need to specify the procedure that is to be carried out, that will produce one of these outcomes.

We call such a phenomenon a **random procedure** or, alternatively, a **random process**. It need not be an experiment; all that is required is that the process for obtaining the outcome is well defined, and the possible outcomes are able to be specified. Why is the word 'random' used? Its purpose is to indicate that the process is inherently uncertain: before we make the observation, we do not know which of the possible outcomes will occur, but we do know what is on the list of possible outcomes. Here are some examples:

1. A standard 20 cent coin is tossed: a rapid spin is delivered by the thumb to the coin as it sits on the index finger; when it lands on the palm of the catching hand and is then flipped onto the back of the other hand, we observe the uppermost face. The possible outcomes are H (head) or T (tail).
2. A fair six-sided die is shaken vigorously in a cup and rolled onto a table; the number of spots on the top side is recorded. The possible outcomes are $1,2,3,4,5,6$.
3. A lottery such as Powerball is observed and the number of the 'Powerball' is recorded. The possible outcomes for Powerball are $1,2,3,\dots,45$.

We start with these simple examples because probability statements associated with them have an obvious intuitive meaning. For example, it is commonly accepted that the chance of obtaining a head when you toss a coin equals $\frac{1}{2}$, or 50%. Indeed, in many sports, a coin toss is used to determine who makes an initial choice that may involve an advantage, such as which end to kick to in football, whether to serve or receive in tennis, and whether to bat or bowl in cricket.

Similarly, when dice are used in board games, it is usually implicitly assumed that each possible outcome is equally likely, and therefore has probability $\frac{1}{6}$.

In commercial lotteries such as Tattslotto and Powerball, it is a regulatory requirement that each outcome is equally likely.

But for the moment, we are just considering these examples as random procedures; we will come to the issue of probability later in this module. And we will see that there is more to it than these simple intuitions.

Here are some examples that are more complex:

1. In a game of Tetris, a sequence of three consecutive pieces is observed.
2. Five babies born in 1995 are followed up over their lives, and major health and milestone events are recorded. This really involves several random procedures; a specific one is considered later in this section.
3. 273 students attempt Mathematical Methods in 2015. In June 2017, for each student, it is recorded whether or not he or she is a tertiary student.
4. A new mobile phone is tested, and the time until its battery needs recharging is observed.

Example: Tetris

In Tetris there are seven distinct shapes made up of four squares. In serious Tetris discussions they are given letter labels which roughly correspond to their shapes. These are `I` for the long stick, `T` for the T-shape, and so on; the others are `J`, `L`, `O`, `S` and `Z`.

What are the possible sequences of three consecutive pieces that we may observe?

We can list them alphabetically:

`III`, `IIJ`, `IIL`, `IIO`, `IIS`, `IIT`, `IIZ`, `IJI`, `IJJ`, …, `ZZZ`.

There is a large number of possible sequences. How many?

Each position in the sequence can be any of the seven possible shapes, so the number of possible sequences is

$$7 \times 7 \times 7 = 7^3 = 343.$$

The previous example illustrates the **multiplication rule**, which has been covered informally in earlier years. This rule tells us that, if a sequence of k separate processes are considered, and the first can be done in n_1 ways, the second in n_2 ways, and so on, up to the kth process having n_k ways of being carried out, then the number of possible ways that the k processes can be carried out successively is $n_1 \times n_2 \times \cdots \times n_k$.

Example: Five people born in 1995

For five people born in 1995, we may record a number of different health, illness and injury phenomena, and usually a particular random procedure will focus on one of these. We might consider the specific injury status: 'experienced a broken leg at least once prior to reaching age 20'. What are the possible outcomes for this random procedure?

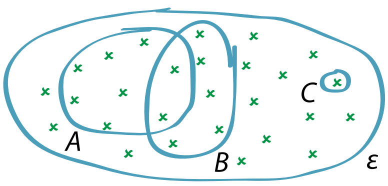
Label the individuals with the letters A to E (the first five letters of the alphabet). We use the symbol A_0 to indicate that individual A does not experience a broken leg prior to reaching age 20, and A_1 to indicate that individual A does have this experience, and similarly for the other four individuals. One possible outcome is $A_0B_0C_0D_0E_0$. In this outcome, none of the individuals has had a broken leg prior to turning 20. Each individual either does or does not have at least one break, so the total number of possible outcomes is $2^5 = 32$.

Events and event spaces

An understanding of sets is crucial for dealing with events and the basic rules of probability: this includes set notation, the subset relationship, and operations on sets such as union and intersection.

The **event space** $\mathscr{E}$ is the set of the possible distinct outcomes of the random process. The event space is sometimes called the **sample space**. When rolling a normal six-sided die and recording the uppermost face, the event space is $\mathscr{E} = \{1,2,3,4,5,6\}$.

An **event** is a collection of possible outcomes, and therefore it is a subset of $\mathscr{E}$. The event A occurs if the observed outcome is in A, and does not occur if the observed outcome is not in A. It is usual to denote an event by a capital letter, commonly near the start of the alphabet: $A, B, C, \ldots$.



The event space.

Events may be expressed in words, inside quotation marks:

- In the die-rolling example, we can consider the event A = "a prime number is obtained" = $\{2, 3, 5\}$.
- In the example of five people born in 1995, we can consider the event B = "exactly two of the individuals experience at least one broken leg prior to turning 20".

When expressing an event in words like this, the description must determine a subset of $\mathcal{E}$. We can regard something as an event only when it is true that, for every possible outcome, we can say definitively whether or not the event has occurred.

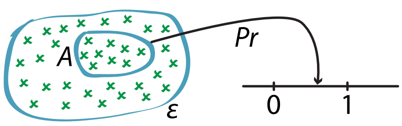
An outcome is sometimes called an **elementary event**.

The following are events in any context:

- The event space $\mathcal{E}$ is itself an event, because it is a collection of outcomes. We can think of $\mathcal{E}$ as the event "the random procedure occurs".
- The empty set $\varnothing$ is a subset of $\mathcal{E}$. The event $\varnothing$ is not very interesting; it contains no outcomes.

The mathematical approach to probability requires that we have to be able to specify the possible outcomes. Outcomes that we may be able to imagine, such as the die coming to rest on a corner (leading to no number being uppermost), are often excluded from consideration altogether. If we do want to consider an outcome as possible, it must be included in the event space.

Recall that the domain of probability statements consists of events. Now that we understand that events are subsets of the event space $\mathcal{E}$, we see that probability is a set function. It maps subsets of $\mathcal{E}$ into the interval $[0, 1]$.



Probability as a set function.

Operations on events and relations between events

Events in random procedures are sets of possible outcomes. We now apply ideas about sets to events; this will help us to think clearly about the events and, subsequently, their probabilities.

For a subset A of the event space $\mathscr{E}$, the complementary set is

$$A' = \{x \in \mathscr{E} : x \notin A\}.$$

In words, A' is the set of all elements of the universal set $\mathscr{E}$ that are *not* in A.

For the event A, the **complementary event** is A'. For example:

- In the die-rolling example, if A = "a prime number is obtained", then $A' = \{1,4,6\}$.
- In the example of five people born in 1995, if B = "exactly two of the individuals experience at least one broken leg prior to turning 20", then B' = "the number of individuals experiencing at least one broken leg prior to turning 20 is 0, 1, 3, 4 or 5".

We are interested in relations between events, and therefore in considering more than one event.

For two events A and B, the **intersection** $A \cap B$ is the event "A and B both occur". In the die-rolling example, if A = "a prime number is obtained" and B = "an even number is obtained", then $A = \{2,3,5\}$ and $B = \{2,4,6\}$, so $A \cap B = \{2\}$. As is the case with sets generally, we can consider the intersection of more than two events: the event $A \cap B \cap C$ is the event "A and B and C all occur", and so on.

For two events A and B, the **union** $A \cup B$ is the event "either A or B occurs, or both". Equivalently, the union $A \cup B$ is the event "at least one of A and B occurs". For the die-rolling example again, with $A = \{2,3,5\}$ and $B = \{2,4,6\}$, the union is $A \cup B = \{2,3,4,5,6\}$. We may be interested in the union of more than two events, such as $A \cup B \cup C$.

We say that an event A is a **subset** of an event B, and write $A \subseteq B$, when all outcomes in A are also in B. For example, suppose two dice are rolled. If A = "the sum of the two numbers obtained is 12" and B = "a six is obtained on the first die", then $A = \{(6,6)\}$ and $B = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$, so $A \subseteq B$.

We say that two events A and B are **mutually exclusive** if their intersection is the empty set, which means they have no outcomes in common. For example, an event A and its complementary event A' are mutually exclusive, since A' consists of all outcomes that are not in A. In the example of five people born in 1995, if we define X to be the number of individuals experiencing a broken leg prior to turning 20, then the event "$X = 1$" and the event "$X = 0$" are mutually exclusive: they cannot both occur.

Exercise 1

Consider rolling two normal dice and observing the uppermost faces obtained.

a For this random procedure, which of the following are events?

- **i** A = "the sum of the two numbers observed is a prime number"
- **ii** B = "Barry sneezes just before the dice are rolled"
- **iii** C = "at least one of the two dice shows an odd number"
- **iv** D = "at least one of the two dice shows a six"
- **v** E = "the outcome $(6,6)$ is observed"

b Which outcomes are in the event A?

c How many outcomes are in D?

d Describe $A \cap E$ in words.

Exercise 2

Observations are made by the Bureau of Meteorology on a city's weather every day, in a systematic fashion. Suppose that we define a random procedure to be: record the daily amounts of rainfall, to the nearest tenth of a millimetre, in August 2017.

a For this random procedure, which of the following are events?

- **i** A = "the only days with recorded rainfall greater than 0.0 mm are Mondays"
- **ii** B = "the main street floods on Tuesday 15 August"
- **iii** C = "the total rainfall for the month is greater than 20.0 mm"

b Consider the following events:

- D = "the rainfall amount recorded on 31 August is greater than the total recorded for the first 30 days"
- E = "no rainfall is recorded before the last seven days of the month"
- F = "the rainfall amount on Wednesday 16 August is 18.7 mm".

Which of the following pairs of events are mutually exclusive?

i A and E **ii** A and F **iii** D and E **iv** E and F

c Describe the following events in words:

i A' **ii** $A \cap E$ **iii** $A \cup C$.

The probability axioms

We now consider probabilities of events. What values can a probability take, and what rules govern probabilities?

Some notation: We have used capital letters $A, B, \dots$ for events, and we have observed that probability can be thought of as a function that has an event as its argument. The domain is the collection of all events, that is, all possible subsets of $\mathscr{E}$. By analogy with usual function notation, such as $f(x)$, we use the notation $\Pr(A)$ to denoted the probability of the event A.

There are just three basic rules or axioms for probabilities, from which several other important rules can be derived. These fundamental rules were first spelled out in a formal way by the Russian mathematician Andrey Kolmogorov.

The three axioms of probability

1. $\Pr(A) \geq 0$, for each event A.
2. $\Pr(\mathscr{E}) = 1$.
3. If events A and B are mutually exclusive, that is, if $A \cap B = \varnothing$, then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B).$$

The first axiom fits with our experience of measuring things, like length and area. Just as the lowest possible value for a length is zero, the lowest possible value for a probability is zero. Any other choice for the minimum numerical value of a probability would not work.

The second axiom says that the probability that something will happen is one. Another way of thinking about the second axiom is this: When the random process is carried out, we are certain that one of the outcomes in the event space $\mathscr{E}$ will occur.

The third axiom determines the way we work out probabilities of mutually exclusive events. The axiom says that, if A and B are mutually exclusive, then the probability that at least one of them occurs is the sum of the two individual probabilities. While this seems very compelling, it cannot be proved; mathematically, it must be assumed.

Taken together, the three axioms imply that probabilities must be between zero and one (inclusive): $0 \leq \Pr(A) \leq 1$, for any event A. We will prove this in the next section.

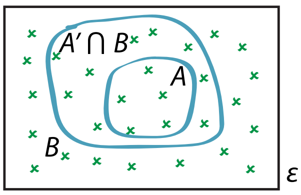
The choice of this numerical range for probabilities fits with relative frequencies from data, which are always in this range. In earlier years, students learned how relative frequencies are estimates of probabilities. We may estimate the probability of being left-handed by finding the relative frequency, or proportion, of left-handed people in a random sample of people. Any relative frequency of this sort must be a fraction between zero and one; it cannot be negative, and it cannot be greater than one. This basic observation fits with the range for probabilities themselves: $0 \leq \Pr(A) \leq 1$.

The probability scale is between zero and one, but in ordinary discourse about probabilities the percentage scale is often used. In the media, in particular, we may read that 'the chance of the government being re-elected is regarded as no better than 40%'. In such usage, the scale for probability is 0% to 100%. There is no real difficulty with this, provided it is very clear which scale is being used. This becomes particularly important for small probabilities: if we say that the chance of an outcome is 0.5%, this is the same as saying that the probability (on the usual zero-to-one scale) is 0.005. It is important to be alert to the potential confusion here.

Technical note. In this module, we only consider examples where the event space $\mathcal{E}$ is finite or countably infinite. The more general axiomatic treatment of probability, which also covers examples where $\mathcal{E}$ is uncountable, is very similar to our treatment here. But in the general situation, we do not insist that *every* subset of $\mathcal{E}$ is an event (and therefore has an associated probability).

Useful properties of probability

We now look at some of the consequences of the three axioms of probability given in the previous section.



Event A is a subset of event B.

Property 1

If the event A is a subset of the event B, then $\Pr(A) \le \Pr(B)$. That is,

$$A \subseteq B \implies \Pr(A) \le \Pr(B).$$

Proof

Consider events A and B such that $A \subseteq B$, as shown in the diagram above. Then $B = A \cup (A' \cap B)$, and the events A and $A' \cap B$ are mutually exclusive. Thus, by the third axiom, $\Pr(B) = \Pr(A) + \Pr(A' \cap B)$. The first axiom tells us that $\Pr(A' \cap B) \ge 0$. Hence, $\Pr(A) \le \Pr(B)$. $\square$

This property has a useful application. We write $A \Rightarrow B$ to mean that, if A occurs, then B occurs. So $A \Rightarrow B$ is the same as $A \subseteq B$, which implies $\Pr(A) \leq \Pr(B)$. For example:

- Indiana Jones can only find the treasure (A)
- if he first solves the puzzle of the seven serpents (B);

so $A \subseteq B$ and therefore $\Pr(A) \leq \Pr(B)$; his chance of finding the treasure is at most equal to his chance of solving the puzzle.

While property 1 is important, it is rather obvious.

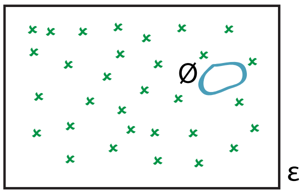
Property 2

$0 \leq \Pr(A) \leq 1$, *for each event* A.

Proof

The first axiom says that $\Pr(A) \geq 0$. It is true for every event A that $A \subseteq \mathscr{E}$. Hence $\Pr(A) \leq \Pr(\mathscr{E})$, by property 1. But $\Pr(\mathscr{E}) = 1$, by the second axiom, and so it follows that $\Pr(A) \leq 1$. □

This property formalises the scale for probabilities, given the axioms.



The empty event.

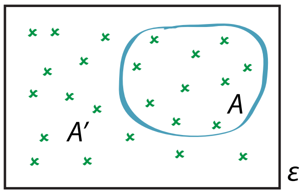
Property 3

$\Pr(\varnothing) = 0$.

Proof

Choose any event A. Then the two events A and the empty set $\varnothing$ are mutually exclusive, so $\Pr(A \cup \varnothing) = \Pr(A) + \Pr(\varnothing)$, by the third axiom. Since $A \cup \varnothing = A$, this gives $\Pr(A) = \Pr(A) + \Pr(\varnothing)$. Hence, $\Pr(\varnothing) = 0$. □

It would be rather strange if the probability of the empty set was anything other than zero, so it is reassuring to confirm that this is not so: $\Pr(\varnothing) = 0$, as expected.



An event A and its complementary event A'.

Property 4

$\Pr(A') = 1 - \Pr(A)$, *for each event* A.

Proof

The events A and A' are mutually exclusive, so $\Pr(A \cup A') = \Pr(A) + \Pr(A')$, by the third axiom. But $A \cup A' = \mathscr{E}$, and therefore $\Pr(A') = \Pr(\mathscr{E}) - \Pr(A)$. According to the second axiom, $\Pr(\mathscr{E}) = 1$. Hence, $\Pr(A') = 1 - \Pr(A)$. $\square$

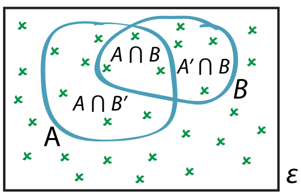
This property is surprisingly useful and is applied frequently. It is most effective when the probability of the event of interest is difficult to calculate directly, but the probability of the complementary event is known or easily calculated.

Exercise 3

Suppose that, in a four-child family, the probability of all four children being boys is 0.07. What is the probability that a four-child family contains at least one girl?

Property 5 (Addition theorem)

$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$, *for all events* A *and* B.



	B	B'
A	$A \cap B$	$A \cap B'$
A'	$A' \cap B$	$A' \cap B'$

Two representations of the addition theorem.

Both the diagram and the table above give representations of the event space $\mathscr{E}$ in terms of the two events A and B and their complements.

The impression we get from the diagram, from the table, or just using basic logic is that, if we add together the probability of A and the probability of B, then $A \cap B$ is included in both events, so it has been counted twice. Hence, to obtain the probability of at least one of A and B occurring, that is, to obtain $\Pr(A \cup B)$, we need to subtract $\Pr(A \cap B)$ from

$\Pr(A) + \Pr(B)$. This justifies the formula given by the addition theorem (property 5). We can demonstrate the result more formally, as follows.

Proof

To devise a formal proof of the addition theorem, we construct mutually exclusive events in a helpful way, and use the third axiom of probability.

We can express the union of A and B as $A \cup B = A \cup (A' \cap B)$, and the events A and $A' \cap B$ are mutually exclusive. Also, $B = (A \cap B) \cup (A' \cap B)$, and the two events $A \cap B$ and $A' \cap B$ are mutually exclusive. By applying the third axiom to the first of these relationships, we get that

$$\Pr(A \cup B) = \Pr(A) + \Pr(A' \cap B).$$

And using the third axiom for the second relationship, we find that

$$\Pr(B) = \Pr(A \cap B) + \Pr(A' \cap B)$$

$$\implies \Pr(A' \cap B) = \Pr(B) - \Pr(A \cap B).$$

Hence, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$, as required. $\square$

Exercise 4

Lego sells ‘minifigures’. There are 16 distinct minifigures, and they are distributed at random among shops (it is said). Furthermore, the minifigures cannot be identified prior to purchase and removal of the packaging. A young child wishes to obtain the ‘Pirate Captain’, one of the 16 figures in Series 8. She persuades her parent to buy a minifigure at two different shops. Define the events:

- A = “Pirate Captain is purchased at the first shop”
- B = “Pirate Captain is purchased at the second shop”.

Assume that the distribution of the minifigures is random across shops, and assume that $\Pr(A \cap B) = \frac{1}{256}$.

a What is $\Pr(A)$? $\Pr(B)$?

b What is the probability of the child’s wishes being satisfied?

c Describe each of the following events in words:

i B' **ii** $A' \cup B'$ **iii** $A \cap B'$ **iv** $A' \cap B'$.

d Which event is more probable: $A \cup B$ or $A' \cap B'$?

Assigning probabilities

So far we have said little about how probabilities can be assigned numerical values in a random procedure. We know that probabilities must be in the interval $[0,1]$, but how do we know what the actual numerical value is for the probability of a particular event?

Symmetry and random mixing

Ideas for the assignment of probabilities have been progressively developed in earlier years of the curriculum. One important approach is based on **symmetry**. The outcome of rolling a fair die is one of the most common examples of this. Even small children will readily accept the notion that the probability of obtaining a three, when rolling a fair die, is $\frac{1}{6}$. Where does this idea come from, and what assumptions are involved?

A classic 'fair die' is a close approximation to a uniform cube. (The word 'approximation' is used here because many dice have slightly rounded corners and edges.) A cube, by definition, has six equal faces, all of which are squares. So if we roll the die and observe the uppermost face when it has come to rest, there are six possible outcomes. The symmetry of the cube suggests there is no reason to think that any of the outcomes is more or less likely than any other. So we assign a probability of one sixth to each of the possible outcomes. Here $\mathscr{E} = \{1,2,3,4,5,6\}$ and $\Pr(i) = \frac{1}{6}$, for each $i = 1,2,\dots,6$.

What assumptions have we made here? One important aspect of this process is the rolling of the die. Suppose my technique for 'rolling' the die is to pick it up and turn it over to the opposite face. This cannot produce all six outcomes. Furthermore, the two outcomes that are produced by this process occur in a non-random systematic sequence.

So we observe that, with randomising devices such as coins, dice and cards, there is an important assumption about **random mixing** involved. When playing games like *Snakes and Ladders* that use dice, we are familiar with players shaking the cup extra vigorously in the interests of a truly random outcome, and on the other hand, those trying to slide the die from the cup in order to produce a six!

Exercise 5

Consider each of the following procedures. How random do you think the mixing is likely to be?

a A normal coin flip that lands on your hand.

b A normal coin flip that lands on the floor.

c A die shaken vigorously inside a cup 'sealed' at the top by your palm, then rolled onto the table.

d An ‘overhand’ shuffle of cards; this is the shuffle that we usually first learn.

e A ‘riffle’ shuffle of cards: in this shuffle, two piles of cards are rapidly and randomly interwoven by gradual release from the two thumbs.

f The blast of air blown into the sphere containing the balls in Powerball.

g The random number generator that produces the next Tetris shape.

h The hat used for drawing raffles at the footy club.

Link

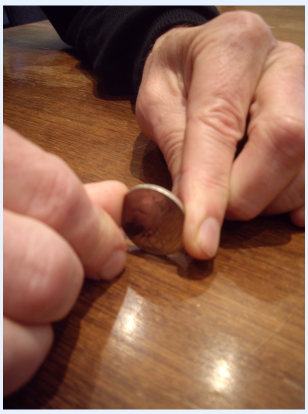
For a guide to various card shuffles (including the overhand and riffle shuffles), see `www.pokerology.com/poker-articles/how-to-shuffle-cards/`

There are also more subtle aspects of symmetry to be considered. In the case of a conventional die, the *shape* may be cubical and hence symmetrical. But is that enough? We also need uniform density of the material of the cube, or at least an appropriately symmetric distribution of mass (some dice are hollow, for example). If a wooden die is constructed with a layer of lead hidden under one face, this would violate the assumption that we make in assigning equal probabilities to all six outcomes.

What about coin tossing and symmetry? We may assume that the coin itself is essentially a cylinder of negligible thickness made from uniform density metal with two perfectly flat sides. No actual coin is like this. The faces of coins have sculpted shapes on them that produce the images we see. These shapes are clearly not actually symmetric. How much does this matter?

Exercise 6

Use an ordinary 20 cent coin to carry out the following experiment. Find the flattest surface you can. Starting with the coin on its edge, spin the coin fast; usually this is most effectively done by delivering the spin from the index finger of one hand on one side of the coin against the thumb of the other hand on the other side, as shown in the photo.



Starting position for non-standard spin.

The spin needs to be uninterrupted by other objects and stay on the flat surface throughout. There is a bit of skill required to get the coin to spin rapidly; practise until you can get it to spin for about ten seconds before coming to rest.

Once you are adept at this technique, on a given spin you can save time near the end of the spin: when it is clear what the result will be, you can stop the motion and record it. Do this 30 times. How many heads and tails do you get?

When we use symmetry to assign probabilities, it is important that the outcomes we consider are truly symmetric, and not just superficially so. It is definitely not enough to know merely the number of possible outcomes of a random process. Just because there are only k possible outcomes does not imply that each one of them has probability equal to $\frac{1}{k}$; there needs to be a basis for assuming symmetry. Often a list of possible 'outcomes' is really a list of events, rather than elementary events. In that case, even if the elementary events are equiprobable, there is no need for the derived events to be so.

Exercise 7

A standard deck of cards is well shuffled and from it I am dealt a hand of four cards. In my hand I can have 0, 1, 2, 3 or 4 aces: five possibilities altogether. So the chance that my hand has four aces is equal to $\frac{1}{5} = 0.2$.

What is wrong with this argument?

Relative frequencies

A second way to assign numerical values for probabilities of events is by using **relative frequencies** from data. In fact, this is probably the most common method. Strictly speaking, what we are doing is *estimating* probabilities rather than assigning the true numerical values.

For example, we may have data from 302 incidents in which school children left their bag at the bus stop briefly to run home and get something that they forgot. In 38 of these incidents, the bag was not there when they returned. So we might estimate the probability of losing a school bag as

$$\frac{38}{302} \approx 0.126,$$

or 12.6% on the percentage scale.

Equally probable outcomes

There are many contexts in which it is desirable to ensure equal probabilities of outcomes. These include commercial games of chance, such as lotteries and games at casinos, and the allocation of treatments to patients in randomised controlled trials. Randomising devices are used to achieve this.

Example: Election ballots (California)

It is understood that there is a 'positional bias' in the voting behaviour of undecided voters in elections; that is, some of them tend to give their voting preference in the order that the candidates are listed on the ballot paper.

For this reason, in California, about three months before an election, the Secretary of State produces a random ordering of the letters of the alphabet. This is used to define the order in which the names of candidates are printed on the ballot papers. What's more, it is applied not just to the first letter of each candidate's name, but to the other letters in the names also.

Example: Powerball

The commercial lottery Powerball has operated in Australia since 1996. There are 45 balls that can appear as the Powerball. A probability model based on symmetry and random mixing implies that the chance of any particular ball appearing as the Powerball is $\frac{1}{45} \approx 0.022$. (We are rounding to three decimal places here.) This is the idealised model that is intended on grounds of fairness for those who play the game, and we may believe that the model applies, to a very good approximation. There are laws governing lotteries that aim to ensure this.

At the time of writing, there have been 853 Powerball draws. The following table shows the relative frequencies of some of the 45 numbers from the 853 draws.

Relative frequencies for the Powerball

Powerball	1	2	3	4	5	…	43	44	45
Relative frequency	$\frac{21}{853}$	$\frac{14}{853}$	$\frac{20}{853}$	$\frac{21}{853}$	$\frac{15}{853}$	…	$\frac{19}{853}$	$\frac{22}{853}$	$\frac{22}{853}$
	0.025	0.016	0.023	0.025	0.018	…	0.022	0.026	0.026

We may check assumptions about randomness by looking at the relative frequencies. We may ask: Are they acceptably close to the probability implied by a fair draw, which is 0.022? Obviously, if we are to do this properly, we must look at the whole distribution, and not just the part shown here.

The previous example is related to testing in statistical inference, in which a model for a random process is proposed and we examine data to see whether it is consistent with the model. This is not part of the topic of probability directly, but it indicates one way in which probability models are used in practice.

Exercise 8

During the Vietnam War, the Prime Minister Robert Menzies introduced conscription to national army service for young men. However, not all eligible men were conscripted: there was a random process involved. Two birthday ballots a year were held to determine who would be called up. Marbles numbered from 1 to 366 were used, corresponding to each possible birthdate. The marbles were placed in a barrel and a predetermined number were drawn individually and randomly by hand. There were two ballots per year, the first for birthdates from 1 January to 30 June and the second for birthdates from 1 July to 31 December. In a given half year, if a birthdate was drawn, all men turning 20 on that date in that year were required by law to present themselves for national service. This occurred from 1965 until 1972, when Gough Whitlam's ALP government abolished the scheme.

Were men born on 29 February more likely to be conscripted, or less likely? Or were their chances the same as other men? What assumptions have you made?

Links

- For a photo of the marbles, see `vrroom.naa.gov.au/print/?ID=19537`
- For the selected birthdates, see `www.awm.gov.au/encyclopedia/viet_app/`

A previous example described the method used to ensure electoral fairness in California. The next example concerns an issue of electoral fairness in Australia.

Example: Election ballots (Australian Senate, 1975)

The 1975 Senate election in Australia occurred in a heightened political climate. The Australian Labor Party (ALP) government, voted into power in 1972 and returned in 1974, was losing popularity. In November 1975, the government was dismissed by the Governor-General, Sir John Kerr, and an election was called which involved a 'double dissolution' of both houses of parliament. Between the dismissal on 11 November and the elections on 13 December, there were many protests, demonstrations and hot political debates. In the election for the Senate, the two main political parties were the ALP and the Liberal–National coalition; between them, they received 80% of the votes cast.

The following table gives some data from the draws for positions on the ballot papers in this Senate election. The table shows the positions of the Liberal–National coalition

(L/N) and the ALP for each of the six states and two territories. Also shown in each case is the number of groups participating in the election. In WA and Tasmania, the Liberal Party and the National Party were separate; elsewhere they had a joint Senate team.

Ballot papers for the 1975 Senate election

State or territory	Position of		Number of groups
	L/N	ALP	
NSW	2	8	10
Vic	1	6	8
Qld	2	6	7
SA	1	3	9
WA	7,1	10	11
Tas	1,2	5	6
ACT	1	2	4
NT	1	3	3

The positions on each ballot paper were determined by drawing envelopes out of a box. Does the result of this appear to you to be appropriately random?

Among those who noted the strangeness of this distribution were two astute Melbourne statisticians, Alison Harcourt and Malcolm Clark. Their analysis of this result formed the basis of a submission to the Joint Select Committee on Electoral Reform in 1983, and this eventually led to a change in the Commonwealth Electoral Act. The process now involves much more thorough random mixing, using 'double randomisation' and a process similar to that used in Tattslotto draws.

Reference

R. M. Clark and A. G. Harcourt, 'Randomisation and the 1975 Senate ballot draw', *The Australian Journal of Statistics* 33 (1991), 261–278.

Conditional probability

Like many other basic ideas of probability, we have an intuitive sense of the meaning and application of conditional probability from everyday usage, and students have seen this in previous years.

Suppose the chance that I leave my keys at home and arrive at work without them is 0.01. Now suppose I was up until 2 a.m. the night before, had to hurry in the morning, and used a different backpack from my usual one. What then is the chance that I don't have my keys? What we are *given* — new information about what has occurred — can change the probability of the event of interest. Intuitively, the more difficult circumstances of the morning described will increase the probability of forgetting my keys.

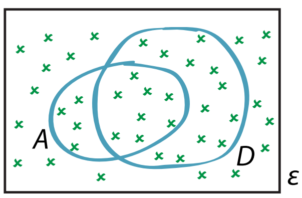
Consider the probability of obtaining 2 when rolling a fair die. We know that the probability of this outcome is $\frac{1}{6}$. Suppose we are told that an even number has been obtained. Given this information, what is the probability of obtaining 2? There are three possible even-number outcomes: 2, 4 or 6. These are equiprobable. One of these three outcomes is 2. So intuition tells us that the probability of obtaining 2, given that an even number has been obtained, is $\frac{1}{3}$. Similar reasoning leads to the following: The probability of obtaining 2, given that an odd number has been obtained, equals zero. If we know that an odd number has been obtained, then obtaining 2 is impossible, and hence has conditional probability equal to zero.

Using these examples as background, we now consider conditional probability formally. When we have an event A in a random process with event space $\mathcal{E}$, we have used the notation $\Pr(A)$ for the probability of A. As the examples above show, we may need to change the probability of A if we are given new information that some other event D has occurred. Sometimes the conditional probability may be larger, sometimes it may be smaller, or the probability may be unchanged: it depends on the relationship between the events A and D.

We use the notation $\Pr(A|D)$ to denote 'the probability of A given D'. We seek a way of finding this conditional probability.

It is worth making the observation that, in a sense, all probabilities are conditional: the condition being (at least) that the random process has occurred. That is, it is reasonable to think of $\Pr(A)$ as $\Pr(A|\mathcal{E})$.

It helps to think about conditional probability using a suitable diagram, as follows.



Illustrating conditional probability.

If the event D has occurred, then the only elementary events in A that can possibly have occurred are those in $A \cap D$. The greater (or smaller) the probability of $A \cap D$, the greater (or smaller) the value of $\Pr(A|D)$. Logic dictates that $\Pr(A|D)$ should be proportional to $\Pr(A \cap D)$. That is,

$$\Pr(A|D) = c \times \Pr(A \cap D),$$

for some constant c determined by D.

We are now in a new situation, where we only need to consider the intersection of events with D; it is as if we have a new, reduced event space. We still want to consider the probabilities of events, but they are all revised to be conditional on D, that is, given that D has occurred.

The usual properties of probability should hold for $\Pr(\cdot|D)$. In particular, the second axiom of probability tells us that we must have $\Pr(\mathscr{E}|D) = 1$. Hence

$$1 = \Pr(\mathscr{E}|D) = c\Pr(\mathscr{E} \cap D) = c\Pr(D),$$

which implies that

$$c = \frac{1}{\Pr(D)}.$$

We have obtained the **rule for conditional probability**: the probability of the event A given the event D is

$$\Pr(A|D) = \frac{\Pr(A \cap D)}{\Pr(D)}.$$

Since $\Pr(D)$ is the denominator in this formula, we are in trouble if $\Pr(D) = 0$; we need to exclude this. However, in the situations we consider, if $\Pr(D) = 0$, then D cannot occur; so in that circumstance it is going to be meaningless to speak of any probability, given that D has occurred.

We can rearrange this equation to get another useful relationship, sometimes known as the **multiplication theorem**:

$$\Pr(A \cap D) = \Pr(D) \times \Pr(A|D).$$

We can interchange A and D and also obtain

$$\Pr(A \cap D) = \Pr(A) \times \Pr(D|A).$$

This is now concerned with the probability conditional on A.

Independence

The concept of independence of events plays an important role throughout probability and statistics. The most helpful way to think about independent events is as follows.

We first consider two events A and B, and assume that $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$. The events A and B are **independent** if

$$\Pr(A|B) = \Pr(A).$$

This equation says that the conditional probability of A, given B, is the same as the unconditional probability of A. In other words, given that we know that B has occurred, the probability of A is unaffected.

Using the rule for conditional probability, we see that the events A and B are independent if and only if

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B).$$

This equation gives us a useful alternative characterisation of independence in the case of two events. The symmetry of this equation shows that independence is not a directional relationship: the events A and B are independent if $\Pr(B|A) = \Pr(B)$. So, for independent events A and B, whether B occurs has no effect on the probability that A occurs, and similarly A has no effect on the probability that B occurs.

If events A and B are not independent, we say that they are **dependent**. This does not necessarily mean that they are directly causally related; it just means that they are not independent in the formal sense defined here.

Events that are physically independent are also independent in the mathematical sense. Mind you, physical independence is quite a subtle matter (keeping in mind such phenomena as the 'butterfly effect' in chaos theory, the notion that a butterfly flapping its wings in Brazil could lead to a tornado in Texas). But we often take physical independence as the working model for events separated by sufficient time and/or space. We may also assume that events are independent based on our consideration of the independence of the random processes involved in the events. Very commonly, we assume that observations made on different individuals in a study are independent, because the individuals themselves are separate and unrelated.

For example, if A = "my train is late" and B = "the clue for 1 across in today's cryptic crossword involves an anagram", we would usually regard A and B as independent. However, if C = "it is raining this morning", then it may well be that A and C are not independent, while B and C are independent.

While physical independence implies mathematical independence, the converse is not true. Events that are part of the same random procedure and not obviously physically independent may turn out to obey the defining relationship for mathematical independence, as the following example demonstrates.

Example

Suppose a fair die is rolled twice. Consider the following two events:

- A = "a three is obtained on the second roll"
- B = "the sum of the two numbers obtained is less than or equal to 4".

In this case, somewhat surprisingly, we can show that $\Pr(B|A) = \Pr(B)$. Thus A and B are mathematically independent, even though at face value they seem related.

To see this, we write the elementary events in this random process as (x, y), with x the result of the first roll and y the result of the second roll. Then:

- $A = \{(1,3),(2,3),(3,3),(4,3),(5,3),(6,3)\}$
- $B = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)\}$.

There are 36 elementary events, each with probability $\frac{1}{36}$, by symmetry. So $\Pr(A) = \frac{6}{36} = \frac{1}{6}$ and $\Pr(B) = \frac{6}{36} = \frac{1}{6}$. We also have $A \cap B = \{(1,3)\}$ and so $\Pr(A \cap B) = \frac{1}{36}$. Hence,

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} = \Pr(B).$$

Thus A and B are independent. Alternatively, we can check independence by calculating $\Pr(A \cap B) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = \Pr(A) \times \Pr(B)$.

Such examples are rare in practice and somewhat artificial. The key point is that physical independence implies mathematical independence.

Exercise 9

For each of the following situations, discuss whether or not the events A and B should be regarded as independent.

a A = "Powerball number is 23 this week"
B = "Powerball number is 23 next week"

b A = "Powerball number is 23 this week"
B = "Powerball number is 24 this week"

c A = "maximum temperature in my location is at least 33 °C today"
B = "maximum temperature in my location is at least 33 °C six months from today"

d A = "the tenth coin toss in a sequence of tosses of a fair coin is a head"
B = "the first nine tosses in the sequence each result in a head"

The previous exercise illustrates some important points. It is usually assumed that games of chance involving gambling entail independence. Some people may suspect that these games (such as lotteries and roulette) are rigged, and there are indeed some famous examples of lottery scandals. But the regulations of such games are designed to create an environment of randomness that involves independence between different instances of the same game. So knowing the Powerball outcome this week should not change the probability distribution for next week; each ball should be equally likely as usual.

On the other hand, the events "Powerball is 23 this week" and "Powerball is 24 this week" are incompatible: on any single Powerball draw there can only be one Powerball number. So these events are mutually exclusive (their intersection is empty) and the probability that they both occur is zero.

Mutually exclusive events are definitely *not* independent; they are dependent. If two events are mutually exclusive, then given that one occurs, the conditional probability of the other event is zero. That is, for mutually exclusive events C and D (both having non-zero probability), we have $\Pr(C|D) = 0 \neq \Pr(C)$. Alternatively, we may observe that $\Pr(C \cap D) = 0 \neq \Pr(C)\Pr(D)$.

Students often confuse the two concepts, but *mutually exclusive* and *independent* are quite different. The following table serves as a reminder.

Mutually exclusive vs independent

Mutually exclusive events	Independent events
If one occurs, the other cannot.	Knowing that one occurs does not affect the probability of the other occurring.
$A \cap B = \varnothing$ and so $\Pr(A \cap B) = 0$	$\Pr(A \cap B) = \Pr(A)\Pr(B)$

What about the maximum-temperature example from exercise 9? This is a great deal more subtle. It might seem that two days six months apart, even at the same location, are sufficiently distant in time for the maximum temperatures to be independent. But for Adelaide (say), a maximum temperature of at least 33 °C could be a fairly clear indication that the day is not in winter. This would make the day six months from now not in summer, which might alter the probability of that day having a maximum temperature of at least 33 °C, compared to not knowing that today's temperature is at least 33 °C.

And the sequence of coin tosses? This rather depends on what we assume about the coin and the tossing mechanism. As indicated in exercises 5 and 6, there is more than one way to toss a coin. Even for the standard method — delivering the spin by a sudden flick of the thumb with the coin positioned on the index finger — successive tosses may not be truly independent if the coin tosser has learned to control the toss.

Just as we may use relative frequencies as estimates of probabilities in general, we may use conditional relative frequencies as estimates of conditional probabilities, and we may examine these to evaluate whether independence is suggested or not.

Recall the example of 302 incidents in which school children left their bag at the bus stop briefly to run home and get something. In 38 of these incidents, the bag was not there when they returned, and we used the relative frequency $\frac{38}{302} \approx 0.126$, or 12.6%, to estimate the chance that a bag will not be there in these circumstances.

In 29 of these incidents, the bag left at the bus stop was a Crumpler bag. Among these 29 incidents, the Crumpler bag was no longer at the bus stop upon return in 10 cases. So the relative frequency of the bag being gone, given that it was a Crumpler bag, was $\frac{10}{29} \approx 0.345$, or 34.5% on the percentage scale. The difference between the conditional relative frequency of 0.345 and the unconditional one of 0.126 suggests that the bag being gone upon return is *not* independent of the bag being a Crumpler bag.

We next prove an important property of independence.

Property

If events A and B are independent, then the events A and B' are independent.

Proof

Suppose that A and B are independent. Then $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$. Therefore

$$\begin{aligned}
\Pr(A \cap B') &= \Pr(A) - \Pr(A \cap B) \\
&= \Pr(A) - \Pr(A) \times \Pr(B) \qquad \text{since } A \text{ and } B \text{ are independent} \\
&= \Pr(A)\big(1 - \Pr(B)\big) \\
&= \Pr(A) \times \Pr(B').
\end{aligned}$$

Hence A and B' are independent. (It follows by symmetry that A' and B are independent, and therefore that A' and B' are independent.) $\square$

Exercise 10

Suppose that two separate experiments with fair dice are carried out. In the first experiment, a die is rolled once. If X is the outcome of the roll, then $Y = X + 1$ is recorded. In the second experiment, two dice are rolled. The sum U of the two outcomes is recorded, and the maximum V of the two outcomes is recorded.

a What are the possible values of Y, U and V?

b Which is more likely: $U > Y$ or $U < Y$?

c Which is more likely: $V > Y$ or $V < Y$?

Independence for more than two events

Our discussion of independence has so far been limited to two events. The extension to an arbitrary number of events is important.

If the events $A_1, A_2, \dots, A_n$ are mutually independent, then

$$\Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = \Pr(A_1)\Pr(A_2)\cdots\Pr(A_n). \qquad (*)$$

This is a necessary condition for mutual independence, but it is not sufficient. As we might reasonably expect, mutual independence is actually characterised by statements about conditional probabilities.

Essentially, the idea is the natural extension of the case of two events: We say that the events $A_1, A_2, \dots, A_n$ are **mutually independent** if, for each event A_i, all of the possible *conditional* probabilities involving the other events are equal to the unconditional probability $\Pr(A_i)$. Informally, this means that regardless of what happens among the other events, the probability of A_i is unchanged; and this must be true for each event A_i, where $i = 1, 2, \dots, n$.

We can express this definition formally as follows: Events $A_1, A_2, \dots, A_n$ are mutually independent if

$$\Pr(A_i) = \Pr(A_i \mid A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_m}),$$

for all i and for every possible combination $j_1, j_2, \dots, j_m$ such that $j_k \neq i$.

What happens, for example, when $n = 3$? This definition says that events A_1, A_2 and A_3 are mutually independent if

$$\Pr(A_1) = \Pr(A_1|A_2) = \Pr(A_1|A_3) = \Pr(A_1|A_2 \cap A_3)$$

and similarly for A_2 and A_3.

The equation $(*)$ above follows from this definition. The reason that $(*)$ is not sufficient to indicate mutual independence, however, is that it is possible for $(*)$ to be satisfied, while at the same time A_1 and A_2 are not independent of each other.

Example: Dice games

A Flemish gentleman called Chevalier de Méré played games of chance using dice in around 1650. He played one game where a fair die is rolled four times; it is assumed that the outcomes are independent. What is the chance of getting at least one six? De Méré reasoned as follows. On any single roll, the probability of getting a six is $\frac{1}{6}$. There are four rolls, so the probability of getting a six at some stage is $4 \times \frac{1}{6} = \frac{2}{3}$. The flaw in this reasoning should be obvious: What if there were seven rolls?

Another game he played was to roll two dice 24 times, and consider the chance of getting a double six at least once. He reasoned the same way for this game. On any single roll, the chance of a double six is $\frac{1}{36}$. There are 24 rolls, so the probability of getting a double six at some stage in the sequence of 24 rolls is $24 \times \frac{1}{36} = \frac{2}{3}$.

Because of the second calculation in particular, he was betting on this outcome occurring, when playing the game . . . and losing in the long run. Why? Rather than persisting doggedly with his strategy, he posed this question to Blaise Pascal, who correctly analysed his chances, as shown below. This was the beginning of the systematic study of probability theory.

In the second game, the chance of a double six on any single roll of the two dice is $\frac{1}{36}$, which we obtain by using the independence of the outcomes on the two dice:

$$\begin{aligned}\Pr(\text{double six}) &= \Pr(\text{six on first die}) \times \Pr(\text{six on second die})\\ &= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.\end{aligned}$$

To work out the probability of at least one double six in 24 rolls, we are going to apply a general rule which we have met in a number of contexts already.

The probability of an event occurring at least once in a sequence of n repetitions is equal to one minus the probability that it does not occur. If X is the number of times the event occurs, then

$$\Pr(X \geq 1) = 1 - \Pr(X = 0).$$

This rule is an application of property 4 (from the section *Useful properties of probability*), since "at least one" and "none" are complementary events. Note that the rule does not require that the events in the sequence of repetitions are independent (although when they are, the calculations are easier).

The application to de Méré's second game is this: On any given roll of the two dice, the probability of *not* obtaining a double six is $\frac{35}{36}$. If the 24 rolls are mutually independent, the probability that each of the 24 rolls does not result in a double six can be obtained as the product of the individual probabilities, using equation $(*)$, and is therefore equal to $\left(\frac{35}{36}\right)^{24} \approx 0.509$. This means that the probability of at least one roll resulting in a double six is approximately $1 - 0.509 = 0.491$. The fact that this probability is less than $\frac{1}{2}$ (and considerably less than the value $\frac{2}{3}$ calculated by de Méré) explains why he was losing money on his bets.

Exercise 11

a A boat offers tours to see dolphins in a partially enclosed bay. The probability of seeing dolphins on a trip is 0.7. Assuming independence between trips with regard to the sighting of dolphins, what is the probability of not seeing dolphins:

  i on two successive trips?

  ii on sevens trips in succession?

b A machine has four components which fail independently, with probabilities of failure 0.1, 0.01, 0.01 and 0.005. Calculate the probability of the machine failing if:

  i all components have to fail for the machine to fail

  ii any single component failing leads to the machine failing.

c Opening the building of an organisation in the morning of a working day is a responsibility shared between six people, each of whom has a key. The chances that they arrive at the building before the required time are, respectively, 0.95, 0.90, 0.80, 0.75, 0.50 and 0.10. Do you think it is reasonable to assume that their arrival times are mutually independent? Assuming they are, find the chance that the building is opened on time.

d In the assessment of the safety of nuclear reactors, calculations such as the following have been made.

> In any year, for one reactor, the chance of a large loss-of-coolant accident is estimated to be 3×10^{-4}. The probability of the failure of the required safety functions is 2×10^{-3}. Therefore the chance of reactor meltdown via this mode is 6×10^{-7}.

What do you think of this argument?

The next example illustrates the somewhat strange phenomenon of events that are *not* mutually independent but nevertheless satisfy equation $(*)$.

Example

Consider the random procedure of tossing a fair coin three times. Define the events:

- A = "at least two heads"
- B = "the last two tosses give the same result"
- C = "the first two results are heads or the last two results are tails".

Then, using an obvious notation:

- $A = \{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}, \mathrm{HHH}\}$
- $B = \{\mathrm{HHH}, \mathrm{THH}, \mathrm{HTT}, \mathrm{TTT}\}$
- $C = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTT}, \mathrm{TTT}\}$.

Thus $A \cap B \cap C = \{\mathrm{HHH}\}$. Assuming independence of the tosses, there are eight elementary outcomes (two for each of the three tosses), all equally probable, so the probability of each of them equals $\frac{1}{8}$. Hence

$$\Pr(A \cap B \cap C) = \tfrac{1}{8} = \left(\tfrac{1}{2}\right)^3 = \Pr(A)\Pr(B)\Pr(C).$$

So equation $(*)$ is satisfied. However, $B \cap C = \{\mathrm{HHH}, \mathrm{HTT}, \mathrm{TTT}\}$ and hence

$$\Pr(B \cap C) = \tfrac{3}{8} \neq \tfrac{1}{4} = \left(\tfrac{1}{2}\right)^2 = \Pr(B)\Pr(C).$$

It follows that B and C are not independent, and hence the events A, B and C are not mutually independent.

Examples like the previous one are not really of practical importance. Of much greater importance is the fact that, if events are physically independent, then they are mutually independent, and so then equation $(*)$ is true. This is the result that is useful in practice.

Tree diagrams

Probability problems sometimes involve more than one step or stage, and at the different stages various outcomes may occur. It can be tricky to keep track of the stages, outcomes and probabilities, and a 'tree diagram' is often helpful. Such a diagram is usually labelled in a sensible way with the outcomes that may occur at each stage and their probabilities. The probabilities of the outcomes at the ends of the branches of the tree may be found using the multiplication rule.

Example: Traffic lights

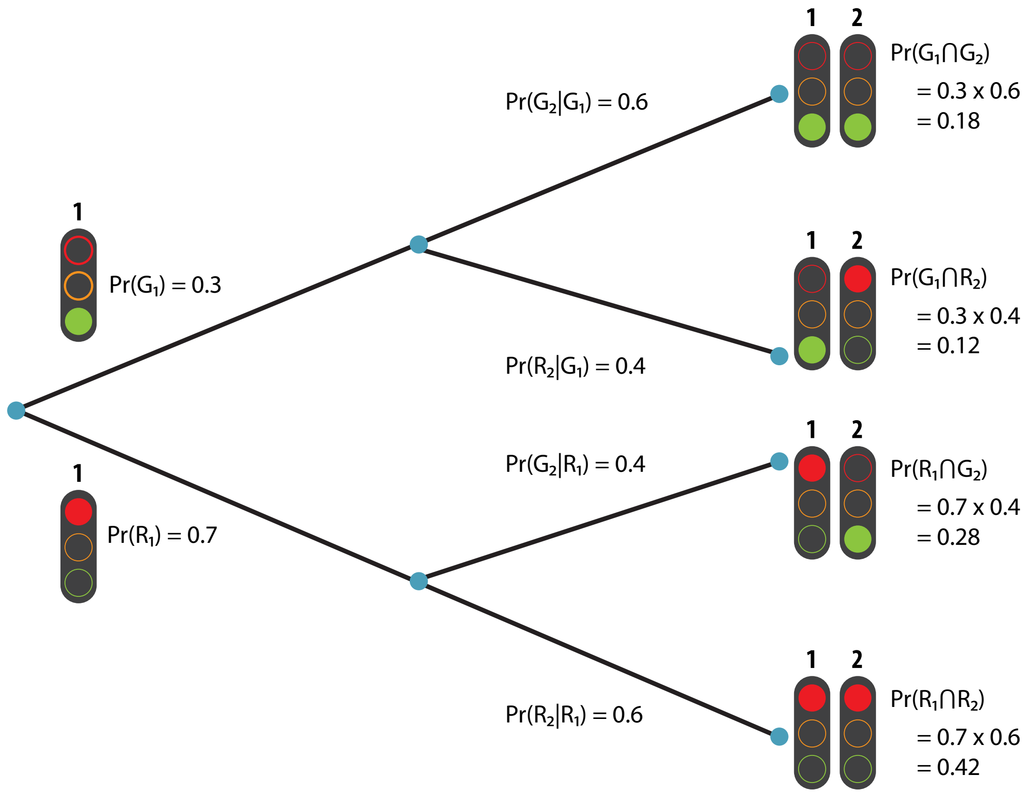
Rose rides a bicycle to work. There is a sequence of two sets of traffic lights on her route, not far apart. For simplicity, we ignore amber and assume the lights are only red or green. Suppose the probability that the first set is green when she arrives at it is 0.3. If she gets through on green at the first set, the probability that the second set is green when she arrives is 0.6; if she has to stop on red at the first set, the probability that the second set is green when she arrives is 0.4. What is the probability that she gets through both sets on green?

Define G_i to be the event that set i is green on her arrival, and similarly for R_i (red at set i). We need to find $\Pr(G_1 \cap G_2)$; note how the word 'both' in the question translates to a calculation of the probability of an intersection between two events. We use the multiplication theorem:

$$\Pr(G_1 \cap G_2) = \Pr(G_1)\Pr(G_2|G_1) = 0.3 \times 0.6 = 0.18.$$

In this calculation it is important to recognise the conditional probabilities we are given. 'If she gets through on green at the first set' is language that tells us that a conditional probability is being provided, the condition being the event G_1, that is, she meets a green light at the first set.

There are several things that can happen, as shown in the following tree diagram.



Tree diagram for the traffic-lights example.

With the probabilities at the first stage and the conditional probabilities at the second stage clearly labelled, it is a straightforward exercise to move along the branches of the tree, multiplying, to arrive at the probabilities of outcomes at the end of the branches; each case uses the multiplication theorem. The outcomes at the ends of the branches are intersections between events at the first and second stages.

What is the probability that the second set is green when she meets it? Note that the problem, as posed, does not specify this. We are only given conditional probabilities for the second set of lights. However, we can obtain the probability $\Pr(G_2)$ that she meets green at the second set by considering the two events $G_1 \cap G_2$ and $R_1 \cap G_2$. These are mutually exclusive, and $G_2 = (G_1 \cap G_2) \cup (R_1 \cap G_2)$. Hence $\Pr(G_2) = 0.18 + 0.28 = 0.46$.

Note that the events G_1 and G_2 are not independent, since

$$\Pr(G_1 \cap G_2) = 0.18 \qquad \text{and} \qquad \Pr(G_1)\Pr(G_2) = 0.3 \times 0.46 = 0.138,$$

and therefore $\Pr(G_1 \cap G_2) \neq \Pr(G_1)\Pr(G_2)$.

Exercise 12

A player pays \$4 to play the following game. A fair die is rolled once. If the outcome is not a prime number, the player loses. If a prime number is rolled (2, 3 or 5), the die is rolled again. If the outcome of the second roll is not a prime number, the player loses. If a prime number is rolled the second time, the player is paid an amount (in dollars) equal to the sum of the outcomes of the two rolls of the die.

Draw the applicable tree diagram and answer the following:

a What is the probability of losing at the first roll?

b What is the probability of winning \$6?

c What is the probability of winning?

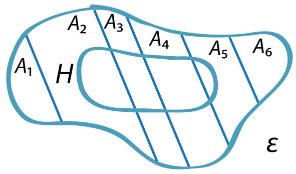
The law of total probability

We saw in the traffic-lights example from the previous section that it can be convenient to calculate the probability of an event in a somewhat indirect manner, by summing probabilities involving intersections. We now consider that process in more detail.

Events $A_1, A_2, \dots, A_k$ are said to be a **partition** of the event space if they are mutually exclusive and their union is the event space $\mathscr{E}$. That is, if events $A_1, A_2, \dots, A_k$ are such that

- $A_i \cap A_j = \varnothing$, for all $i \neq j$, and
- $A_1 \cup A_2 \cup \cdots \cup A_k = \mathscr{E}$,

then $A_1, A_2, \dots, A_k$ is said to be a partition. This means that exactly one of the events in the partition occurs: one and only one. This is illustrated in the following diagram.



A partition of the event space and the intersection of an event H with the partition.

The simplest version of a partition is any event A and its complement, since $A \cap A' = \varnothing$ and $A \cup A' = \mathscr{E}$.

Note the event H represented on the diagram of the event space $\mathscr{E}$ above. It appears from the diagram that the probability of H can be obtained by summing the probabilities of the intersection of H with each event A_i in the partition. We can show this formally:

$$
\begin{aligned}
H &= H \cap \mathscr{E} \\
&= H \cap (A_1 \cup A_2 \cup \cdots \cup A_k) \\
&= (H \cap A_1) \cup (H \cap A_2) \cup \cdots \cup (H \cap A_k).
\end{aligned}
$$

Since $A_1, A_2, \dots, A_k$ are mutually exclusive, it follows that the events

$$H \cap A_1,\ H \cap A_2,\ \dots,\ H \cap A_k$$

are also mutually exclusive. Hence, by the third axiom of probability,

$$
\begin{aligned}
\Pr(H) &= \Pr(H \cap A_1) + \Pr(H \cap A_2) + \cdots + \Pr(H \cap A_k) \\
&= \Pr(A_1)\Pr(H|A_1) + \Pr(A_2)\Pr(H|A_2) + \cdots + \Pr(A_k)\Pr(H|A_k) \\
&= \sum_{i=1}^{k} \Pr(A_i)\Pr(H|A_i).
\end{aligned}
$$

This result is known as the **law of total probability**. Note that it does not matter if there are some events A_j in the partition for which $H \cap A_j = \varnothing$. (For example, see A_1 and A_6 in the diagram above.) For these events, $\Pr(H \cap A_j) = 0$.

A table can provide a useful alternative way to represent the partition and the event H shown in the diagram above. In the following table, the event A_3 is used as an example.

	H	H'	
A_1			$\Pr(A_1)$
A_2			$\Pr(A_2)$
A_3	$\Pr(A_3 \cap H) = \Pr(A_3)\Pr(H \mid A_3)$		$\Pr(A_3)$
A_4			$\Pr(A_4)$
A_5			$\Pr(A_5)$
A_6			$\Pr(A_6)$
	$\Pr(H)$	$\Pr(H')$	

Example: Traffic lights, continued

We have seen an illustration of the law of total probability, in the traffic-lights example from the previous section. We obtained the probability of the rider meeting green at the second set of traffic lights by using the possible events at the first set of traffic lights, G_1 and R_1. These two events are a partition: note that $R_1 = G_1'$. Hence,

$$\begin{aligned}\Pr(G_2) &= \Pr(G_1 \cap G_2) + \Pr(R_1 \cap G_2)\\ &= \Pr(G_1)\Pr(G_2 \mid G_1) + \Pr(R_1)\Pr(G_2 \mid R_1)\\ &= (0.3 \times 0.6) + (0.7 \times 0.4)\\ &= 0.18 + 0.28\\ &= 0.46.\end{aligned}$$

In contexts with sequences of events, for which we use tree diagrams, we may be interested in the conditional probability of an event at the first stage, given what happened at the second stage. This is a reversal of what seems to be the natural order, but it is sometimes quite important. In the traffic-lights example, it would mean considering a conditional probability such as $\Pr(G_1 \mid G_2)$: the probability that the rider encounters green at the first set of traffic lights, given that she encounters green at the second set.

Again, consider the general situation involving a partition $A_1, A_2, \dots, A_k$ and an event H. The law of total probability enables us to find $\Pr(H)$ from the probabilities $\Pr(A_i)$ and $\Pr(H \mid A_i)$. However, we may be interested in $\Pr(A_i \mid H)$. This is found using the standard rule for conditional probability:

$$\Pr(A_i \mid H) = \frac{\Pr(A_i \cap H)}{\Pr(H)}.$$

We can use the multiplication theorem in the numerator and the law of total probability in the denominator to obtain

$$\Pr(A_i|H) = \frac{\Pr(A_i)\Pr(H|A_i)}{\sum_{j=1}^{k}\Pr(A_j)\Pr(H|A_j)}.$$

This process involves a 'reversal' of the conditional probabilities.

It is most important to avoid confusion between the two possible conditional probabilities: Are we thinking of $\Pr(C|D)$ or $\Pr(D|C)$? The language of any conditional probability statement must be examined closely to get this right.

These ideas are applied in diagnostic testing for diseases, as illustrated by the following example.

Example: Diagnostic testing

Suppose that a diagnostic screening test for breast cancer is used in a population of women in which 1% of women actually have breast cancer. Let C = "woman has breast cancer", so that $\Pr(C) = 0.01$. Suppose that the test finds that a woman has breast cancer, given that she actually does, in 85% of such cases. Let T_+ = "test is positive, that is, indicates cancer"; then $\Pr(T_+|C) = 0.85$. This quantity is known in diagnostic testing as the **sensitivity** of the test.

Suppose that when a woman actually does not have cancer, the test gives a negative result (indicating no cancer) in 93% of such cases; that is, $\Pr(T_-|C') = 0.93$. This is called the **specificity** of the test.

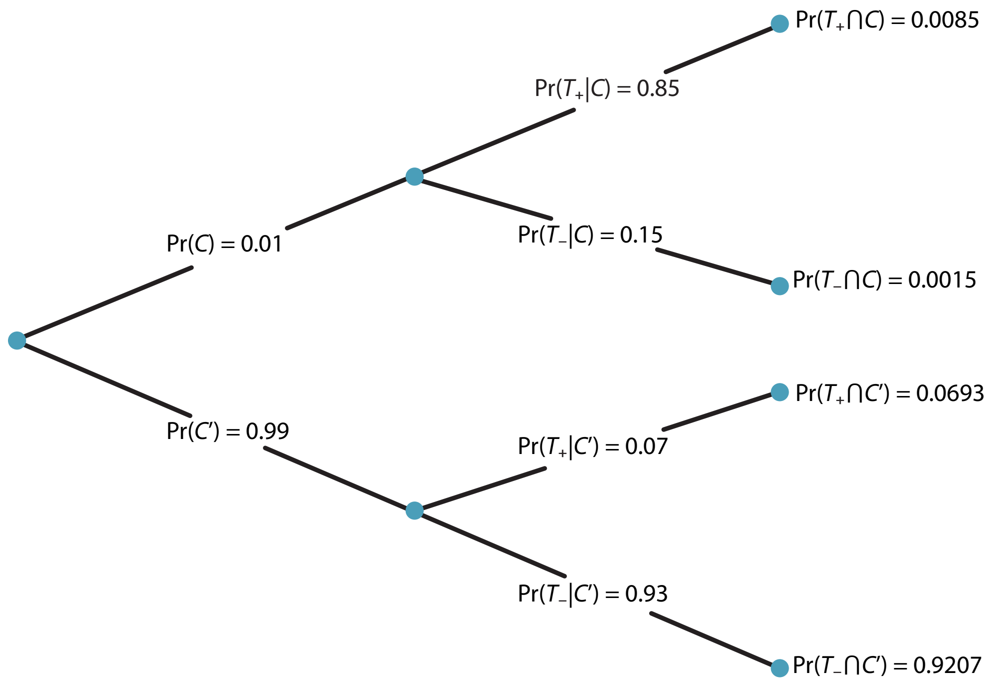
Clearly, we want both the sensitivity and the specificity to be as close as possible to 1.0. But — particularly for relatively cheap screening tests — this cannot always be achieved. (You may wonder how the sensitivity and specificity of such a diagnostic test can ever be determined, or estimated. This is done by using definitive tests of the presence of cancer, which may be much more invasive and costly than the non-definitive test under scrutiny. The non-definitive test can be applied in a large group of women whose cancer status is clear, one way or the other, to estimate the sensitivity and specificity.)

Before reading further, try to guess how likely it is that a woman (from this population) who tests positive on this test actually has cancer?

People's intuition on this question is often alarmingly awry. Notice that this is a question of genuine interest and concern in practice: the screening program will need to follow up with women who test positive, leading to more testing and also to obvious concern on the part of these women.

This crucial clinical question is: What is $\Pr(C|T_+)$?

This is a natural context for a tree diagram, which is shown below. This is because there are two stages. The first stage concerns whether or not the woman has cancer, and the second stage is the result of the test, positive (indicating cancer) or not. We find, for example, that $\Pr(T_+ \cap C) = \Pr(C)\Pr(T_+|C) = 0.01 \times 0.85 = 0.0085$.



Tree diagram for the diagnostic-testing example.

We can now apply the usual rule for conditional probability to find $\Pr(C|T_+)$:

$$\Pr(C|T_+) = \frac{\Pr(C \cap T_+)}{\Pr(T_+)} = \frac{0.0085}{\Pr(T_+)}.$$

To find $\Pr(T_+)$, we use the law of total probability:

$$\begin{aligned}\Pr(T_+) &= \Pr(C \cap T_+) + \Pr(C' \cap T_+)\\ &= \Pr(C)\Pr(T_+|C) + \Pr(C')\Pr(T_+|C')\\ &= (0.01 \times 0.85) + (0.99 \times 0.07)\\ &= 0.0085 + 0.0693\\ &= 0.0778.\end{aligned}$$

Hence

$$\Pr(C|T_+) = \frac{\Pr(C \cap T_+)}{\Pr(T_+)} = \frac{0.0085}{0.0778} \approx 0.1093.$$

Note what this implies: Among women with positive tests (indicating cancer), about 11% actually do have cancer. This is known as the **positive predictive value**. Many people are surprised by how low this probability is. Women who actually do have cancer and test positive are known as **true positives**. In this example, the overwhelming majority (about 89%) of women with a positive test do *not* have cancer; women who do not have cancer but who test positive are known as **false positives**.

The information in the tree diagram can also be represented in a table, as follows.

	T_+	T_-	
C	$\Pr(C \cap T_+) = 0.0085$	$\Pr(C \cap T_-) = 0.0015$	$\Pr(C) = 0.01$
C'	$\Pr(C' \cap T_+) = 0.0693$	$\Pr(C' \cap T_-) = 0.9207$	$\Pr(C') = 0.99$
	$\Pr(T_+) = 0.0778$	$\Pr(T_-) = 0.9222$	

Exercise 13

Consider the diagnostic-testing example above.

a Find $\Pr(C'|T_-)$. This quantity is known as the **negative predictive value**: it concerns **true negatives**, that is, women who do not have cancer and who have a negative test.

b For $\Pr(C) = 0.01$ and $\Pr(T_-|C') = 0.93$, as in the example, what is the largest possible value for the positive predictive value $\Pr(C|T_+)$?

c For $\Pr(C) = 0.01$ and $\Pr(T_+|C) = 0.85$, as in the example, what value of the specificity $\Pr(T_-|C')$ gives a positive predictive value $\Pr(C|T_+)$ of:

i 0.20 **ii** 0.50 **iii** 0.80 **iv** 0.99?

d Breast cancer is many times rarer in men than in women. Suppose that a population of men are tested with the same test, and that the same sensitivity and specificity apply (so $\Pr(T_+|C) = 0.85$ and $\Pr(T_-|C') = 0.93$), but that in men $\Pr(C) = 0.0001$. Find $\Pr(C|T_+)$ and $\Pr(C'|T_-)$.

Answers to exercises

Exercise 1

a A, C, D and E are all events, because they each consist of a collection of possible outcomes. B is not an event for this random procedure, because if we only know the outcome of the roll of the dice, we do not know whether or not B has occurred.

b $$A = \{(1,1),(1,2),(2,1),(1,4),(4,1),(2,3),(3,2),$$
$$(1,6),(6,1),(2,5),(5,2),(3,4),(4,3),(5,6),(6,5)\}.$$

c $|D| = 11$.

d $A \cap E = \varnothing$, the empty set. None of the possible outcomes is in both A and E: there is only one outcome in E, and it is clearly not in A (since $6+6 = 12$, which is not prime).

Exercise 2

For this random procedure, each possible outcome consists of an ordered sequence of 31 numbers, such as $(0.1, 1.2, 0.7, 3.4, \dots)$. The first number is the rainfall on 1 August, the second is the rainfall on 2 August, and so on. There are infinitely many possible outcomes, in principle. Cases like this are not at all unusual. Any time we take a sequence of measurements of a quantity of interest, we have a random procedure of this type.

a
- **i** We know which dates are Mondays in August 2017. So, if we have observed an outcome, we can determine whether or not A has occurred. Thus A is an event.
- **ii** Although B seems related to rainfall, it is not an event for this random procedure. For any given outcome, we cannot say for sure whether B has occurred.
- **iii** The outcomes in C are those sequences for which the sum of the 31 numbers (the daily rainfall amounts) is greater than 20.0. Thus C is an event.

b The pairs of mutually exclusive events are: A and F; E and F.

c
- **i** A' = "there is a rainfall amount greater than zero recorded on at least one day that is not a Monday".
- **ii** $A \cap E$ = "either no rainfall is recorded for the whole month, or else the only day with a rainfall amount greater than zero is the last Monday of the month".
- **iii** $A \cup C$ is a complicated event to put in words: the best way of describing it is to say that either A occurs, or C occurs, or both. An example of an outcome that is not in $A \cup C$ is rainfall of 0.1 mm every day; for such an outcome there is non-zero rainfall on days other than Mondays (so A does not occur) and the total rainfall for the month is 3.1 mm (so C does not occur). Since neither A nor C occurs, the union $A \cup C$ does not occur for this outcome.

Exercise 3

This problem requires you to recognise that, in a four-child family, the event "at least one girl" is the complement of "all boys". Let A = "all four children are boys". Then A' = "there is at least one girl". Since $\Pr(A) = 0.07$, it follows by property 4 that $\Pr(A') = 1 - 0.07 = 0.93$.

Exercise 4

a Using the specified assumption of random mixing, we get $\Pr(A) = \Pr(B) = \frac{1}{16}$.

b The child's wishes will be satisfied if at least one of the two purchases is a Pirate Captain; this event is $A \cup B$. By the addition theorem (property 5), we have

$$\begin{aligned}\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B)\\ &= \frac{1}{16} + \frac{1}{16} - \frac{1}{256} = \frac{31}{256} \approx 0.121,\end{aligned}$$

using part (a) and the specified assumption that $\Pr(A \cap B) = \frac{1}{256}$.

c i B' = "a figure other than a Pirate Captain is purchased at the second shop"

ii $A' \cup B'$ = "either the first purchase is not a Pirate Captain, or the second purchase is not a Pirate Captain, or both"

iii $A \cap B'$ = "the first purchase is a Pirate Captain, and the second purchase is not a Pirate Captain"

iv $A' \cap B'$ = "neither purchase is a Pirate Captain"

d In part (b), we found that $\Pr(A \cup B) = \frac{31}{256} \approx 0.121$. Note that $A' \cap B' = (A \cup B)'$. This is an important general result: *The complement of the union is the intersection of the complements.* By an application of property 4 to this result, we find that

$$\Pr(A' \cap B') = 1 - \Pr(A \cup B) = 1 - \frac{31}{256} = \frac{225}{256} \approx 0.879.$$

So *not* purchasing a Pirate Captain is the more probable event.

Exercise 5

This question reminds us that we are often using an idealised model in probability, in which random mixing is assumed. Whether the mixing is sufficiently random for this assumption to be reasonably satisfied is party a matter of physics (more specifically, the concrete physical processes involved in the mixing). But in practice it usually boils down to an assumption, informed by consideration of the procedure involved.

a A normal flip with many spins is likely to be random, if executed by a person with no particular skill. However, it is (apparently) possible to learn how to flip a coin to produce the desired result. In those cases the mixing is not random at all.

b For reasons associated with exercise 6, a flip that lands on the floor may show a bias towards heads or tails, depending on the specific coin. This does not mean the mixing is not random, but it demonstrates the subtleties of randomising devices.

c With a sufficiently vigorous and lengthy shake, the mixing is random, leading to no preference for any outcome, assuming a symmetric die.

d An overhand shuffle needs to be done for a very long time to produce randomness in the order of the pack, relative to the starting order. If the shuffle is done just a few times, traces of the original order will persist.

e A riffle shuffle produces a random order much more effectively than an overhand shuffle.

f It seems clear that this blast of air produces such chaotic and extensive mixing of the balls that, assuming the balls are of uniform size, shape and weight, all outcomes will be (close to) equally likely.

g Historically, random number generators have been of varying quality; it depends on the method used. Most computer games now use algorithms that will not produce detectable non-randomness.

h A simple physical device like this may well have inadequate mixing.

Exercise 6

What you get depends on the specific coin. Were you surprised by your results?

Exercise 7

There are indeed five possibilities for the number of aces in the hand. But these different possibilities are not equally likely. A standard deck has 52 cards. There are thousands of equiprobable distinct hands of four cards that can be dealt: 270 725, to be precise. Only one of these possible hands contains all 4 aces, whereas there are many hands that contain no aces. So the chance of four aces is $\frac{1}{270\,725}$, while the probability of a hand with no aces is much greater.

Exercise 8

There is some simple reasoning to be applied here. Consider the ballot for the first half of the year, in which 29 February could be selected. In most years, no men born on 29 February could be recruited, even if that marble came up, since their birth year was not a leap year (for example, in 1965 the birth year was 1945). In years such as 1968, the men turning 20 were born in the leap year of 1948, and some would have been born on 29 February. For those men born on 29 February 1948, their chances of conscription were the same as other men born in the first half of 1948.

In fact, there were only two relevant leap years, 1948 and 1952, involving conscription in the years 1968 and 1972. The date 29 February came up in the selected marbles in 1972 and not in 1968.

Exercise 9

These answers are discussed further in the section *Independence*.

a Two lottery draws are usually regarded as mathematically independent because of their physical independence. In principle, there are issues to be considered here. Does the outcome depend at all on the initial position of the balls? If so, could this in any way be related to the position they finished after last week's draw?

b A and B are dependent, because they are mutually exclusive. Any mutually exclusive events are not independent.

c A and B are not likely to be independent, given the seasonal cycle of weather in most places, as discussed after the exercise.

d A and B are independent if the toss is truly random.

Exercise 10

To answer this exercise it is important to recognise that the independence of the experiments means that their outcomes are independent.

a Y can take the values $2,3,4,5,6,7$; U can take the values $2,3,4,\dots,12$; V can take the values $1,2,3,4,5,6$.

b By careful consideration of the outcomes, and the use of independence, we find that $\Pr(U > Y) = \frac{160}{216}$ and $\Pr(U < Y) = \frac{35}{216}$. So it is much more likely that the sum U will be greater than $X+1$, rather than the other way around. Note also that $\Pr(U = Y) = \frac{21}{216}$.

c $\Pr(V > Y) = \frac{90}{216}$ and $\Pr(V < Y) = \frac{91}{216}$. (Very close!) Hence $\Pr(V > Y) < \Pr(V < Y)$. Note also that $\Pr(V = Y) = \frac{35}{216}$.

Exercise 11

a
- **i** Assuming independence, we can calculate the probability of two non-sightings as $0.3 \times 0.3 = 0.09$.
- **ii** $0.3^7 \approx 0.0002$.

b
- **i** If all components have to fail for the machine to fail, the probability of failure of the machine is obtained by multiplying the individual failure probabilities, so the probability equals $0.1 \times 0.01 \times 0.01 \times 0.005 = 5 \times 10^{-8}$. (Very small indeed.)
- **ii** The probability of at least one component failing is one minus the probability that none fails. Assuming independence, the chance that none fails is equal to $0.9 \times 0.99 \times 0.99 \times 0.995 \approx 0.8777$. Hence the probability of failure of the machine is approximately $1 - 0.8777 = 0.1223$. (Much higher.)

c Independence seems doubtful here; the six people could all be delayed by a common causative event, such as inclement weather. Assuming independence, the chance of the building being opened on time is $1 - (0.05 \times 0.1 \times 0.2 \times 0.25 \times 0.5 \times 0.9) \approx 0.9999$.

d This assessment entails the multiplication of probabilities, which assumes independence. It seems plausible that circumstances that precipitate a loss-of-coolant accident might also predispose towards the failure of safety functions: for example, an earthquake or human sabotage.

Exercise 12

a The player loses at the first roll if 1, 4 or 6 is obtained. The probability of this occurring is 0.5.

b To win \$6, the player needs to be paid \$10. This only occurs if the first roll is 5 and the second roll is 5; the probability of this is $\frac{1}{36}$.

c The probability of winning is obtained by adding up the probabilities of outcomes for which the payout is greater than \$4. There are eight such outcomes: $(2,3)$, $(2,5)$, $(3,2)$, $(3,3)$, $(3,5)$, $(5,2)$, $(5,3)$ and $(5,5)$, each with probability equal to $\frac{1}{36}$. So the probability of winning equals $\frac{8}{36} = \frac{2}{9}$.

Exercise 13

a $\Pr(C'|T_-) \approx 0.9984$. So, if the test is negative, there is a very high probability that the woman does not have breast cancer.

b The positive predictive value is calculated as follows:

$$\begin{aligned}\Pr(C|T_+) &= \frac{\Pr(C)\Pr(T_+|C)}{\Pr(C)\Pr(T_+|C) + \Pr(C')\Pr(T_+|C')} \\ &= \frac{0.01 \times \Pr(T_+|C)}{0.01 \times \Pr(T_+|C) + 0.99 \times 0.07}.\end{aligned}$$

This is maximised when the sensitivity $\Pr(T_+|C)$ is 1. For $\Pr(T_+|C) = 1$, the positive predictive value is $\Pr(C|T_+) \approx 0.1261$, which is still quite small.

c Some algebra shows that, for $\Pr(C) = 0.01$ and $\Pr(T_+|C) = 0.85$:

- **i** $\Pr(C|T_+) = 0.20$ requires $\Pr(T_-|C') \approx 0.9657$
- **ii** $\Pr(C|T_+) = 0.50$ requires $\Pr(T_-|C') \approx 0.9914$
- **iii** $\Pr(C|T_+) = 0.80$ requires $\Pr(T_-|C') \approx 0.9979$
- **iv** $\Pr(C|T_+) = 0.99$ requires $\Pr(T_-|C') \approx 0.9999$.

The specificity values required to achieve high positive predictive values are very close to 1.

d The positive predictive value is $\Pr(C|T_+) \approx 0.0012$ and the negative predictive value is $\Pr(C'|T_-) \approx 0.99998$.

0 1 2 3 4 5 6 7 8 9 10 11 12