22 Probability

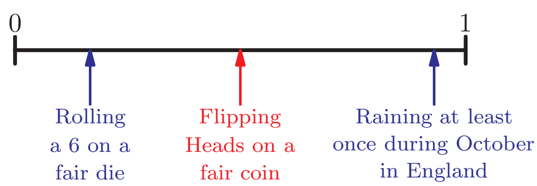
In this chapter you will learn:

- how probability can be estimated from data
- how probability can be predicted theoretically
- how to work out probabilities when you are interested in more than one outcome
- how to work out the probability of a sequence of events occurring
- how to work out probabilities of simple functions of independent random events
- how counting principles can be used to calculate probabilities
- how being given additional information changes our estimate of a probability
- how to infer information from changes in probabilities.

Introductory problem

A woman gives birth to non-identical twins. One of them is a girl. What is the probability that the other one is a girl?

In real life we often deal with uncertain events, but not all events are equally uncertain. It is not certain that the next Steven Spielberg film will be a big hit, nor is it certain that it will snow in India next summer. However, intuitively, these two events do not seem equally likely. We can put events on a scale with impossible at one end and certain at the other and assign a number between 0 and 1 to indicate where the event is on this line. This number is called the probability of the event.



For probability to be useful, it must be more than just a reflection of past experience. We also have to be able to predict probabilities of events in the future, and use these predictions to make decisions. This is the focus of this chapter.

Probability is one of the most recent additions to the field of mathematics. It was formalised by the mathematician Pierre de Fermat (1601–1665) in response to a request from his patron, the Chevalier de Méré, a notorious gambler who wanted help at the gambling table.

22A Introduction to probability

We can estimate the probability of an event practically by doing an experiment repeatedly; this is known as empirical or experimental probability:

KEY POINT 22.1

The probability of an event 'A' occurring is denoted by $P(A)$. From observation, we can estimate $P(A)$ as:

$$P(A) = \frac{\text{Number of times } A \text{ occurs}}{\text{Number of times } A \text{ could have occurred}}.$$

Probability is revised in Prior learning section H on the CD-ROM.

However, in some situations it is possible to predict the probability *before* the experiment (theoretical probability). To do this we need to be able to list all the possible outcomes. This list is called the **sample space**.

For example, when you toss a coin there are two possible outcomes: Heads or Tails. Since both are equally likely (for a 'fair' coin), the probability of each must be one half. This leads to a theoretical definition of probability:

KEY POINT 22.2

$$P(A) = \frac{\text{number of times } A \text{ occurs in the sample space}}{\text{number of items in the sample space}}$$

You might think it seems obvious that these two definitions are equivalent. However, it is quite tricky to prove. If you would like to see how it is done you might like to research the law of large numbers.

Again we have $0 \le P(A) \le 1$. We can now add an interpretation to P(A). If $P(A) = 0$ then the event A is impossible. If $P(A) = 1$ then event A is certain. As P(A) rises the likelihood of A occurring increases.

There are two possible outcomes when you enter a lottery. Either you win or you do not win, but this does not mean that the probability of winning is one half, since there is no reason to believe that both outcomes are equally likely. Many mistakes in probability come from this type of error.

Worked example 22.1

(a) For a family with two children, list the sample space for the sexes of the children, assuming no twins.

(b) Hence find the theoretical probability that the two children are a boy and a girl.

A table provides a systematic way to list the possibilities

(a)

First child	Second child
Boy	Boy
Boy	Girl
Girl	Boy
Girl	Girl

There are four outcomes, and two of them are a boy and a girl

(b) $P(\text{a boy and a girl}) = \frac{2}{4} = \frac{1}{2}$

EXAM HINT

Although the 'boy first, girl second' and 'girl first, boy second' cases can both be described as 'a boy and a girl', we must count them separately in the sample space.

EXAM HINT

'Die' is the singular of 'dice'.

Worked example 22.2

What is the probability of getting a prime number when you roll a six-sided die?

We can list all the possible outcomes and they are all equally likely	Possible outcomes are 1, 2, 3, 4, 5, 6
We can identify how many of them are prime	2, 3 and 5 are prime, which is 3 outcomes out of a sample space of 6.
We can write this as a probability	So probability is $\frac{3}{6} = \frac{1}{2}$

An event either happens or it does not. Everything other than the event happening is called the **complement** of the event. For example, the complement of rolling a 6 on a die is rolling 'not a 6' so 1, 2, 3, 4 or 5. The complement of A is given the symbol A'.

An event and its complement are mutually exclusive (they can't both happen), but we know that one of the two must happen so they have a total probability of 1. We can therefore deduce a formula linking the probability of an event and its complement.

Saying that an event either happens or it does not is called the law of the excluded middle, and it is a basic axiom of standard logic. However, there is an alternative logical system called fuzzy logic where an event could be in a state of 'maybe happening'. As well as lots of real world applications there is a philosophical physics problem called Schrödinger's Cat which applies this idea.

KEY POINT 22.3

$$P(A) + P(A') = 1$$

Suppose that in a board game, players face a penalty whenever the sum of two rolled dice is 7. We do not care how the 7 is achieved – it may be from a 1 and a 6, or a 3 and a 4. In such a situation it is useful to use a probability grid diagram which lists the sample space with each possibility of the first event on one axis, each possibility of the second event on the other axis, giving all possible results in the cells of the table.

Worked example 22.3

What is the probability of getting a sum of 7 when two dice are rolled?

Draw a probability grid diagram showing all possible totals when two dice are rolled

Die B \ Die A	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Count how many items are in the sample space

36 items in the sample space, each of equal probability.

Count how many sums of 7 there are

6 ways to get 7 therefore the probability of a sum of 7 is $\frac{6}{36} = \frac{1}{6}$

EXAM HINT

Notice in the grid in Worked example 22.3 that a '3' on die A and a '1' on die B has to be counted as a separate event from a '1' on die A and '3' on die B. However, a '3' on die A and a '3' on die B is only one event. This often causes confusion.

We are often interested in more complicated events, such as 'getting a grade *A* in Extended Essay or TOK'.

In everyday language the word 'or' can be ambiguous. If you say 'Everyone at the party is a doctor or a lawyer' you generally do not mean that each person could be both. However, in a game, if you say 'I win if I get a black number or an even number' you would also expect to win if you got a black even number. In probability we use the word 'or' in the second sense: '*A* or *B*' means *A* or *B or both* could happen.

Worked example 22.4

When a die is rolled find the probability that the outcome is odd or a prime number.

List the sample space	Possible outcomes: 1, 2, 3, 4, 5, 6
List the events which satisfy the condition	Odd or prime: 1, 2, 3, 5
	Probability $= \frac{4}{6} = \frac{2}{3}$

Exercise 22A

1. List the sample space for each of the following:
   (a) a fair 6 sided die
   (b) arrangements of the letters 'RED'
   (c) the sexes of a 3-child family
   (d) a six sided die with 3 sides labelled '1' and the remaining sides labelled 2, 3, 4.

2. In a standard pack of 52 playing cards there are 4 different suits (red hearts, red diamonds, black clubs and black spades). In each suit there are number cards from 2 to 10, then four 'picture cards'; jack, queen, king and ace. Find the probability that a randomly chosen card is:
   (a) (i) red (ii) a spade
   (b) (i) a jack (ii) a picture card
   (c) (i) a black number card (ii) a club picture card
   (d) (i) not a heart (ii) not a picture card
   (e) (i) a club or a picture card
   (ii) a red card or a number card
   (f) (i) a red number card strictly between three and nine
   (ii) a picture card that is not a jack

EXAM HINT

You can give probability as a fraction, a decimal or as a percentage.

3. A bag contains three different kinds of marble: six are red, four are blue and five are yellow. One marble is taken from the bag. Calculate the probability that it is:
   (a) (i) red (ii) yellow
   (b) (i) not blue (ii) not red

(c) (i) blue or yellow (ii) red or blue

(d) (i) green (ii) not green

(e) (i) neither red nor yellow (ii) neither yellow nor blue

(f) (i) red or green (ii) neither blue nor green

(g) (i) red and green (ii) red and blue

4. By means of an example show that $P(A)+P(B)=1$ does not mean that B is the complement of A.

5. Two fair six-sided dice numbered one to six are rolled. By drawing probability grid diagrams find the probability that:

(a) the sum is 8

(b) the product is greater than or equal to 8

(c) the product is 24 or 12

(d) the largest value is 4

(e) the largest value is more than twice the smaller value

(f) the value on the first die divided by the value on the second die is a whole number.

6. A fair four-sided die (numbered 1 to 4) and a fair eight-sided die (numbered 1 to 8) are rolled. Find the probability that:

(a) the sum is 8

(b) the product is greater than or equal to 8

(c) the product is 24 or 12

(d) the largest value is 4

(e) the largest value is more than twice the smaller value

(f) the value on the eight-sided die divided by the value on the four-sided die is a whole number.

7. Two fair six-sided dice are thrown and the score is the highest common factor of the two outcomes. If this were done 180 times how many times would you expect the score to be 1? *[5 marks]*

8. Three fair six-sided dice are thrown, and the score is the sum of the three results. What is the probability that the score is less than 6? *[5 marks]*

22B Combined events and Venn diagrams

In this section we shall generalise the sample space method so we can efficiently calculate probabilities when we are interested in more than one outcome.

Which is more likely when you roll a die once:

- getting a prime number *and* an odd number?
- getting a prime number *or* an odd number?

The first possibility is restrictive; we have to satisfy both conditions. The second has more possibilities – we can satisfy either condition. So the second must be more likely.

These are examples of two of the most common ways of combining events: Intersection (in normal language 'and') and Union (in normal language 'or'). These are given the following symbols, which might be used in the examination, so get used to them!

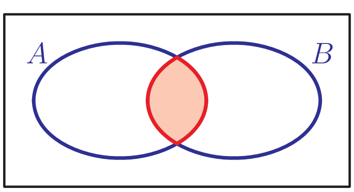
$A \cap B$ is the **intersection** of A and B, meaning when both A and B happens.

$A \cup B$ is the **union** of A and B, meaning when either A happens, or B happens, or both happen.

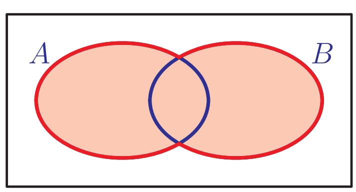
If we have neither apples nor pears then we have no apples and no pears. In set notation this can be written as $(A \cup B)' = A' \cap B'$. This is one of De Morgan's laws - a description of some of the algebraic rules obeyed by sets and hence probability. You will meet these again if you take the Discrete mathematics option (Topic option 10) or the Sets, relations and groups option (Topic option 8).

Why do mathematicians not just use simple words? One of the reasons for this is that everyday language can be ambiguous. If I say that I play rugby or hockey some people may think this means I do not play both. Mathematicians hate ambiguity.

We can use **Venn diagrams** to illustrate the concepts of union and intersection:



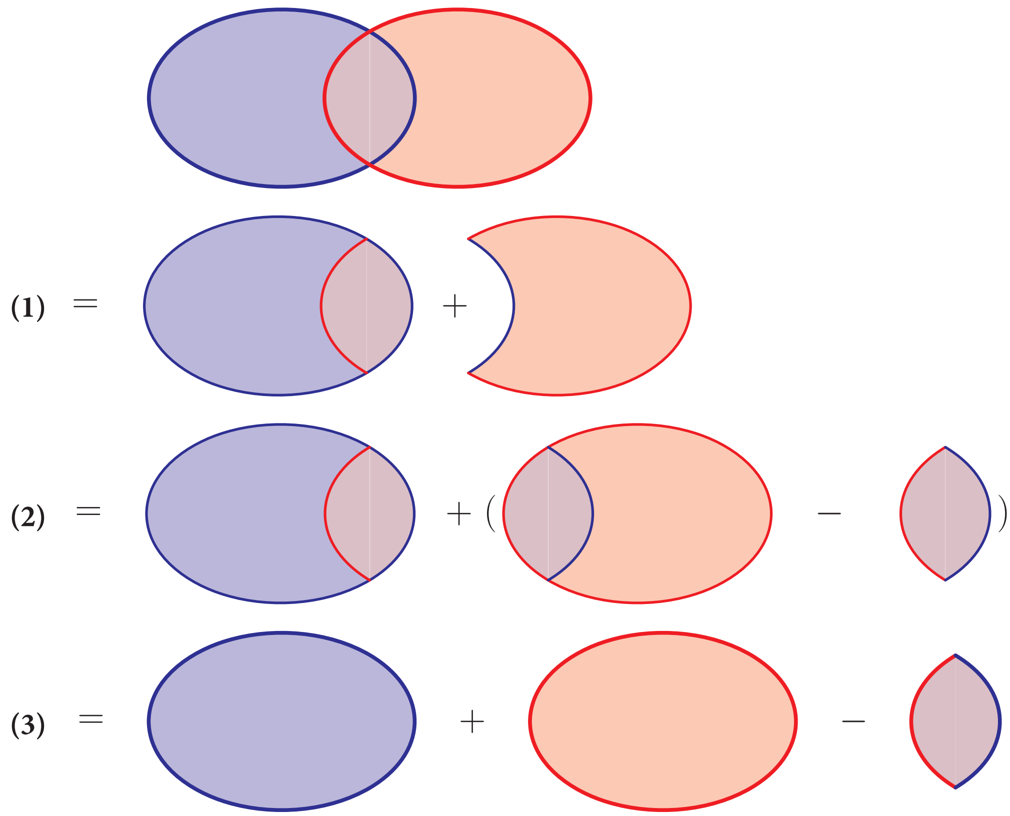
$A \cap B$



$A \cup B$

See Prior learning J on the CD-ROM which explains the concept of Venn diagrams.

Venn diagrams can help us calculate the probability of the union of two events. In the following diagram, the region of the overlap can be seen to be composed of three sections, which can be broken down into their different regions.



This illustrates a very important formula.

KEY POINT 22.4

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

EXAM HINT

Be careful that you do not use the formula for mutually exclusive events unless you are sure that the events cannot both occur simultaneously.

We can interpret this formula as 'If you want to count the number of ways of getting A or B, count the number of ways of getting A and add to that the number of ways of getting B. However, we have then counted the number of ways of getting A and B twice, so we need to compensate by subtracting it'.

If there is no possibility of A and B occurring at the same time, then $P(A \cap B) = 0$. We call these events **mutually exclusive**, and the formula reduces to $P(A \cup B) = P(A) + P(B)$.

Worked example 22.5

A chocolate is selected randomly from a box. The probability of it containing nuts is $\frac{1}{4}$. The probability of it containing caramel is $\frac{1}{3}$. The probability of it containing both nuts and caramel is $\frac{1}{6}$. What is the probability of a randomly chosen chocolate containing either nuts or caramel or both?

Use the formula

$$P(\text{nuts} \cup \text{caramel}) = P(\text{nuts}) + P(\text{caramel}) - P(\text{nuts} \cap \text{caramel})$$

$$= \frac{1}{4} + \frac{1}{3} - \frac{1}{6} = \frac{5}{12}$$

Exercise 22B

1. (a) (i) $P(A) = 0.4$, $P(B) = 0.3$ and $P(A \cap B) = 0.2$.

Find $P(A \cup B)$.

(ii) $P(A) = \frac{3}{10}$, $P(B) = \frac{4}{5}$ and $P(A \cap B) = \frac{1}{10}$.

Find $P(A \cup B)$.

(b) (i) $P(A) = \frac{2}{3}$, $P(B) = \frac{1}{8}$ and $P(A \cup B) = \frac{5}{8}$.

Find $P(A \cap B)$.

(ii) $P(A) = 0.2$, $P(B) = 0.1$ and $P(A \cup B) = 0.25$.

Find $P(A \cap B)$.

(c) (i) $P(A \cap B) = 20\%$, $P(A \cup B) = 0.4$ and $P(A) = \frac{1}{3}$.

Find $P(B)$.

(ii) $P(A \cup B) = 1$, $P(A \cap B) = 0$ and $P(B) = 0.8$.

Find $P(A)$.

(d) (i) Find $P(A \cup B)$ if $P(A) = 0.4$, $P(B) = 0.3$ and A and B are mutually exclusive.

(ii) Find $P(A \cup B)$ if $P(A) = 0.1$, $P(B) = 0.01$ and A and B are mutually exclusive.

2. (a) (i) When a fruit pie is selected at random, $\mathrm{P}(\text{it contains pears}) = \frac{1}{5}$ and $\mathrm{P}(\text{it contains apples}) = \frac{1}{4}$. 10% contain both apples and pears. What is the probability that it contains either apples or pears?

(ii) In a library 80% of books are classed as fiction and 70% were written in the 20th century. Half of the books are 20th century fiction. What proportion of the books are either fiction or from the 20th century?

(b) (i) 95% of pupils in a school play either football or tennis. The probability of a randomly chosen pupil playing football is $\frac{6}{10}$ and the probability of playing tennis is $\frac{5}{8}$. What percentage of pupils play both football and tennis?

(ii) 2 in 5 people in a school study Spanish and 1 in 3 study French. Half of the school study either French or Spanish. What fraction study both French and Spanish?

(c) (i) 90% of pupils in a school have a Facebook account and 3 out of 5 have a Twitter account. One twentieth of pupils have neither a Facebook account nor a Twitter account. What percentage are on both Facebook and Twitter?

(ii) 25% of teams in a football league have French players and a third have Italian players. 60% have neither French nor Italian players. What percentage have both French and Italian players?

3. Simplify the following expressions where possible:

(a) $\mathrm{P}(x > 2 \cap x > 4)$

(b) $\mathrm{P}(y \leq 3 \cup y < 2)$

(c) $\mathrm{P}(a < 3 \cap a > 4)$

(d) $\mathrm{P}(a < 5 \cup a \geq 0)$

(e) $\mathrm{P}(\text{apple} \cup \text{fruit})$

(f) $\mathrm{P}(\text{apple} \cap \text{fruit})$

(g) $\mathrm{P}(\text{multiple of } 4 \cap \text{multiple of } 2)$

(h) $\mathrm{P}(\text{square} \cup \text{rectangle})$

(i) $\mathrm{P}(\text{blue} \cap (\text{blue} \cup \text{red}))$

(j) $\mathrm{P}(\text{blue} \cap (\text{blue} \cap \text{red}))$

4. In a survey, 60% of people are in favour of building a new primary school and 85% are in favour of building a new library. Half of all those surveyed would like both a new primary school and a new library. What percentage supported neither a new library nor a new primary school? *[5 marks]*

5. If $P(A) = 0.2$, $P(A \cap B) = 0.1$ and $P(A \cup B) = 0.7$, find $P(B')$. *[5 marks]*

6. Events A and B satisfy $P((A \cap B)') = 0.2$, $P(A) = P(B) = 0.5$. Find $P(A \cap B')$. *[5 marks]*

7. An integer is chosen at random from the first one thousand positive integers. Find the probability that the integer chosen is:

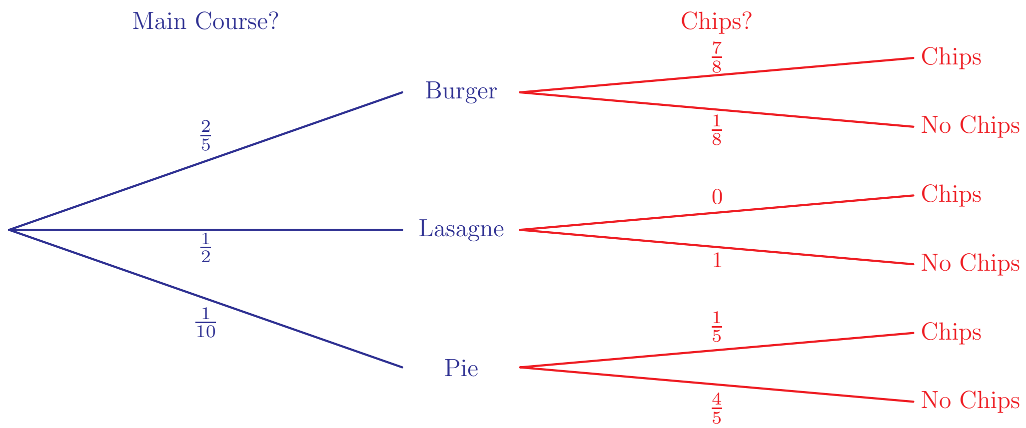
(a) a multiple of 6

(b) a multiple of both 6 and 8. *[5 marks]*

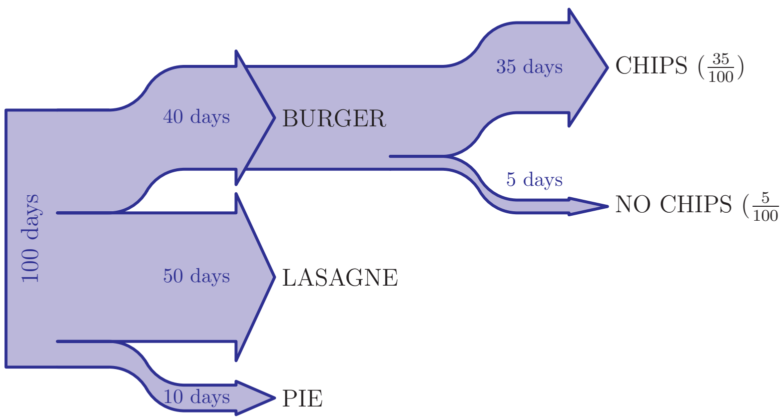
22C Tree diagrams and finding the intersection

Venn diagrams provide a formula linking intersection and union. However they do not help us find a formula for either intersection or union by themselves. To do this it is useful to consider a second method: tree diagrams. When several events happen, either in succession or simultaneously, a **tree diagram** is a way of listing all the possible outcomes. It starts with branches for all the possible outcomes for one of the events, then from each branch we list all possible outcomes for the next event. Along each branch we write the probability of taking that branch.

Tree diagrams have an advantage over the sample space method as they can cope with more than two events, and with outcomes which are not equally likely.



We can consider there to be a filtering process at each branch. Suppose we started with 100 days of school food. On $\frac{2}{5}$ of these days (i.e. 40 days) there will be burgers. On $\frac{7}{8}$ of these burger days there will also be chips. So overall there will be 35 out of 100 days with burgers and chips – or a probability of $\frac{7}{20}$.



To find the probability of travelling along each branch we have found a fraction of a fraction. To do this we need to multiply the two probabilities; but what do these two probabilities represent? In our burger and chips case the $\frac{2}{5}$ represents the probability of burgers, and it might be tempting to say that the $\frac{7}{8}$ is the probability of chips. However, this is not true. Looking at the completed tree diagram we can see that 37 out of 100 days have chips, and this is certainly not $\frac{7}{8}$. The $\frac{7}{8}$ represents the probability of having chips if you already know that there are burgers. This is called **conditional probability**. We use the notation P(chips | burgers), read as 'the probability of chips given burgers'. This leads to a very important rule.

KEY POINT 22.5

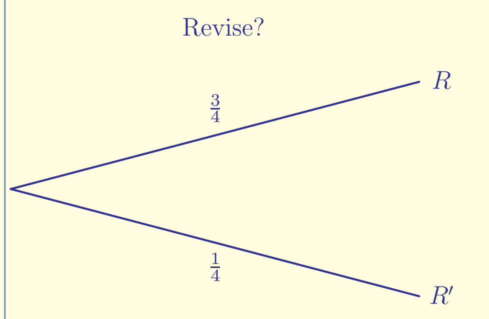
Conditional probability

$$P(A \cap B) = P(B)P(A|B)$$

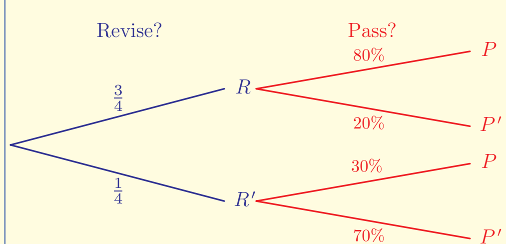
Worked example 22.6

If I revise there is an 80% chance I will pass the maths test, but if I do not revise there is only a 30% chance of passing. I revise for $\frac{3}{4}$ of tests. What proportion of tests do I pass?

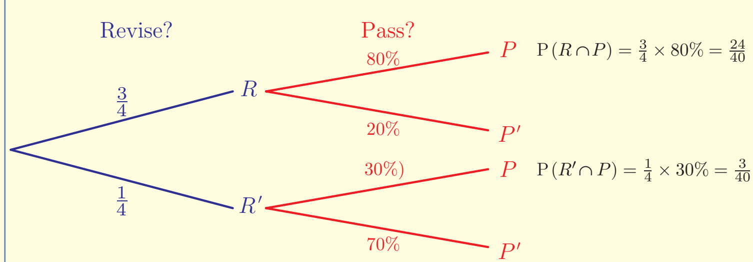
Decide which probability (revise or pass) is not conditional. Start the tree diagram with this event. The probability of passing the test is conditional on revision, so the revision branches have to come first



Add the conditional event



Identify which branches result in passing the test
Multiply to find the probability at the end of each branch



$$P(\text{passing}) = P(\text{revising} \cap \text{passing}) + P(\text{not revising} \cap \text{passing})$$

$$= \frac{24}{40} + \frac{3}{40}$$

$$= \frac{27}{40}$$

EXAM HINT

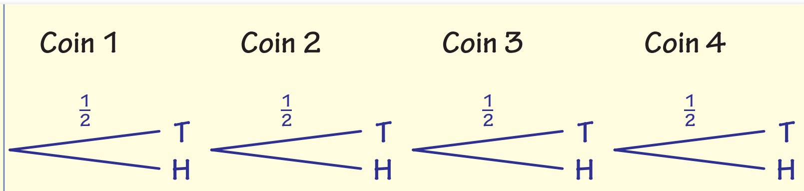
The question in Worked example 22.6 uses the words chance and proportion. These are just other words for probability. Make sure you do not get put off by unusually worded questions.

Sometimes a question may ask for the probability of an event which is quite difficult to find directly, but it is much easier to find the probability of its complement.

Worked example 22.7

Find the probability of getting at least 1 head when you toss four fair coins.

We could draw a tree diagram to find the probability of 1, 2, 3 or 4 heads, but it is easier to find the complement (0 heads)



$$P(0\text{heads}) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

$$P(\geq 1\text{heads}) = 1 - P(0\text{heads})$$

$$= \frac{15}{16}$$

EXAM HINT

When you are asked to find the probability of 'at least…' or 'at most…' finding the complement is frequently a good idea.

Exercise 22C

1. (a) (i) $P(A) = 0.4$ and $P(B \mid A) = 0.3$. Find $P(A \cap B)$.

   (ii) $P(X) = \frac{3}{5}$ and $P(Y \mid X) = 0$. Find $P(X \cap Y)$.

   (b) (i) $P(A) = 0.3$, $P(B) = 0.2$ and $P(B \mid A) = 0.8$.

   Find $P(A \cap B)$.

(ii) $\text{P}(A) = 0.4$, $\text{P}(B) = 0.8$ and $\text{P}(A \mid B) = 0.3$.

Find $\text{P}(A \cap B)$.

(c) (i) $\text{P}(A) = \frac{2}{5}$, $\text{P}(B) = \frac{1}{3}$ and $\text{P}(A \mid B) = \frac{1}{4}$.

Find $\text{P}(A \cup B)$.

(ii) $\text{P}(A) = \frac{3}{4}$, $\text{P}(B) = \frac{1}{4}$ and $\text{P}(B \mid A) = \frac{1}{3}$.

Find $\text{P}(A \cup B)$.

2. A class contains 6 boys and 8 girls. Two are picked at random. What is the probability that they are both boys? *[4 marks]*

3. A bag contains 4 red balls, 3 blue balls and 2 green balls. A ball is chosen at random from the bag and is not replaced. A second ball is chosen. Find the probability of choosing one green ball and one blue ball in any order. *[5 marks]*

4. Given that $\text{P}(X) = \frac{1}{3}$, $\text{P}(Y \mid X) = \frac{2}{9}$ and $\text{P}(Y \mid X') = \frac{1}{3}$, find:

(a) $\text{P}(Y')$

(b) $\text{P}(X' \cup Y')$ *[6 marks]*

5. A factory has two machines making widgets. The older machine has larger capacity, so it makes 60% of the widgets, but 6% are rejected by quality control. The newer machine has only a 3% rejection rate. Find the probability that a randomly selected widget is rejected. *[5 marks]*

6. The school tennis league consists of 12 players. Daniel has a 30% chance of winning any game against a higher-ranked player, and a 70% chance of winning any game against a lower-ranked player. If Daniel is currently in third place, find the probability that he wins his next game against a random opponent. *[5 marks]*

7. There are 36 disks in a bag. Some of them are black and the rest are white. Two are simultaneously selected at random. Given that the probability of selecting two disks of the same colour is equal to the probability of selecting two disks of different colour, how many black disks are there in the bag? *[6 marks]*

22D Independent events

We can now evaluate the intersection and union of two events A and B if we know $P(A|B)$ or $P(B|A)$, but finding this can be quite difficult. There is one important exception, when the two events do not affect each other. These are called **independent events**. In this case knowing that B has occurred has no impact on the probability of A occurring, so for independent events $P(A|B) = P(A)$. However, we know that $P(A \cap B) = P(A|B)\,P(B)$.

KEY POINT 22.6

For independent events: $P(A \cap B) = P(A)P(B)$.

As well as being true for all independent events, Key point 22.6 is actually a defining feature of independent events. So to show that two events are independent we need to show that they satisfy this equation.

All strawberries are red, but not all red things are strawberries. There can be problems distinguishing between a property and a defining feature. A circle has a constant width, but does having a constant width make a plane shape a circle?

Worked example 22.8

$P(A) = 0.5$, $P(B) = 0.2$ and $P(A \cup B) = 0.6$. Are the events A and B independent?

Use the information to find $P(A \cap B)$

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

$\Rightarrow 0.6 = 0.5 + 0.2 - P(A \cap B)$

$\Rightarrow P(A \cap B) = 0.1$

Evaluate $P(A)P(B)$

$P(A)P(B) = 0.1$

Compare the two

$\therefore P(A \cap B) = P(A)P(B)$

So A and B are independent

If we know that two events are independent we can use this to help calculate other probabilities.

Worked example 22.9

A and B are independent events with $P(A \cup B) = 0.8$ and $P(A) = 0.2$. Find $P(B)$.

Write $P(A \cap B)$ in terms of other probabilities

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$\Rightarrow 0.8 = 0.2 + P(B) - P(A \cap B)$$

Use independence

$$= 0.2 + P(B) - 0.2 \times P(B)$$

$$\Rightarrow 0.6 = 0.8 \times P(B)$$

$$\Rightarrow P(B) = \frac{3}{4} \text{ (or 0.75)}$$

Exercise 22D

1. Events A and B are independent.

(a) (i) $P(A) = 0.3$ and $P(B) = 0.7$. Find $P(A \cap B)$.

(ii) $P(A) = \frac{1}{5}$ and $P(B) = \frac{1}{3}$. Find $P(A \cap B)$.

(b) (i) $P(A) = \frac{4}{5}$ and $P(A \cap B) = \frac{3}{7}$. Find $P(B)$.

(ii) $P(A \cap B) = 0.5$ and $P(B) = 0.9$. Find $P(A)$.

(c) (i) $P(A) = 40\%$ and $P(B) = 16\%$. Find $P(A \cup B)$.

(ii) $P(A) = 0.2$ and $P(B) = \frac{1}{4}$. Find $P(A \cup B)$.

(d) (i) $P(A \cup B) = 0.6$ and $P(A) = 0.4$. Find $P(B)$.

(ii) $P(A \cup B) = 0.5$ and $P(A) = 0.1$. Find $P(B)$.

2. Determine which of the following pairs of events are independent:

(a) (i) $P(A) = 0.6$, $P(B) = 0.5$ and $P(A \cap B) = 0.3$

(ii) $P(A) = 0.8$, $P(B) = 0.1$ and $P(A \cap B) = 0.05$

(b) (i) $P(A) = \frac{1}{5}$, $P(B) = \frac{1}{4}$ and $P(A \cup B) = \frac{2}{5}$

(ii) $P(A) = 56\%$, $P(B) = 32\%$ and $P(A \cup B) = 72\%$

3. The independent events A and B are such that $P(A) = 0.6$ and $P(A \cup B) = 0.72$. Find:

(a) $P(B)$

(b) the probability that either A occurs or B occurs, but not both. *[6 marks]*

4. A school has two photocopiers, one for teachers and one for pupils. The probability of the teachers' photocopier working is 92%. The probability of the students' photocopier working is 68%. The two outcomes do not affect each other. What is the probability that:

(a) both photocopiers are working

(b) neither photocopier is working

(c) at least one photocopier is working. *[7 marks]*

5. As part of a promotion a toy is put in each packet of crisps sold. There are eight different toys available. Each toy is equally likely to be found in any packet of crisps.

David buys four packets of crisps.

(a) Find the probability that the four toys in these packets are all different.

(b) Of the eight toys in the packets, his favourites are the yo-yo and the gyroscope. Find the probability that he finds at least one of his favourite toys in these four packets. *[7 marks]*

6. Given that events A and B are independent with $P(A \cap B) = 0.3$ and $P(A \cap B') = 0.6$, find $P(A \cup B)$. *[5 marks]*

22E Counting principles in probability

In chapter 1 you learnt how to count the number of ways in which events can occur.

Counting principles can be used to calculate probabilities in some problems that would be extremely difficult in any other way. A random arrangement of a collection of objects forms a sample space so we can find the probability of particular condition A occurring.

KEY POINT 22.7

$$P(A) = \frac{\text{number of ways in which } A \text{ occurs}}{\text{total number of ways}}$$

Worked example 22.10

A committee of 4 is randomly chosen from 6 girls and 5 boys. What is the probability that the committee contains exactly 3 girls?

First decide how many different committees can be made up. This is a selection where order does not matter

Total number of committees $= \binom{11}{4} = 330$

Then see how may committees have exactly three girls

Choosing 3 girls from 6 and 1 boy from 5 can be done in $\binom{6}{3} \times \binom{5}{1} = 100$ ways

Find the ratio

$P(\text{exactly 3 girls}) = \frac{100}{330} = \frac{10}{33}$

Exercise 22E

1. A set of four alphabet cubes bearing the letters A, R, S and T are dropped in a line at random. What is the probability that they spell out one of the words STAR, RATS or ARTS? *[4 marks]*

2. Consider the word PARTING. What is the probability that a sequence of four letters chosen from this word contains the letter P? *[4 marks]*

3. (a) A team of 11 is chosen randomly from a squad of 18. What is the probability that both the captain and the vice captain are selected?

(b) Two of the squad are goalkeepers and one of them must be chosen. If neither of the goalkeepers is captain or vice captain, what now is the probability that both the captain and the vice captain are selected? *[6 marks]*

4. A team of five students is to be chosen at random to take part in a debate. The team is to be chosen from a group of six history students and three philosophy students. Find the probability that:

(a) only history students are chosen

(b) all three philosophy students are chosen. *[6 marks]*

5. Six boys sit at random in six seats arranged in a row. Two of the boys are brothers. Find the probability that the brothers:

(a) sit at the ends of the row

(b) sit next to each other. *[6 marks]*

See Worked example 1.13 for a similar problem.

6. A rugby team consists of 8 forwards, 7 backs and 5 substitutes. They all line up at random in one row for a picture. What is the probability that:

(a) the forwards are all next to each other?

(b) no two forwards are next to each other? *[8 marks]*

22F Conditional probability

Estimate the probability that a randomly chosen person is a dollar millionaire. Would your estimate change if you were told that they live in a mansion?

When we get additional information, probabilities change. In the above example, P(millionaire) is very different to P(millionaire|lives in a mansion). The second is a conditional probability, and we used it in Section 22 C when looking at tree diagrams.

One important method for finding conditional probabilities is called restricting the sample space. We write out a list of all the equally likely possibilities before we are given any information, and then cross out any possibilities the information rules out.

Worked example 22.11

Given that the number rolled on a die is prime, show that the probability that it is odd is $\frac{2}{3}$.

Write out sample space for a single roll of a die	On one roll we could get 1, 2, 3, 4, 5 or 6
But we are told that the number is prime	If it is prime it can only be 2, 3 or 5
Decide how many of these are odd	Two of these are odd, so the probability is $\frac{2}{3}$

EXAM HINT

Note that P (A ∩ B) = P (B ∩ A), therefore:

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

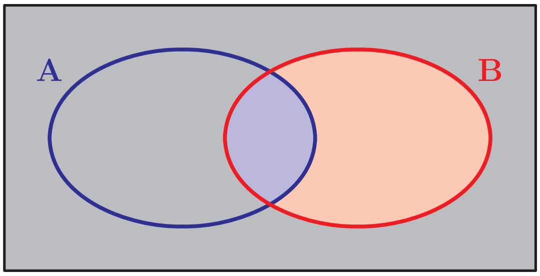
In Key point 22.5 we saw that $P(A \cap B) = P(B)P(A \mid B)$. This formula can be rearranged to get a very important formula for conditional probability.

KEY POINT 22.8

Conditional probability

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

This can be visualised using Venn diagrams.



If B has been given we can ignore all of the Venn diagram except B. The portion of this region in which A occurs is $A \cap B$. We will look at this in more detail in Section 22G.

When calculating a probability it is not always obvious which sample space needs to be used. Supplementary sheet 11 'Measuring risk', on the CD-ROM explores this difficulty in the context of estimating the risk of a transport accident.

Worked example 22.12

The probability that a randomly chosen resident of a city in Japan is a millionaire is $\frac{1}{10000}$. The probability that a randomly chosen resident lives in a mansion is $\frac{1}{30000}$. Only 1 in 40 000 are millionaires who live in mansions. What is the probability of a randomly chosen individual being a millionaire given that they live in a mansion?

Write required probability in 'given' notation and apply the formula

$$\mathrm{P}(\text{millionaire} \mid \text{mansion}) = \frac{\mathrm{P}(\text{millionaire} \cap \text{mansion})}{\mathrm{P}(\text{mansion})}$$

$$= \frac{\left(\frac{1}{40000}\right)}{\left(\frac{1}{30000}\right)}$$

$$= \frac{3}{4}$$

EXAM HINT

It can sometimes be difficult to interpret questions to decide whether they want conditional probability or combined probability. For example, if the question tells you that a boy has green eyes and asks what is the probability that he also has brown hair, it is tempting to find P(green eyes $\cap$ brown hair). However the fact that he has green eyes has been given, so we should actually find P(brown hair | green eyes).

Exercise 22F

1. For each of the questions below write the probability required in mathematical notation. An expression rather than a number is required.

(a) Find the probability that the outcome on a dice is prime and odd.

(b) Find the probability that a person is from either Senegal or Taiwan.

(c) A student is studying the International Baccalaureate®. Find the probability that he is also studying French.

(d) If a playing card is a red card, find the probability that it is a heart.

(e) What proportion of German people live in Munich?

(f) What is the probability that someone is wearing neither black nor white socks?

(g) What is the probability that a vegetable is a potato if it is not a cabbage?

(h) What is the probability that a ball drawn is red given that the ball is either red or blue.

2. (a) (i) If $P(X) = 0.3$ and $P(X \cap Y) = \frac{1}{5}$, find $P(Y \mid X)$.

(ii) If $P(Y) = 0.8$ and $P(X \cap Y) = \frac{3}{7}$, find $P(X \mid Y)$.

(b) (i) If $P(X) = 0.4$, $P(Y) = 0.7$ and $P(X \cap Y) = \frac{1}{4}$, find $P(X \mid Y)$.

(ii) If $P(X) = 0.6$, $P(Y) = 0.9$ and $P(X \cap Y) = \frac{1}{2}$, find $P(Y \mid X)$.

3. The events A and B are such that
$P(A) = 0.6,\ P(B) = 0.2,\ P(A \cup B) = 0.7$.

(a) (i) Find the value of $P(A \cap B)$.

(ii) Hence show that A and B are not independent.

(b) Find the value of $P(B \mid A)$. *[7 marks]*

4. Let A and B be events such that $\mathrm{P}(A)=\frac{2}{3}$, $\mathrm{P}(B\mid A)=\frac{1}{2}$ and $\mathrm{P}(A\cup B)=\frac{4}{5}$.

(a) Find $\mathrm{P}(A\cap B)$.

(b) Find $\mathrm{P}(B)$.

(c) Show that A and B are not independent. *[7 marks]*

5. $\mathrm{P}(A)=\frac{2}{3}$, $\mathrm{P}(A\mid B)=\frac{1}{5}$ and $\mathrm{P}(A\cup B)=\frac{4}{5}$. Find $\mathrm{P}(B)$. *[6 marks]*

22G Further Venn diagrams

If we have information about the number of people in many overlapping groups, a Venn diagram is a very useful way of representing the information.

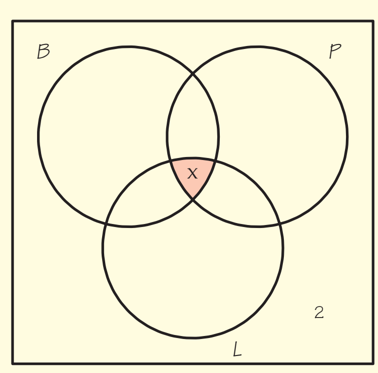
It is helpful to use the convention that the number we put into each region is the number or probability unique to that region. To do this it is often a good idea to label the intersection of all the groups with an unknown and work outwards. Do not try to label the total for regions joined together.

Worked example 22.13

In a class of 32, 19 have a bicycle, 21 have a mobile phone and 16 have a laptop computer. 11 have both a bike and a phone, 12 have both a phone and a laptop and 6 have both a bike and a laptop. 2 have none of these objects.

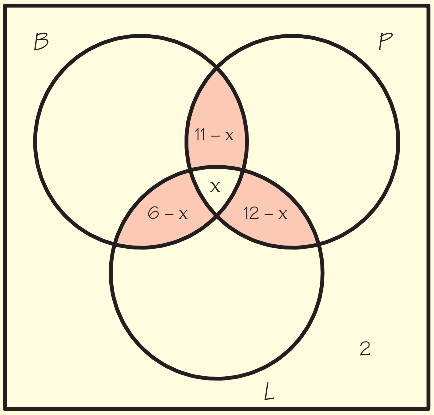
How many have a bike, phone and a laptop?

Draw a Venn diagram showing three overlapping groups, and label the size of the central region as x. We know that 2 people are outside these regions

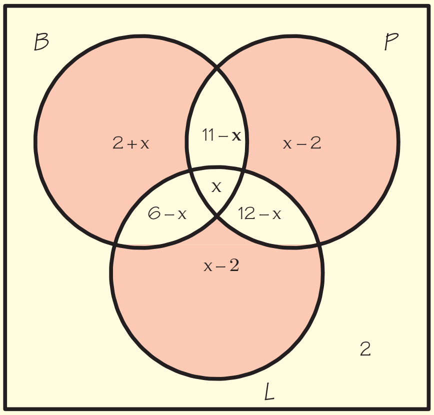


continued . . .

Work outwards. For example the number who have a bicycle and a phone but not a laptop will be $11 - x$



Continue working outwards. For example, the total of all the bicycle regions must be 19, so the remaining section is $19-(11-x)-(6-x)-x$ which is $2+x$



Use the fact that there are 32 people in the class to form an equation

$$(2+x)+(11-x)+(6-x)+x+(x-2)+(12-x)+(x-2)+2=32$$

$$\Leftrightarrow 29+x=32$$

$$\Leftrightarrow x=3$$

Therefore three people have a bicycle, a phone and a laptop.

Venn diagrams are particularly useful when thinking about conditional probability. We can use the given information to exclude parts of the Venn diagram which are not relevant.

Worked example 22.14

Daniel has 18 toys. 12 are made of plastic and 13 are red. 2 are neither red nor plastic.

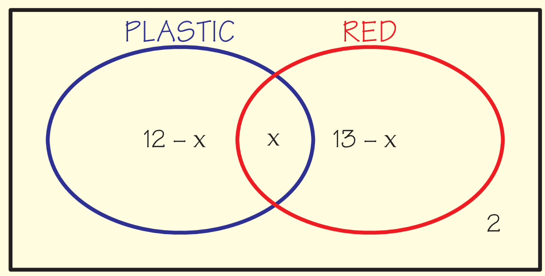
Daniel chooses a toy at random.

(a) Find the probability that it is a red plastic toy.

(b) If it is a red toy, find the probability that it is plastic.

We need to find the size of the intersection. To do this put the given information into a Venn diagram

(a)



We know that there are 18 toys

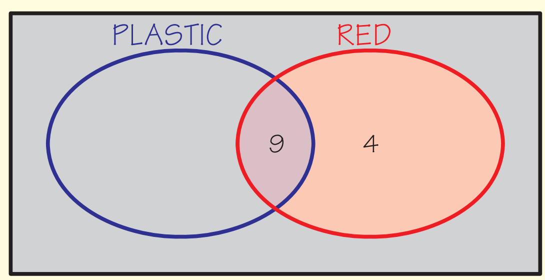
$(12 - x) + x + (13 - x) + 2 = 18 \Leftrightarrow 27 - x = 18$

$\Leftrightarrow x = 9$

$\therefore P(\text{plastic and red}) = \frac{9}{18} = \frac{1}{2}$

We can focus on the red toys

(b)



9 out of 13 red toys are plastic

$\therefore P(\text{plastic} \mid \text{red}) = \frac{9}{13}$

As well as frequencies, Venn diagrams can also represent probabilities. We can still apply the same methods.

Worked example 22.15

Events A and B are such that $P(A) = 0.6$, $P(B) = 0.7$ and $P(A \cup B) = 0.9$. Find $P(B' \mid A')$.

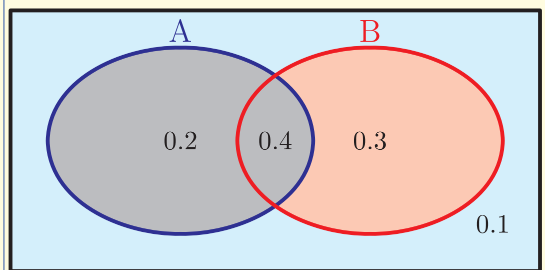
Find $P(A \cap B)$

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

$\therefore 0.9 = 0.6 + 0.7 - P(A \cap B)$

$\Rightarrow P(A \cap B) = 0.4$

Draw a Venn diagram



The probability of not being in A is 0.4. Out of this, the probability of not being in B is 0.1

$$\therefore P(B' \mid A') = \frac{0.1}{0.4} = \frac{1}{4}$$

Some people find this is a useful way of thinking about the formula $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. Are visual arguments clearer than mathematical arguments? If so, why?

Exercise 22G

1. Out of 145 students in a college, 34 play football, 18 play badminton, and 5 play both sports.
   (a) Draw a Venn diagram showing this information.
   (b) How many students play neither sport?
   (c) What is the probability that a randomly chosen student plays badminton?
   (d) If we know that the chosen student plays football, what is the probability that they also play badminton?

2. Out of 145 students in a college, 58 study mathematics, 47 study economics and 72 study neither of the two subjects.
   (a) Draw a Venn diagram to show this information.

(b) How many students study both subjects?

(c) A student tells you that he studies mathematics. What is the probability that he studies both mathematics and economics?

3. Denise conducts a survey about food preferences in the college. She asks students which of the three meals (spaghetti bolognese, chilli con carne, and vegetable curry) they would eat. She finds out that, out of the 145 students:

43 would eat spaghetti bolognese

80 would eat vegetable curry

20 would eat both the bolognese and the curry

24 would eat both curry and chilli

35 would eat both chilli and bolognese

12 would eat all three meals

10 would not eat any of the three meals

(a) Draw a Venn diagram showing this information.

(b) How many students would eat only bolognese?

(c) How many students would eat chilli?

(d) What is the probability that a randomly selected student would eat only one of the three meals?

(e) Given that a student would eat only one of the three meals, what is the probability that they would eat curry?

(f) Find the probability that a randomly selected student would eat at least two of the three meals.

4. The probability that a girl has dark hair is 0.7, the probability that she has blue eyes is 0.4 and the probability that she has both dark hair and blue eyes is 0.2.

(a) Draw a Venn diagram showing this information.

(b) Find the probability that a girl has neither dark hair nor blue eyes.

(c) Given that a girl has dark hair, find the probability that she also has blue eyes.

(d) Given that a girl does not have dark hair, find the probability that she has blue eyes.

(e) Are the characteristics of having dark hair and having blue eyes independent? Explain your answer. *[11 marks]*

5. The probability that it rains on any given day is 0.45, and the probability that it is cold is 0.6. The probability that it is neither cold nor raining is 0.25.

(a) Find the probability that it is both cold and raining.

(b) Draw a Venn diagram showing this information.

(c) Given that it is raining, find the probability that it is not cold.

(d) Given that it is not cold, find the probability that it is raining.

(e) Are the events 'it's raining' and 'it's cold' independent? Explain your answer and show any supporting calculations. *[12 marks]*

22H Bayes' theorem

A test for a rare medical disease is 99% accurate in the sense that:

P (positive result | you have the disease) = 0.99 and

P (negative result | you don't have the disease) = 0.99

If you take the test and a positive result comes back, how likely is it that you have the disease? Most peoples' instinctive reaction to this is 99%, but that is not necessarily even close to the right answer. The problem is that the probability you are interested in is not P (positive result | you have the disease) but P (you have the disease | positive result). These two probabilities can be very different, and **Bayes' theorem** is a formula for relating the two.

Notice that the two given probabilities do not both have to be the same (in fact, they can be very different), so you have to be careful what you mean when you say that a test is 99% accurate. The two cases when the test gives a wrong result are referred to as false positives and false negatives, and are related to errors is hypothesis testing, which you will study if you do Statistics Option 7.

The starting point is the formula for conditional probability:

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

We can replace $P(A \cap B)$ with $P(A)P(B \mid A)$ to get Bayes' theorem:

KEY POINT 22.9

Bayes' theorem:

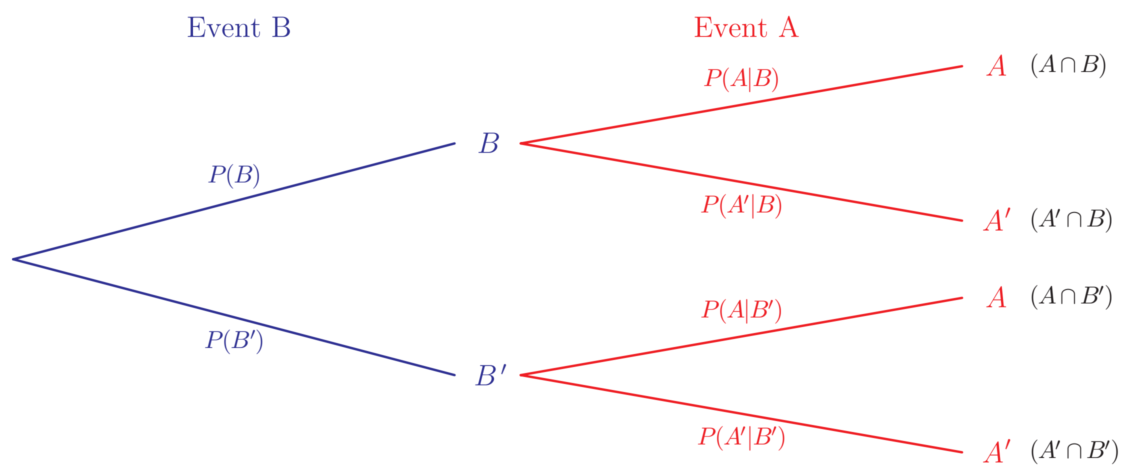
$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$$

EXAM HINT

Note that for $P(B \mid A)$ it is:

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

Often the hardest thing to calculate in Bayes' theorem turns out to be the denominator. This is often done by drawing out a tree diagram:



From the tree diagram we can see that there are two ways to get A, either after B has happened or after B has not happened. This can be expressed as $\mathrm{P}(A) = \mathrm{P}(B)\mathrm{P}(A \mid B) + \mathrm{P}(B')\mathrm{P}(A \mid B')$, leading to the formula quoted in your Formula booklet:

KEY POINT 22.10

Bayes' theorem

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(B)\mathrm{P}(A \mid B)}{\mathrm{P}(B)\mathrm{P}(A \mid B) + \mathrm{P}(B')\mathrm{P}(A \mid B')}$$

Applying Bayes' theorem to the disease detection example above, if we denote A = positive result and B = you have the disease, then the required probability is $\mathrm{P}(B \mid A)$ and the information given tells us that

$\mathrm{P}(A \mid B) = 0.99$ and $\mathrm{P}(A' \mid B') = 0.99$, implying that

$\mathrm{P}(A \mid B') = 0.01$.

Hence:

$$\mathrm{P}(\text{you have the disease} \mid \text{positive result}) = \frac{0.99\ \mathrm{P}(B)}{0.99\ \mathrm{P}(B) + 0.01\ (1 - \mathrm{P}(B))}$$

We do not have sufficient information to calculate this probability, the answer depends on $\mathrm{P}(B)$ which represents the incidence of the disease in the population. It turns out that if $\mathrm{P}(B)$ is very small then so is the required probability. For example, if $\mathrm{P}(B) = \frac{1}{100}$ (so 1% of the population have the disease), the required probability is 50%, but if $\mathrm{P}(B) = \frac{1}{10000}$ it is under 1%. We can make sense of this by arguing 'the disease

The answer to this question may not be what you expected, but it was derived logically using mathematical formulae. When intuition and logic conflict, how do you know which one to trust? Mathematicians have responded to this by developing a rigorous system of proof so that they don't need to rely on intuition.

is so rare that a positive test is far more likely to be a faulty test result of a healthy person than an accurate result of a diseased person'.

The next example illustrates the use of Bayes' theorem when all the required information is given.

Worked example 22.16

Marika travels to school by bus on 70% of school days and by car on the other 30%. The probability that she is late is 0.05 if she travels by bus and 0.12 if she travels by car. On a particular day, Marika is late for school. What is the probability that she travelled by bus?

Write given information and required probability using mathematical notation

Let L = late for school and B = travels by bus

Then:

$P(B) = 0.7,\ P(L \mid B) = 0.05,\ P(L \mid B') = 0.12$

To find: $P(B \mid L)$

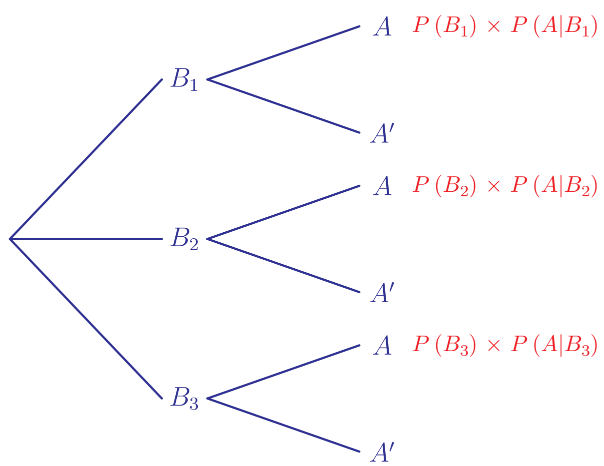
Need to 'reverse the conditional probability' so use Bayes' theorem

$$P(B \mid L) = \frac{P(B)P(L \mid B)}{P(B)P(L \mid B) + P(B')P(L \mid B')}$$

$$= \frac{0.7 \times 0.05}{0.7 \times 0.05 + 0.3 \times 0.12}$$

$$= 0.493$$

It is possible to extend Bayes' theorem to situations which have more options than simply B occurs or B does not occur. Suppose that there were three possible outcomes for the first event, called B_1, B_2 and B_3:



From this we can see that the total probability of A occurring can be expressed as:

$$P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + P(B_3)P(A \mid B_3)$$

Substituting this into Key point 22.9 gives the second version of Bayes' theorem quoted in the Formula booklet:

KEY POINT 22.11

Bayes' theorem

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + \ldots + P(B_n)P(A \mid B_n)}$$

Importantly, outcomes B_1, B_2 up to and including B_n must cover all the possible outcomes.

Worked example 22.17

The Adelaide Eagles want to be sponsored by the International Baccalaureate©. If they come first in the league there is a 90% chance that they will be sponsored. If they come second there is a 20% chance that they will be sponsored and if they come lower than second there is a 5% chance that they will be sponsored. There is a 30% chance that they will come first in the league and a 20% chance that they will come second. At the end of the season they are sponsored by the International Baccalaureate©. What is the probability that they came first in the league?

Decide what conditional probability is being asked for and then write out Bayes' theorem in terms of the events mentioned

Let Sp be the event 'being sponsored'

$$P(1^{st} \mid Sp) = \frac{P(Sp \mid 1^{st})P(1^{st})}{P(1^{st})P(Sp \mid 1^{st}) + P(2^{nd})P(Sp \mid 2^{nd}) + P(< 2^{nd})P(Sp \mid < 2^{nd})}$$

$$= \frac{0.9 \times 0.3}{0.3 \times 0.9 + 0.2 \times 0.2 + 0.5 \times 0.05}$$

$$= 0.806$$

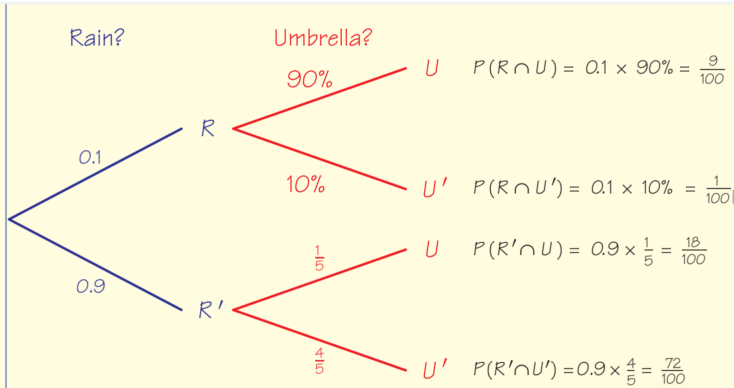
The formula for Bayes' theorem is quite complicated, and you may be worried about identifying all the probabilities correctly. The following example shows that you can also answer these questions without using the formula, you just need the ideas about tree diagrams and conditional probability (this is essentially repeating what we did in deriving the formula).

This illustrates one of the functions of proof in mathematics: To suggest a method for tackling related problems. Does this change your opinion about the usefulness of proof?

Worked example 22.18

If it is raining in the morning there is a 90% chance that I will bring my umbrella. If it is not raining in the morning there is only a $\frac{1}{5}$ chance of me taking my umbrella. On any given morning the probability of rain is 0.1. If you see me with an umbrella, what is the probability that it was raining that morning?

First draw a tree diagram



Write what is required in probability notation: we realise that it is a conditional probability, so use the correct formula

We want $P(\text{rain} \mid \text{umbrella}) = \dfrac{P(\text{rain} \cap \text{umbrella})}{P(\text{umbrella})}$

Use the tree diagram to find relevant probabilities

$$P(\text{rain} \cap \text{umbrella}) = \frac{9}{100}$$

$$P(\text{umbrella}) = \frac{9}{100} + \frac{18}{100} = \frac{27}{100}$$

Use the conditional probability formula

$$P(\text{rain} \mid \text{umbrella}) = \frac{P(\text{rain} \cap \text{umbrella})}{P(\text{umbrella})}$$

$$\frac{\frac{9}{100}}{\frac{27}{100}} = \frac{1}{3}$$

Exercise 22H

1. Box A contains 6 red balls and 4 green balls. Box B contains 5 red balls and 3 green balls. A standard fair cubical die is thrown. If a 6 is scored, a ball is selected from box A; otherwise a ball is selected from box B.

(a) Calculate the probability that the ball selected was red.

(b) Given that the ball selected was red, calculate the probability that it came from box B. *[7 marks]*

2. Robert travels to work by train every weekday from Monday to Friday. The probability that he catches the 7.30 a.m. train on Monday is $\frac{2}{3}$. The probability that he catches the 7.30 a.m. train on any other weekday is 90%. A weekday is chosen at random.

(a) Find the probability that he catches the 7.30 a.m train on that day.

(b) Given that he catches the 7.30 a.m. train on that day, find the probability that the chosen day is Monday. *[7 marks]*

3. Bag 1 contains 6 red cubes and 10 blue cubes. Bag 2 contains 7 red cubes and 3 blue cubes.

Two cubes are drawn at random, the first from bag 1 and the second from bag 2.

(a) Find the probability that the cubes are of the same colour.

(b) Given that the cubes selected are of different colours, find the probability that the red cube was selected from bag 1. *[8 marks]*

4. On any day in April there is a $\frac{2}{3}$ chance of rain in the morning. If it is raining there is a $\frac{4}{5}$ chance I will remember my umbrella, but if it is not raining there is only a $\frac{2}{5}$ chance of remembering my umbrella.

(a) On a random day in April, what is the probability I have my umbrella?

(b) Given that I have an umbrella on a day in April, what is the probability that it was raining? *[6 marks]*

5. The probability that a man leaves his umbrella in any shop he visits is $\frac{1}{5}$. After visiting two shops in succession, he finds he has left his umbrella in one of them. What is the probability that he left his umbrella in the second shop? *[4 marks]*

6. Only two international airlines fly daily into an airport. Pi Air has 40 flights a day and Lambda Air has 25 flights a day. Passengers flying with Pi Air have a $\frac{1}{10}$ probability of losing their luggage and passengers flying with Lambda Air have a $\frac{1}{4}$ probability of losing their luggage. Someone complains that their luggage has been lost. Find the probability that they travelled with Pi Air. *[6 marks]*

7. A girl walks to school every day. If it is not raining, the probability that she is late is $\frac{1}{5}$. If it is raining, the probability that she is late is $\frac{2}{3}$. The probability that it rains during her walk to school on a particular day is $\frac{1}{4}$. On one particular day the girl is late. Find the probability that it was raining on that day. *[7 marks]*

8. If $\mathrm{P}(A) = 0.3$, $\mathrm{P}(B \mid A') = 0.4$ and $\mathrm{P}(A \mid B) = \frac{3}{17}$ find $\mathrm{P}(B \mid A)$. *[6 marks]*

9. Lisa enters a chess tournament in which the result of every match is either win, lose or draw. The probability that she wins the tournament if she wins her first match is 60%. The probability that she wins the tournament if she draws her first match is 50%. The probability that she wins the tournament if she loses her first match is 10%. There is a 50% chance that she wins her first match and a 30% chance that she draws her first match. Given that she wins the tournament, find the probability that she drew her first match. *[6 marks]*

10. When Omar goes to school he walks $\frac{1}{4}$ of the time, catches the bus $\frac{1}{3}$ of the time and cycles the remaining times. If he walks there is a 50% chance of being late. If he catches the bus there is a 25% chance of being late and if he cycles there is a 10% chance of being late. If he is late, what was the probability that he caught the bus? *[6 marks]*

11. Two events A and B are such that $\mathrm{P}(B \mid A) = \frac{3}{5}$ and $\mathrm{P}(B' \cap A) = \frac{1}{3}$. Find $\mathrm{P}(A \cap B)$. *[6 marks]*

12. A new blood test has been devised for early detection of a disease. Studies show that the probability that the blood test correctly identifies someone with this disease is 0.95, and the probability that the blood test correctly identifies someone without that disease is 0.99. The incidence of this disease in the general population is 0.0003.

The result of the blood test on one patient indicates that he has the disease. What is the probability that this patient has the disease? *[8 marks]*

13. You have two coins: one is a normal fair coin with heads on one side and tails on the other. The second coin has heads on both sides. You randomly pick a coin and flip it. The result comes up heads. What is the probability that you chose the fair coin? *[7 marks]*

Summary

- Probability is a measure of how likely an event is, varying from 0 for impossible events up to 1 for certain events.
- The **sample space** is a list of all the possible outcomes.
- The probability of an event A not happening is the **complement** of A (A') and $P(A) + P(A') = 1$.
- Probability can be estimated by looking at previous data (experimental probability) or it can be predicted by finding what proportion of the sample space contains the event (theoretical probability).
- If we have to look at a function of two events, such as their sum or their product, a convenient way of illustrating this is in a probability grid diagram.
- **Venn diagrams** are a useful tool for illustrating how different events can be combined; for representing the information when there are many overlapping groups; and for conditional probability. To calculate probabilities from a Venn diagram, label the intersection of all the groups with an unknown and work outwards.
- In combined probability, the probability of events A or B or both occurring is known as the **union** of A and B: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; in the case when the events are mutually exclusive this becomes: $P(A \cup B) = P(A) + P(B)$.
- The probability of events A and B both occurring is the **intersection** of A and B: $P(A \cap B)$.
- The probability of an event A happening given that an event B has already happened is known as **conditional probability** and is denoted by $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ where $P(A|B)$ means the probability of A given B.
- When two events, A and B, do not affect one another, they are said to be **independent events**:

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

- A sequence of events can be illustrated by a **tree diagram**. To calculate the probability following along a particular path of a tree diagram we use $P(A \cap B) = P(A)P(B \mid A)$, (for conditional probability) and $P(A \cap B) = P(A)P(B)$ and for independent events.

- **Bayes' theorem** is a formula for relating conditional probabilities given that $P(A|B)$ could be very different from $P(B|A)$:

$$P(A|B)=\frac{P(A\cap B)}{P(B|A)P(A)+P(B|A')P(A')}$$ (when there are only two outcomes, e.g. event B occurs or does not occur)

or $$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)}$$ (where there are more than two outcomes, e.g. B_1, B_2, … B_n.)

- The same outcome provided by Bayes' theorem can be achieved using a tree diagram, which can often be the far less complicated method.

- Counting principles, learned in chapter 1, can be helpful when calculating probabilities: $P(A)=\frac{\text{number of ways } A \text{ occurs}}{\text{total number of ways}}$.

Introductory problem revisited

A woman gives birth to non-identical twins. One of them is a girl. What is the probability that the other one is a girl?

There are four equally likely possibilities for the how twins can be born:

First child	Second child
Boy	Boy
Boy	Girl
Girl	Boy
Girl	Girl

If you find this result intriguing, you may like the famous Monty Hall problem.

If you gave the answer $\frac{1}{2}$ to this question, would you be making a mathematical mistake or an error of interpretation? Are they the same thing?

If we are told that one of them is a girl we can exclude the first situation. This leaves three equally likely situations in which only one is two girls, therefore the probability is $\frac{1}{3}$.

We end with a word of warning: it is very tempting to argue that the probability should be $\frac{1}{2}$, as the probability of the second child being a girl is independent of the gender of the first child. This would indeed be the case if the question had said that the *first child* is a girl, rather than that *one of the children* is a girl. Our intuition about probabilities is often flawed, which is why it is important to have precise mathematical methods and use language accurately.

Mixed examination practice 22

Short questions

1. A drawer contains 6 red socks, 4 black socks and 8 white socks. Two socks are picked at random. What is the probability that two socks of the same colour are drawn? *[5 marks]*

2. In a bilingual school there is a class of 21 pupils. In this class, 15 of the pupils speak Spanish as their first language and 12 of these 15 pupils are Argentine. The other 6 pupils in the class speak English as their first language and 3 of these 6 pupils are Argentine.

   A pupil is selected at random from the class and is found to be Argentine. Find the probability that the pupil speaks Spanish as his/her first language. *[4 marks]*

   (IB Organization 1999)

3. The probability that it rains during a summer's day in a certain town is 0.2. In this town, the probability that the daily maximum temperature exceeds 25°C is 0.3 when it rains and 0.6 when it does not rain. Given that the maximum daily temperature exceeded 25°C on a particular summer's day, find the probability that it rained on that day. *[6 marks]*

4. A set of five alphabet blocks bearing the letters A, C, H, R and T are dropped in a line at random.

   (a) What is the probability that they spell out the word CHART?

   (b) What is the probability that the word HAT is formed by three consecutive letters? *[6 marks]*

5. Given that $(A \cup B)' = \varnothing$, $\mathrm{P}(A' \mid B) = \frac{1}{5}$ and $\mathrm{P}(A) = \frac{14}{15}$, find $\mathrm{P}(B)$. *[5 marks]*

Long questions

1. (a) A large bag of sweets contains 8 red and 12 yellow sweets. Two sweets are chosen at random from the bag without replacement. Find the probability that 2 red sweets are chosen.

(b) A small bag contains 4 red and n yellow sweets. Two sweets are chosen without replacement from this bag. If the probability that two red sweets are chosen is $\frac{2}{15}$, show that $n = 6$.

Ayesha has one large bag and two small bags of sweets. She selects a bag at random and then picks two sweets without replacement.

(c) Calculate the probability that two red sweets are chosen.

(d) Given that two red sweets are chosen, find the probability that Ayesha had selected the large bag. *[15 marks]*

2. (a) If $P(X)$ represents a probability, state the range of $P(X)$.

(b) Express $P(A) - P(A \cap B)$ in terms of $P(A)$ and $P(B \mid A)$.

(c) (i) Show that:
$$P(A \cup B) - P(A \cap B) = P(A)\big(1 - P(B \mid A)\big) + P(B)\big(1 - P(A \mid B)\big).$$

(ii) Hence show that $P(A \cup B) \geq P(A \cap B)$. *[9 marks]*

3. The probability that a student plays badminton is 0.3.The probability that a student plays neither football nor badminton is 0.5, and the probability that a student plays both sports is x.

(a) Draw a Venn diagram showing this information.

(b) Find the probability that a student plays football, but not badminton.

Given that a student plays football, the probability that they also play badminton is 0.5.

(c) Find the probability that a student plays both badminton and football.

(d) Hence complete your Venn diagram. What is the probability that a student plays only badminton?

(e) Given that a student plays only one sport, what is the probability that they play badminton? *[13 marks]*

23 Discrete probability distributions

In this chapter you will learn:

- what a random variable is and how to list its possible values
- how to predict the average value and spread of a random variable
- how to calculate the probability of getting a given number of successes over a fixed number of trials (Binomial distribution)
- how to calculate probabilities for events which occur at a fixed rate (Poisson distribution).

Introductory problem

A casino offers a game where a coin is tossed repeatedly. If the first head occurs on the first throw you get £2, if the first head occurs on the second throw you get £4, if the first head is on the third throw you get £8, and so on with the prize doubling each time. How much should the casino charge for this game if they want to make a profit?

In statistics we often find the mean or standard deviation of data we have already collected. However, in real life it is often useful to be able to predict these quantities in advance. Even though it is impossible in a random situation to predict the outcome of a single event, such as one roll of a die, it turns out that if you look at enough events, the average can be predicted quite precisely.

23A Random variables

A **random variable** is a quantity whose value depends upon chance; for example, the outcome when a die is rolled. If the probabilities associated with each possible value are known, useful mathematical calculations can be made. A random variable is conventionally represented by a capital letter and the values which the random variable can take are represented by the equivalent small letter. We may use X to represent the random variable outcome when a die is rolled and in one particular experiment you may find that $x = 2$.

The list of all values in the sample space of a random variable, together with their corresponding probabilities, is called the **probability distribution** or **probability mass function** of the variable; this information is often best displayed in a table.

The idea of being able to predict averages but not individual events is central to many areas of knowledge. Economists cannot predict what an individual will do when interest rates increase, but they can predict how it will affect the economy. In a waterfall a physicist knows that any given water molecule may actually be moving up, but on average the flow is definitely going to be downwards.

Worked example 23.1

Draw a table to show the probability mass function of the outcomes of a fair six-sided die.

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The probabilities cannot be just any numbers:

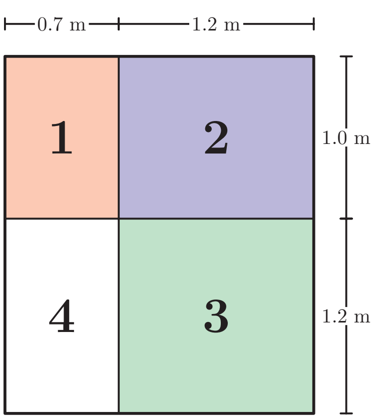
KEY POINT 23.1

The total of all the probabilities of a probability distribution must always equal 1.

This fact is useful if we do not have complete information about the probabilities.

Worked example 23.2

In a game at a fair, a ball is thrown at a rectangular target. The dimensions of the target (in metres) are as shown. The probability of hitting each region is proportional to its area. The prize for hitting a region is the number of chocolates equal to the number shown in that region. Find the probability distribution of the number of chocolates won.



Show the possible values in the table

Express the fact that the probability is proportional to the area by writing $p = k \times \text{area}$

Let X = the number of chocolates won.

x	1	2	3	4
$P(X = x)$	$0.7k$	$1.2k$	$1.44k$	$0.84k$

continued . . .

Use the fact that the probabilities add up to 1

$0.7k + 1.2k + 1.44k + 0.84k = 1$
$\therefore k = 0.239$

We can now calculate all the probabilities
We are not asked for exact values, so round them to 3SF

x	1	2	3	4
$P(X = x)$	0.167	0.287	0.344	0.201

One of the most obvious questions to ask about a random variable is what value it is most likely to have. This value is called the **mode**. The random variable X in the above example has mode 3; the most likely number of chocolates you will win is three. A random variable may not have a mode (for example, the outcomes of a fair die are all equally likely) or it may have more than one mode. In particular, if the largest probability corresponds to two of the outcomes, the random variable is called **bimodal**.

Another question we could ask is, if we were to play the above game many times, on average how many chocolates would we expect to win? The answer is not necessarily the same as the most likely outcome. We will see how to answer this question in the next section.

Exercise 23A

In this exercise you will need to use ideas from chapter 22, particularly tree diagrams. For Question 2(c) you may want to look at chapter 7 on Geometric sequences.

1. For each of the following, draw out a table to represent the probability distribution of the random variable described:
   (a) A fair coin is thrown four times. The random variable W is the number of tails obtained.
   (b) Two fair dice are thrown. The random variable D is the difference between the larger and the smaller score, or zero if they are the same.
   (c) A fair die is thrown once. The random variable X is calculated as half the result if the die shows an even number, or one higher than the result if the die shows an odd number.
   (d) A bag contains six red and three green counters. Two counters are drawn at random from the bag without replacement. G is the number of green counters remaining in the bag.

(e) Karl picks a card at random from a standard pack of 52 cards. If he draws a diamond, he stops; otherwise, he replaces the card and continues to draw cards at random, with replacement, until he has either drawn a diamond or has drawn a total of 4 cards. The random variable C is the total number of cards drawn.

(f) Two fair four-sided spinners, each labelled 1, 2, 3 and 4, are spun. The random variable X is the product of the two values shown.

2. Find the missing value k:

(a) (i)

x	3	7	9	11
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	k

(ii)

x	5	6	7	10
$P(X = x)$	0.2	0.3	k	0.5

(b) (i) $P(Y = y) = ky$ for $y = 1, 2, 3, 4$

(ii) $P(X = x) = \frac{k}{x}$ for $y = 1, 2, 3, 4$

(c) (i) $P(X = x) = k(0.1)^x$ for $x \in \mathbb{N}$

(ii) $P(R = r) = k(0.9)^r$ for $y \in \mathbb{N}$

3. In a game a player rolls a biased four-sided die. The probability of each possible score is shown below.

Score	1	2	3	4
Probability	$\frac{1}{3}$	$\frac{1}{4}$	k	$\frac{1}{5}$

Find the probability that the total score is four after two rolls. *[5 marks]*

23B Expectation, median and variance of a discrete random variable

The **expectation** of a random variable is a value which represents the mean result if the variable were to be repeatedly measured an infinite number of times. It is a representation of the 'average' value of the random variable.

KEY POINT 23.2

The expected value of a discrete random variable X is written $\mathrm{E}(X)$ and calculated as:

$$\mathrm{E}(X) = \sum_x x\mathrm{P}(X = x)$$

EXAM HINT

In the Formula booklet, μ is included to denote $\sum_x x\,\mathrm{P}(X = x)$.

Worked example 23.3

The random variable X has probability distribution as shown in the table below. Calculate $\mathrm{E}(X)$.

x	1	2	3	4	5	6
$\mathrm{P}(X = x)$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{10}$

Apply the formula

$$\mathrm{E}(X) = 1\times\frac{1}{10} + 2\times\frac{1}{4} + 3\times\frac{1}{10} + 4\times\frac{1}{4} + 5\times\frac{1}{5} + 6\times\frac{1}{10}$$

$$= \frac{7}{2}$$

Just as the mean of a set of integers could be fractional, so the expectation of a random variable need not be a value which the variable can itself take.

To find the median of a discrete random variable we use the defining property of the median – that half of the data should fall below it. In the context of a probability distribution this means that:

Median, m, is the smallest value of X that satisfies $\mathrm{P}(X \le m)$ is more than $\frac{1}{2}$. If there is a value m such that $\mathrm{P}(X \le m) = \frac{1}{2}$ then the median is the mean of this value and the next largest value of X.

Probabilities of the form $\mathrm{P}(X \le x)$ which give the probability of being less than or equal to a certain value are called **cumulative probabilities**.

Worked example 23.4

Find the median of the probability distribution below:

x	1	3	6	8
$P(X=x)$	0.2	0.4	0.3	0.1

To find the median evaluate the probability of being below each value until you get above 0.5

$P(X \le 1) = 0.2$

$P(X \le 3) = 0.6$

Therefore the median is 3.

In the above example if the distribution had been

x	1	3	6	8
$P(X=x)$	0.2	0.3	0.4	0.1

then $P(X \le 3)$ is exactly 0.5. The median is the average of 3 and 6, so it is 4.5.

As well as knowing the expectation and median, we may also be interested in how far away from the average we can expect an outcome to be. The variance of a random variable is a value representing the degree of variation that would be seen if the variable were to be repeatedly measured an infinite number of times. It is related to how spread out the variable is.

Standard deviation is a much more meaningful representation of the spread of the variable. So why do we bother with variance at all? The answer is purely to do with mathematical elegance. If you do the statistics option (Topic option 7) you will see that the algebra of variance is far neater than the algebra of standard deviations.

KEY POINT 23.3

The variance of a random variable X is written Var (X) and is calculated as

$$\text{Var}(X) = \text{E}(X^2) - [\text{E}(X)]^2$$

where $\text{E}(X^2) = \sum x^2 \text{P}(X = x)$

(in the Formula booklet this includes an interim notation: $\text{E}(X - \mu)^2$

This is the same idea as the variance of the set of data from Section 21B.

This formula is often quoted as 'the mean of the squares minus the square of the mean.'

Worked example 23.5

Calculate Var(X) for the probability distribution in Worked example 23.3.

Find the expectation

From above, $E(X) = 3.5$

Apply the values from the distribution

$$E(X^2) = 1^2 \times \frac{1}{10} + 2^2 \times \frac{1}{4} + 3^2 \times \frac{1}{10} + 4^2 \times \frac{1}{4} + 5^2 \times \frac{1}{5} + 6^2 \times \frac{1}{10}$$
$$= 14.6$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$
$$= 14.6 - 12.25 = 2.35$$

Exercise 23B

1. Calculate the expectation, median and variance of each of the following random variables:

(a) (i)

x	1	2	3	4
$P(X = x)$	0.4	0.3	0.2	0.1

(ii)

w	8	9	10	11
$P(W = w)$	0.4	0.3	0.2	0.1

(b) (i) $P(X = x) = \frac{x^2}{14}$, $x = 1, 2, 3$

(ii) $P(X = x) = \frac{1}{x}$, $x = 2, 3, 6$

2. A discrete random variable X is given by

$$P(X = x) = k(x + 1) \text{ for } x = 2, 3, 4, 5, 6.$$

(a) Show that $k = 0.04$.

(b) Find $E(X)$. *[5 marks]*

3. The discrete random variable V has the probability distribution shown below and $E(V) = 6.1$. Find the value of k and the median of V.

v	1	2	5	8	k
$P(V = v)$	0.2	0.3	0.1	0.1	0.3

[6 marks]

4. A discrete random variable X has its probability mass function given by

$$P(X = x) = k(x + 3), \text{ where } x \text{ is } 0, 1, 2, 3.$$

(a) Show that $k = \dfrac{1}{18}$.

(b) Find the exact value of $E(X)$. *[6 marks]*

5. The probability distribution of a discrete random variable X is defined by:

$$P(X = x) = kx(4 - x), x = 1, 2, 3$$

(a) Find the value of x.

(b) Find $E(X)$. *[6 marks]*

6. A fair six-sided die, with sides numbered 1, 1, 2, 2, 2, 5 is thrown. Find the mean and variance of the score. *[6 marks]*

7. The table below shows the probability distribution of a discrete random variable X.

x	0	1	2	3
$P(X = x)$	0.1	p	q	0.2

(a) Given that $E(X) = 1.5$, find the values of p and q.

(b) Calculate $Var(X)$. *[9 marks]*

8. A biased die with four faces is used in a game. A player pays 5 counters to roll the die. The table below shows the possible scores on the die, the probability of each score and the number of counters the player wins for each score.

Score	1	2	3	4
Probability	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{20}$
Number of counters player receives	4	5	15	n

Find the value of n in order for the player to get an expected return of 3.25 counters per roll. *[5 marks]*

9. In a game a player pays an entrance fee of $\$n$. He then selects one number from 1, 2, 3 or 4 and rolls three standard dice.

If his chosen number appears on all three dice he wins four times his entrance fee.

If his number appears on exactly two of the dice he wins three times the entrance fee.

If his number appears on exactly one die he wins \$1.

If his number does not appear on any of the dice he wins nothing.

(a) Copy and complete the probability table below.

Profit (\$)	$-n$		$2n$	$3n$
Probability		$\frac{27}{64}$		

(b) The game organiser wants to make a profit over many plays of the game. Given that he must charge a whole number of cents, what is the minimum amount the organiser must charge? *[10 marks]*

23C The binomial distribution

Some discrete probability distributions are met so often that they have been given names and formal notation. One of the most important of these is the **binomial distribution**. There are several others, some of which you will meet in this chapter and some if you study the statistics option (Topic option 7).

A binomial distribution occurs in situations where you have a set number of 'experiments' (or 'trials') each of which have two possible outcomes. The number of trials is usually denoted n. One outcome is conventionally called a 'success' and the other a 'failure'. The probability of success is denoted p. If the probability of success in a trial is constant, and trials are conducted independently of each other, then the number of successes can be modelled using the binomial distribution.

The symbol ~ is used to denote the concept 'follows this distribution', and one or two letter abbreviations are used for the standard distributions. So if a random variable X follows the binomial distribution with n trials and probability of success p, we would write $X \sim \mathrm{B}(n, p)$.

So what is this distribution? Let us consider a specific example: suppose a die is rolled four times, what is the probability of getting exactly two fives? There are four trials so $n = 4$ and if we label a five as a success then $p = \frac{1}{6}$. The probability of a failure is therefore $\frac{5}{6}$.

One way of getting two fives is if the first two times we get a five and the last two times we get something else. The probability of this happening is $\frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} \times \frac{5}{6}$. But this is not the only way

Counting the number of possible selections was discussed in chapter 1.

in which two fives can occur. The two fives may be first and third or second and fourth. In fact, we have to consider all the ways in which we pick two trials out of the four for the 5 to occur. This can happen in $\binom{4}{2}$ ways. Each of them has the same probability as the first case. If X is the random variable 'number of 5s thrown when four dice are rolled' then we can say that:

$$P(X=2)=\binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$$

The useful thing about identifying a binomial distribution is that you can then apply standard results without having to go through this argument every time. In particular, the expectation and variance of the binomial distribution can just be quoted using the formulae below. The proofs of these are beyond what is expected in the International Baccalaureate®, but if you are interested they are on Fill-in proof 25 'Expectation and variance of the binomial distribution' on the CD-ROM.

KEY POINT 23.4

Standard results of the binomial distribution

Statement of distribution	$X \sim B(n, p)$
Probability formula	$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$ for $x = 0, 1, 2, \ldots, n$
Expectation (E(X))	np
Variance (Var(X))	$np(1-p)$

(Note: in the Formula booklet, the expectaion is referred to as the mean)

Worked example 23.6

Rohir has a 30% chance of correctly answering a multiple-choice question. In a test there are ten questions.

(a) What is the probability that Rohir gets exactly four of them correct? Give your answer to five significant figures.

(b) What is the probability that Rohir gets at least one correct in the first five questions?

(c) Suggest which requirements for a binomial distribution might not be satisfied in this situation?

continued . . .

Define the random variable if not already defined in the question	(a) Let X be the number of correct answers in the first ten questions
Give the probability distribution, checking that the conditions are met	$X \sim B(10, 0.3)$
Express the formula for the probability required, and calculate the answer	$P(X=4)=\binom{10}{4}(0.3)^4(0.7)^6=0.20012$ (5SF)
Define the random variable if not already defined in the question	(b) Let Y be the number of correct answers in the first five questions
Give the probability distribution	$Y \sim B(5, 0.3)$
Write down the probability required We are interested in $X \geq 1$, which means that X = 1, 2, 3, 4 or 5. Remember that a quicker way to do the calculation is to find $1-P(X<1)$	$P(X \geq 1)=1-P(X=0)$
Express the formula for the probability required, and calculate the answer	$=1-\binom{5}{0}0.3^0\,0.7^5$ $=0.832$ (3SF)
Consider the requirements for the distribution	(c) Binomial requires: • two outcomes at each trial • constant probability of success in each trial • trial results independent of each other
Identify a requirement which is failed in this context: there are two outcomes, and trials are independent (answering one question does not make it easier or harder to answer another)	All questions are not of the same difficulty, so there might not be a constant probability of success.

If you need to find a probability of a range of successes, you could in theory add up the probabilities of individual outcomes. This can be very time consuming, so your calculator has a function giving the probability of getting up to and including any number of successes.

EXAM HINT

Most GDCs can calculate binomial probabilities automatically given n and p, see Calculator sheet 13 on the CD-ROM. But you may also be tested on applying the formula, which is given in the Formula booklet.

Worked example 23.7

Random variable X has distribution B(15, 0.6). Find $P(5 < X \leq 10)$.

The calculator can give us probabilities of the form $P(X \leq k)$

$X \sim B(15, 0.6)$

$P(5 < X \leq 10) = P(X \leq 10) - P(X \leq 5)$

$= 0.7827 - 0.0338$

$= 0.749$ (3SF) (from GDC)

EXAM HINT

Even when you are using a calculator to find probabilities, you should still use correct mathematical notation (not calculator notation) in your answer. You do not need to explain how you did things on the calculator – just state the distribution you used, the probabilities calculated, and give the answer (usually to 3 significant figures).

Exercise 23C

Remember to round your answer to three significant figures when using the calculator.

1. The random variable X has a binomial distribution with $n = 8$ and $p = 0.2$. Calculate:

   (a) (i) $P(X = 3)$ (ii) $P(X = 4)$

(b) (i) $P(X \le 3)$ (ii) $P(X \le 2)$

(c) (i) $P(X > 3)$ (ii) $P(X > 4)$

(d) (i) $P(X < 5)$ (ii) $P(X < 3)$

(e) (i) $P(X \ge 3)$ (ii) $P(X \ge 1)$

(f) (i) $P(3 < X \le 6)$ (ii) $P(1 \le X < 4)$

2. Given that $Y \sim B(5, 0.5)$, find the exact value of:

(a) (i) $P(Y = 1)$ (ii) $P(Y = 0)$

(b) (i) $P(Y \ge 1)$ (ii) $P(Y \le 1)$

(c) (i) $P(Y > 4)$ (ii) $P(Y \le 3)$

3. Find the mean and standard deviation of each of the following variables:

(a) (i) $Y \sim B\left(100, \frac{1}{10}\right)$ (ii) $X \sim B\left(16, \frac{1}{2}\right)$

(b) (i) $X \sim B(15, 0.3)$ (ii) $Y \sim B(20, 0.35)$

(c) (i) $Z \sim B\left(n-1, \frac{1}{n}\right)$ (ii) $X \sim B\left(n, \frac{2}{n}\right)$

4. (a) Jake beats Marco at chess in 70% of their games. Assuming that this probability is constant and that the results of games are independent of each other, what is the probability that Jake will beat Marco in at least 16 of their next 20 games?

(b) On a television channel, the news is shown at the same time each day; the probability that Salia watches the news on a given day is 0.35. Calculate the probability that on 5 consecutive days she watches the news on exactly 3 days.

(c) Sandy is playing a computer game and needs to accomplish a difficult task at least three times in five attempts in order to pass the level. There is a 1 in 2 chance that he accomplishes the task each time he tries, unaffected by how he has done before. What is the probability that he will pass to the next level?

5. 15% of students at a large school travel by bus. A random sample of 20 students is taken.

(a) Explain why the number of students in the sample who travel by bus is only approximately a binomial distribution.

(b) Use the binomial distribution to estimate the probability that exactly five of the students travel by bus. *[3 marks]*

6. A coin is biased so that when it is tossed the probability of obtaining heads is $\frac{2}{3}$. The coin is tossed 4050 times. Let X be the number of heads obtained. Find:

(a) the mean of X

(b) the standard deviation of X. *[3 marks]*

7. A biology test consists of eight multiple-choice questions. Each question has four answers, only one of which is correct. At least five correct answers are required to pass the test. Sheila does not know the answers to any of the questions, so answers each question at random.

(a) What is the probability that Sheila answers exactly five questions correctly?

(b) What is the expected number of correct answers Sheila will give?

(c) What is the standard deviation in the number of correct answers Sheila will give?

(d) What is the probability that Sheila manages to pass the test? *[7 marks]*

8. 0.8% of people in the country have a particular cold virus at any time. On a single day, a doctor sees 80 patients.

(a) What is the probability that exactly 2 of them have the virus?

(b) What is the probability that 3 or more of them have the virus?

(c) State an assumption you have made in these calculations. *[5 marks]*

9. Given that $Y \sim \mathrm{B}(12,0.4)$:

(a) Find the expected mean of Y.

(b) Find the mode of Y. *[3 marks]*

10. On a fair die, which is more likely: rolling 3 sixes in 4 throws or rolling a five or a six in 5 out of 6 throws? *[6 marks]*

11. Over a one month period, Ava and Sven play a total of x games of tennis. The probability that Ava wins any game is 0.4. The result of each game played is independent of any other game played. Let X denote the number of games won by Ava over a one month period.

(a) Find an expression for $P(X = 2)$ in terms of n.

(b) If the probability that Ava wins two games is 0.121 correct to three decimal places, find the value of n.

[5 marks]

12. A coin is biased so that the probability of it showing tails is p. The coin is tossed n times. Let X be a random variable representing the number of tails. It is known that the mean of X is 19.5 and the variance is 6.825. Find the values of n and p. *[5 marks]*

13. A die is biased so that the probability of rolling a 6 is p. If the probability of rolling 2 sixes in 12 throws is 0.283 (to three significant figures), find the possible values of p correct to two decimal places. *[5 marks]*

14. In an experiment, a trial is repeated n times. The trials are independent and the probability p of success in each trial is constant. Let X be the number of successes in the n trials. The mean of X is 12 and the standard deviation is 2. Find n and p. *[5 marks]*

15. X is a binomial random variable, where the number of trials is 4 and the probability of success of each trial is p. Find the possible values of p if $P(X = 3) = 0.3087$. *[5 marks]*

16. X is a binomial random variable, where the number of trials is 4 and the probability of success of each trial is p. Find the possible values of p if $P(X = 2) = \frac{96}{625}$. *[6 marks]*

Question 10 is the problem which was posed to Pierre de Fermat in 1654 by a professional gambler who could not understand why he was losing. It inspired Fermat (with the assistance of Pascal) to set up probability as a rigorous mathematical discipline.

23D The Poisson distribution

When you are waiting for a bus there are at any given moment two possible outcomes – it either arrives or it does not. We can try modelling this situation using a binomial distribution, but it is not clear what an individual trial is. Instead we have a rate of success – the number of buses that arrive in a fixed time period.

EXAM HINT

If a question mentions average rate of success, or events occurring at a constant rate, you should use Poisson distribution. If you can identify a fixed number of trials then binomial distribution is required.

There are many situations in which we know the rate of events within a given space or time, in contexts ranging from commercial (such as the number of calls through a telephone exchange per minute) to biological (such as the number of clover plants seen per square metre in a pasture). Where the events occur singly (one at a time) and can be considered independent of each other (so that the probability of each event is not affected by what has already happened), the number of events in a fixed space or time interval can be modelled using **Poisson distribution**. This distribution is fully defined once we know the rate of success, which is conventionally called m.

KEY POINT 23.5

Standard results of the Poisson distribution

Statement of distribution	$X \sim \text{Po}(m)$
Probability formula	$\text{P}(X = x) = \dfrac{\text{e}^{-m}m^x}{x!}$ *for* $x = 0, 1, 2, \ldots$
Expectation $\text{E}(X)$	m
Variance $\text{Var}(X)$	m

(Note: in the Formula booklet, $\text{E}(X)$ is called the mean)

Worked example 23.8

Recordable accidents occur in a factory at an average rate of 7 every year, independently of each other. Find the probability that in a given year exactly 3 recordable accidents occurred.

Define the random variable

Let X be the number of accidents in a year

Give the probability distribution

$X \sim \text{Po}(7)$

Write down the probability required, and calculate the answer

$$\text{P}(X = 3) = \frac{e^{-7}7^3}{3!}$$

$$= 0.521 \text{ (3SF)}$$

The Poisson distribution is scaleable. If the number of butterflies seen on a flower in 10 minutes follows a Poisson distribution with mean (expectation) m, then the number of butterflies seen on a flower in 20 minutes follows a Poisson distribution with mean $2m$; the number of butterflies seen on a flower in 5 minutes follows a Poisson distribution with mean $\frac{m}{2}$.

EXAM HINT

See Calculator sheet 13 on the CD-ROM. Your GDC can calculate Poisson probabilities and cumulative probabilities, but you may be explicitly asked to use the formula. Remember to round your answers to 3SF.

Worked example 23.9

If there are an average of 12 buses per hour arriving at a bus stop, find the probability that there are more than 6 buses in 30 minutes.

Define the random variable — Let X be the number of buses in 30 minutes

Give the probability distribution — $X \sim Po(6)$

Write down the probability required. To use the calculator we must relate it to $P(X \leq k)$ — $P(X > 6) = 1 - P(X \leq 6)$

$= 0.161$ (3SF) from GDC

Exercise 23D

1. State the distribution of the variable in each of the following situations:

   (a) Cars pass under a motorway bridge at an average rate of 6 per 10 second period.

      (i) the number of cars passing under the bridge in 1 minute

      (ii) the number of cars passing under the bridge in 15 seconds

   (b) Leaks occur in water pipes at an average rate of 12 per kilometre.

      (i) the number of leaks in 200 m

      (ii) the number of leaks in 10 km

(c) A widget machine manufactures on average 96 functional widgets out of 100.

(i) the number of faulty widgets in a sample of 10

(ii) the number of functioning widgets in sample of 20

(d) 12 worms are found on average in a 1 m^2 area of a garden.

(i) the number of worms found in a 0.3 m^2 area

(ii) the number of worms found in a 2 m by 2 m area

2. Calculate the following probabilities:

(a) If $X \sim \text{Po}(2)$

(i) $\text{P}(X = 3)$ (ii) $\text{P}(X = 1)$

(b) If $Y \sim \text{Po}(1.4)$

(i) $\text{P}(Y \le 3)$ (ii) $\text{P}(Y \le 1)$

(c) If $Z \sim \text{Po}(7.9)$

(i) $\text{P}(Z < 6)$ (ii) $\text{P}(Z < 10)$

(d) If $X \sim \text{Po}(5.9)$

(i) $\text{P}(X \ge 3)$ (ii) $\text{P}(X > 1)$

(e) If $X \sim \text{Po}(11.4)$

(i) $\text{P}(8 < X < 11)$ (ii) $\text{P}(8 \le X \le 12)$

3. A random variable X follows a Poisson distribution with mean 1.7. Copy and complete the following table of probabilities, giving results to 3 significant figures:

x	0	1	2	3	4	>4
$\text{P}(X = x)$	0.183					

4. From a particular observatory, shooting stars are observed in the night sky at an average rate of one every 5 minutes. Assuming that this rate is constant and that shooting stars occur (and are observed) independently of each other, what is the probability that more than 20 are seen over a period of 1 hour? *[4 marks]*

5. When examining blood from a healthy individual under a microscope, a haematologist knows that on average he should see 4 white blood cells in each high power field. Find the probability that blood from a healthy individual will show:

(a) 7 white blood cells in a single high power field

(b) a total of 28 white blood cells in 6 high power fields, selected independently. *[5 marks]*

6. Salah is sowing flower seeds in his garden. He scatters seeds randomly so that the number of seeds falling on any particular region is a random variable with a Poisson distribution, with mean value proportional to the area. He will sow fifty thousand seeds over an area of 2 m^2.

(a) Calculate the expected number of seeds falling on a 1 cm^2 region.

(b) Calculate the probability that a given 1 cm^2 area receives no seeds. *[4 marks]*

7. A wire manufacturer is looking for flaws. Experience suggests that there are on average 1.8 flaws per metre in the wire.

(a) Determine the probability that there is exactly 1 flaw in 1 metre of the wire.

(b) Determine the probability that there is at least one flaw in 2 metres of the wire. *[5 marks]*

8. The random variable X has a Poisson distribution with mean 5. Calculate:

(a) $P(X \leq 5)$

(b) $P(3 < X \leq 5)$

(c) $P(X \neq 4)$

(d) $P(3 < X \leq 5 \mid X \leq 5)$ *[8 marks]*

9. Patients arrive at random at an emergency room in a hospital at the rate of 14 per hour throughout the day.

(a) Find the probability that exactly 4 patients will arrive at the emergency room between 18:00 and 18:15.

(b) Given that fewer than 15 patients arrive in one hour, find the probability that more than 12 arrive. *[6 marks]*

10. The number of eagles observed in a forest in one day follows a Poisson distribution with mean 1.4.

(a) Find the probability that more than three eagles will be observed on a given day.

(b) Given that at least one eagle is observed on a particular day, find the probability that exactly two eagles are seen that day. *[6 marks]*

11. The random variable X follows a Poisson distribution. Given that $P(X \geq 1) = 0.4$, find:

(a) the mean of the distribution

(b) $P(X > 2)$. *[5 marks]*

12. The random variable X is Poisson distributed with mean m and satisfies $P(X = 3) = P(X < 3)$.

(a) Find the value of m, correct to four decimal places.

(b) For this value of m evaluate $P(2 \le X < 4)$. *[6 marks]*

13. Let X be a random variable with a Poisson distribution, such that $P(X > 2) = 0.3$. Find $P(X < 2)$. *[5 marks]*

14. The number of emails you receive per day follows a Poisson distribution with mean 6. Let D be the number of emails received in one day and W the number of emails received in a week.

(a) Calculate $P(D = 6)$ and $P(W = 42)$.

(b) Find the probability that you receive 6 emails every day in a seven-day week.

(c) Explain why this is not the same as $P(W = 42)$. *[8 marks]*

15. The number of mistakes a teacher makes whilst marking homework has a Poisson distribution with a mean of 1.6 errors per piece of homework.

(a) Find the probability that there are at least two marking errors in a randomly chosen piece of homework.

(b) Find the most likely number of marking errors occurring in a piece of homework. Justify your answer.

(c) Find the probability that in a class of 12 pupils fewer than half of them have errors in their marking. *[9 marks]*

16. A car company has two limousines that it hires out by the day. The number of requests per day has a Poisson distribution with mean 1.3 requests per day.

(a) Find the probability that neither limousine is hired.

(b) Find the probability that some requests have to be denied.

(c) If each limousine is to be equally used, on how many days in a period of 365 days would you expect a particular limousine to be in use? *[8 marks]*

17. A shop has 4 copies of the book ‘Ballroom Dancing’ delivered each week. The demand for the book follows a Poisson distribution with mean 3.2 requests per week.

(a) Calculate the probability that the shop cannot meet the demand in a given week.

(b) Find the most probable number of books sold in one week.

(c) Find the expected number of books sold in one week.

(d) Determine the smallest number of copies of the book that should be ordered each week to ensure that the demand is met with a probability of at least 98%. *[8 marks]*

18. The random variable X follows Poisson distribution with mean λ. If $P(X=2)=P(X=0)+P(X=1)$, find the exact value of λ. *[4 marks]*

19. The random variable X follows a Poisson distribution with mean λ.

(a) Show that $P(Y=y+2)=\dfrac{\lambda^2}{(y+1)(y+2)}P(Y=y)$.

(b) Given that $\lambda=6\sqrt{2}$, find the value of y such that $P(Y=y+2)=P(Y=y)$. *[4 marks]*

Summary

- A **random variable** is a quantity whose value depends on chance. A list of all possible outcomes and their associated probabilities is called a **probability distribution** or **probability mass function**.
- The total of all the probabilities of a probability distribution must always equal 1.
- Even though the outcome of any one observation of a random variable is impossible to predict with any certainty, the **expectation** (of the mean) and variance of observations can be predicted quite accurately, using:

$$E(X)=\sum_x xP(X=x)$$

$$\text{Var}(X)=E(X^2)-\left[E(X)\right]^2$$

- If there is a fixed number of trials (each with two possible outcomes) with constant and independent probability of success in each trial then the number of successes follows a **Binomial distribution**: $X \sim B(n,p)$.
- If events occur singly, independently and at a constant rate, then the number of events in a given period follows a **Poisson distribution**: $X \sim Po(m)$, where m is the **rate of success**.
- Once the distribution has been identified then probabilities and statistics for the distribution can be immediately quoted:

Distribution	Notation	$P(X=x)$	$E(X)$	$\text{Var}(X)$
Binomial	$X \sim B(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$
Poisson	$X \sim Po(m)$	$\dfrac{e^{-m}m^x}{x!}$	m	m

- Probabilities in the form $P(X \le x)$ give the probability of being less than or equal to a certain value and are called **cumulative probabilities**.
- You can use your GDC to calculate probabilities in the Binomial and Poisson distributions. Make sure you use the correct setting depending on whether the probability is cumulative or not.

Introductory problem revisited

A casino offers a game where a coin is tossed repeatedly. If the first head occurs on the first throw you get £2, if the first head occurs on the second throw you get £4, if the first head is on the third throw you get £8, and so on with the prize doubling each time. How much should the casino charge for this game if they want to make a profit?

The probability of getting heads on the first throw is $\frac{1}{2}$, so $P(\text{win £2}) = \frac{1}{2}$. The probability of the first head being on the second throw is $P(\text{tails}) \times P(\text{heads}) = \frac{1}{4}$, so $P(\text{win £4}) = \frac{1}{4}$; and so on.

If X is the random variable 'amount of money won', the probability distribution is:

X	2	4	8	…
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	…

$E(X)$ is therefore $2 \times \frac{1}{2} + 4 \times \frac{1}{4} + 8 \times \frac{1}{8} \ldots$ which is $1 + 1 + 1 \ldots$

This sum continues for ever, therefore $E(X) = \infty$. The expected payout over a long period of time is infinite, the casino could not charge sufficient money to cover the expected payout.

Even if you were offered the opportunity to play this game, you should think twice. The result assumes that you can play the game infinitely many times, and in reality this is not the case. It is an example of a famous fallacy called 'Gambler's Ruin'.

Mixed examination practice 23

Short questions

1. A factory making bottles knows that on average, 1.5% of its bottles are defective. Find the probability that, in a randomly selected sample of 20 bottles, at least 1 bottle is defective. *[4 marks]*

2. A biased die with four faces is used in a game. A player pays 10 counters to roll the die and receives a number of counters equal to the value shown on the die. The table below shows the different values on the die and the probability of each.

Value	1	5	10	N
Probability	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{10}$

Find the value represented by N, given that the player has an expected loss of 1 counter each time he plays the game. *[5 marks]*

3. When a boy bats at baseball, the probability that he hits the ball is 0.4. In practice he gets pitched 12 balls. Let X denote the total number of balls he hits. Assuming that his hits are independent, find:

(a) $\mathrm{E}(X)$

(b) $\mathrm{P}(X \leq \mathrm{Var}(X))$ *[5 marks]*

4. A receptionist at a hotel answers on average 35 phone calls a day.

(a) Find the probability that on a particular day she will answer more than 40 phone calls.

(b) Find the probability that she will answer more than 35 phone calls every day during a five-day week. *[5 marks]*

5. When Robyn shoots an arrow at a target, the probability that she hits the target is 0.6. In a competition she has eight attempts to hit the target. If she gets at least seven hits on target she will qualify for the next round.

(a) Find the probability that she hits the target exactly four times.

(b) Find the probability that she fails to qualify for the next round.

(c) Find the probability that she hits the target for the first time on her third attempt. *[6 marks]*

6. During the month of August in Bangalore, India, there are on average 11 rainy days.

(a) Find the probability that there are fewer than seven rainy days during the month of August in a particular year.

(b) Find the probability that in ten consecutive years, exactly five have fewer than seven rainy days in August. *[5 marks]*

7. A company producing light bulbs knows that the probability that a new light bulb is defective is 0.5%.

(a) Find the probability that a pack of 6 light bulbs contains at least one defective one.

(b) Mario buys 20 packs of six light bulbs. Find the probability that more than 4 of the boxes contain at least one defective light bulb. *[6 marks]*

8. (a) Given that $X \sim \text{Po}(m)$ and $\text{P}(X=0)=0.305$ find the value of m.

(b) $Y \sim \text{Po}(k)$. Find the possible values of k such that $\text{P}(X=1)=0.2$.

(c) If $W \sim \text{Po}(\lambda)$ and $\text{P}(W = w + 1) = \text{P}(W = w)$ express w in terms of λ. *[8 marks]*

9. When a fair die is rolled n times, the probability of getting at most two sixes is 0.532 correct to three significant figures.

(a) Find the value of n.

(b) Find the probability of getting exactly two sixes. *[7 marks]*

10. Sonja rolls a single die until she has seen a six twice. Find the probability that she needs more than 5 rolls to do this. *[6 marks]*

Long questions

1. A bag contains a very large number of ribbons. One quarter of the ribbons are yellow and the rest are blue. Ten ribbons are selected at random from the bag.

(a) Find the expected number of yellow ribbons selected.

(b) Find the probability that exactly six of these ribbons are yellow.

(c) Find the probability that at least two of these ribbons are yellow.

(d) Find the most likely number of yellow ribbons selected.

(e) What assumption have you made about the probability of selecting a yellow ribbon? *[11 marks]*

2. A geyser erupts randomly. The eruptions at any given time are independent of one another and can be modelled using a Poisson distribution with mean 20 per day.

(a) Determine the probability that there will be exactly one eruption between 10 a.m and 11 a.m.

(b) Determine the probability that there are more than 22 eruptions during one day.

(c) Determine the probability that there are no eruptions in the 30 minutes Naomi spends watching the geyser.

(d) Find the probability that the first eruption of a day occurs between 3 a.m and 4 a.m.

(e) If each eruption produces 12 000 litres of water, find the expected volume of water produced in a week.

(f) Determine the probability that there will be at least one eruption in each of at least six out of the eight hours the geyser is open for public viewing.

(g) Given that there is at least one eruption in an hour, find the probability that there is exactly one eruption in an hour. *[23 marks]*

3. The probability that a student forgets to do their homework is 5%, independent of other students. If at least one student forgets to do homework, the whole class has to do a test.

(a) There are 12 students in a class. Find the probability that the class will have to do a test.

(b) For a class with n students, write down an expression for the probability that the class will have to do a test.

(c) Hence find the smallest number of students in the class such that the probability that the class will have to do a test is at least 80%. *[12 marks]*

4. Two women, Anna and Brigid, play a game in which they take it in turns to throw an unbiased six-sided die. The first woman to throw a '6' wins the game. Anna is the first to throw.

(a) Find the probability that:

(i) Brigid wins on her first throw

(ii) Anna wins on her second throw

(iii) Anna wins on her nth throw.

(b) Let p be the probability that Anna wins the game. Show that $p = \frac{1}{6} + \frac{25}{36}p$.

(c) Find the probability that Brigid wins the game.

(d) Suppose that the game is played six times. Find the probability that Anna wins more games than Brigid. *[17 marks]*

(*© IB Organization 2001*)

5. On a particular road, serious accidents occur at an average rate of two per week and can be modelled using a Poisson distribution.

(a) (i) What is the probability of at least eight serious accidents occurring during a particular four-week period?

(ii) Assume that a year consists of thirteen periods of four weeks. Find the probability that in a particular year, there are more than nine four-week periods in which at least eight serious accidents occur.

(b) Given that the probability of at least one serious accident occurring in a period of n weeks is greater than 0.99, find the least possible value of n, $n \in \mathbb{Z}^+$. *[18 marks]*

(*© IB Organization 2001*)

6. Two children, Aleric and Bala, each throw two fair cubical dice simultaneously. The score for each child is the sum of the two numbers shown on their respective dice.

(a) (i) Calculate the probability that Aleric obtains a score of 9.

(ii) Calculate the probability that Aleric and Bala each obtain a score of 9.

(b) (i) Calculate the probability that Aleric and Bala obtain the same score.

(ii) Deduce the probability that Aleric's score exceeds Bala's score.

(c) Let X denote the largest number shown on the four dice.

(i) Show that $P(X \leq x) = \left(\frac{x}{6}\right)^4$, for $x = 1, 2, \ldots, 6$.

(ii) Copy and complete the following probability distribution table.

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{1296}$	$\frac{15}{1296}$				$\frac{671}{1296}$

(iii) Calculate $E(X)$. *[13 marks]*

(© IB Organization 2007)

7. Adele is a senior typist and makes an average of 2.5 mistakes per letter. Bozena is a trainee typist and makes an average of 4.1 mistakes per letter. Assume that the number of mistakes made by any typist follows a Poisson distribution.

(a) Calculate the probability that on a particular letter:

(i) Adele makes exactly three mistakes

(ii) Bozena makes exactly three mistakes.

(b) Adele types 80% of all the letters.

(i) Find the probability that a randomly chosen letter contains exactly three mistakes.

(ii) Given that a letter contains exactly three mistakes, find the probability that it was typed by Adele.

(c) Adele and Bozena type one letter each. Given that the two letters contain a total of three mistakes, find the probability that Adele made more mistakes than Bozena. *[16 marks]*

5. (a) 1.5 (b) 1.84

6. $p = 45, q = 5$ and $p = 10, q = 40$

Mixed examination practice 21

Short questions

1. (a) 4.08 (b) 2.97

2. (a) $x = 6, y = 4$ (b) 1367

3. $\frac{79}{400}, \frac{\sqrt{390}}{400}$

4. $2\sqrt{k}$

Long questions

1. (a) 11 (b) 8,5,4
   (c) 11.8

2. (a) The first group starts at 39.5
   (b) 18, 6, 0, 1
   (c) 53.7, 8.50
   (d) Natural variation

3. (a) $\frac{a+2b}{a+b}$ (c) 1.25, 0.433

Chapter 22

Exercise 22A

1. (a) 1,2,3,4,5,6
   (b) RED, RDE, ERD, EDR, DRE, DER
   (c) BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGB
   (d) 1,2,3,4

2. (a) (i) $\frac{1}{2}$ (ii) $\frac{1}{4}$
   (b) (i) $\frac{1}{13}$ (ii) $\frac{4}{13}$
   (c) (i) $\frac{9}{26}$ (ii) $\frac{1}{13}$
   (d) (i) $\frac{3}{4}$ (ii) $\frac{9}{13}$
   (e) (i) $\frac{25}{52}$ (ii) $\frac{11}{13}$
   (f) (i) $\frac{5}{26}$ (ii) $\frac{3}{13}$

3. (a) (i) $\frac{2}{5}$ (ii) $\frac{1}{3}$
   (b) (i) $\frac{1}{3}$ (ii) $\frac{3}{5}$
   (c) (i) $\frac{3}{5}$ (ii) $\frac{2}{3}$
   (d) (i) 0 (ii) 1
   (e) (i) $\frac{4}{15}$ (ii) $\frac{2}{5}$
   (f) (i) $\frac{2}{5}$ (ii) $\frac{11}{15}$
   (g) (i) 0 (ii) 0

5. (a) $\frac{5}{36}$ (b) $\frac{11}{18}$
   (c) $\frac{1}{6}$ (d) $\frac{7}{36}$
   (e) $\frac{1}{3}$ (f) $\frac{7}{18}$

6. (a) $\frac{1}{8}$ (b) $\frac{19}{32}$
   (c) $\frac{5}{32}$ (d) $\frac{7}{32}$
   (e) $\frac{7}{16}$ (f) $\frac{1}{2}$

7. 115

8. $\frac{5}{108}$

Exercise 22B

1. (a) (i) 0.5 (ii) 1
   (b) (i) $\frac{1}{6}$ (ii) 0.05
   (c) (i) $\frac{4}{15}$ (ii) 0.2
   (d) (i) 0.7 (ii) 0.11

2. (a) (i) $\frac{7}{20}$ (ii) all (100%)
   (b) (i) 27.5% (ii) $\frac{7}{30}$
   (c) (i) 55% (ii) 18.3%

3. (a) $P(x > 4)$ (b) $P(y \le 3)$
   (c) O (d) $P(a \in \mathbb{R})$
   (e) P(fruit) (f) P(apple)
   (g) P(multiple of 4) (h) P(rectangle)
   (i) P(blue) (j) $P(\text{blue} \cap \text{red})$

4. 5%

5. 0.4

6. 0.3

7. (a) 0.166 (b) 0.041

Exercise 22C

1. (a) (i) 0.12 (ii) 0
   (b) (i) 0.24 (ii) 0.24
   (c) (i) $\frac{13}{20}$ (ii) $\frac{3}{4}$
2. 0.165
3. $\frac{1}{6}$
4. (a) $\frac{19}{27}$ (b) $\frac{25}{27}$
5. 0.048
6. 0.627
7. 15 or 21

Exercise 22D

1. (a) (i) 0.21 (ii) $\frac{1}{15}$
   (b) (i) $\frac{15}{28}$ (ii) 0.556
   (c) (i) 0.496 (ii) 0.4
   (d) (i) 0.333 (ii) 0.444
2. (a) (i) Yes (ii) No
   (b) (i) Yes (ii) No
3. (a) 0.3 (b) 0.54
4. (a) 62.6% (b) 2.56%
   (c) 97.4%
5. (a) 0.410 (b) 0.684

ANSWER HINT(5)
Did you consider finding the complement?

6. 14/15

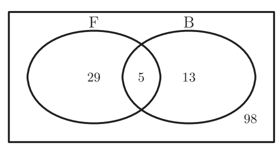
Exercise 22E

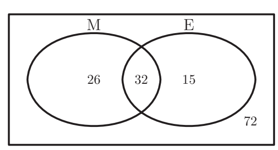
1. $\frac{1}{8}$
2. $\frac{4}{7}$
3. (a) 0.359 (b) 0.375
4. (a) 0.0476 (b) 0.119
5. (a) $\frac{1}{15}$ (b) $\frac{1}{3}$
6. (a) 1.03×10^{-4} (b) 0.0102

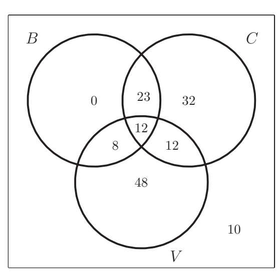
Exercise 22F

1. (a) P(prime ∩ odd)
   (b) P(Senegal ∪ Taiwan)
   (c) P(French|IB)
   (d) P(heart|red)
   (e) P(lives in Munich | German)
   (f) P(not black ∩ not white)
   (g) P(potato|not cabbage)
   (h) P(red|red ∪ blue)
2. (a) (i) $\frac{2}{3}$ (ii) $\frac{15}{28}$
   (b) (i) $\frac{5}{14}$ (ii) $\frac{5}{6}$
3. (a) (i) 0.1 (b) $\frac{1}{6}$
4. (a) $\frac{1}{3}$ (b) $\frac{7}{15}$
5. $\frac{1}{6}$

Exercise 22G

1. (a)


   (b) 98 (c) $\frac{18}{145}$
   (d) $\frac{5}{34}$
2. (a)


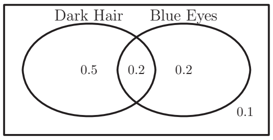
   (b) 32 (c) $\frac{16}{29}$
3. (a)


(b) 0 (c) 79

(d) $\frac{16}{29}$ (e) $\frac{3}{5}$

(f) $\frac{11}{29}$

4. (a)



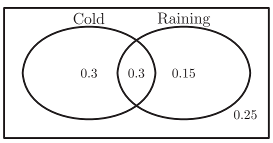
(b) 0.1 (c) $\frac{2}{7}$

(d) $\frac{2}{3}$

(e) No: $P(B \cap D) = 0.2, P(B) \times P(D) = 0.28$

5. (a) 0.3

(b)



(c) $\frac{1}{3}$ (d) $\frac{3}{8}$

(e) No: $P(C \cap R) = 0.3, P(C) \times P(R) = 0.27$

Exercise 22H

1. (a) 0.621 (b) 0.839

2. (a) 0.853 (b) 0.156

3. (a) 0.45 (b) 0.205

4. (a) $\frac{2}{3}$ (b) $\frac{4}{5}$

5. $\frac{4}{9}$

6. $\frac{16}{41}$

7. $\frac{10}{11}$

8. 0.02

9. $\frac{15}{47}$

10. $\frac{1}{3}$

11. 0.5

12. 0.0277

13. $\frac{1}{3}$

Mixed examination practice 22

Short questions

1. 0.320

2. 0.8

3. 0.1111

4. (a) $\frac{1}{120}$ (b) $\frac{1}{20}$

5. $\frac{1}{3}$

Long questions

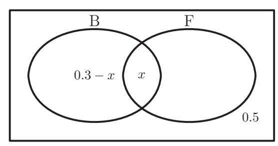
1. (a) $\frac{14}{95}$ (c) 0.138

(d) 0.356

2. (a) $0 \le P(X) \le 1$

(b) $P(A) - P(B|A)P(A)$

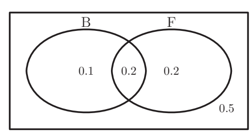
3. (a)



(b) 0.2

(c) 0.2

(d)



, 0.1

(e) $\frac{4}{13}$

Chapter 23

Exercise 23A

1. (a)

w	0	1	2	3	4
$P(W = w)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

(b)

d	0	1	2	3	4	5
$P(D=d)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

(c)

x	1	2	3	4	6
$P(X=x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(d)

g	1	2	3
$P(G=g)$	$\frac{6}{27}$	$\frac{36}{72}$	$\frac{30}{72}$

(e)

c	1	2	3	4
$C=c$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{9}{64}$	$\frac{27}{64}$

(f)

x	1	2	3	4	6	8	9	12	16
$P(X=x)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

2. (a) (i) $\frac{1}{8}$ (ii) 0

(b) (i) $\frac{1}{10}$ (ii) $\frac{12}{25}$

(c) (i) $\frac{9}{10}$ (ii) $\frac{1}{10}$

3. 0.207

Exercise 23B

1. (a) (i) $E(X)=2$, med $=2$, $Var(X)=1$

(ii) $E(X)=9$, med $=9$, $Var(X)=4.2$

(b) (i) $E(X)=2.57$, med $=3$, $Var(X)=0.388$

(ii) $E(X)=3$, med $=2.5$, $Var(X)=2$

2. (b) 4.4

3. $k=\frac{40}{3}$, med = 3.5

4. (b) $\frac{32}{18}$

5. (a) $\frac{1}{10}$ (b) 2

6. $E(X)=\frac{13}{6}$, $Var(X)=\frac{65}{36}$

7. (a) $p=0.5, q=0.2$ (b) 0.35

8. 40

9. (a)

Profit ($)	$-n$	$1-n$	$2n$	$3n$
Probability	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

(b) $0.82

Exercise 23C

1. (a) (i) 0.147 (ii) 0.0459

(b) (i) 0.944 (ii) 0.797

(c) (i) 0.0563 (ii) 0.0104

(d) (i) 0.990 (ii) 0.797

(e) (i) 0.203 (ii) 0.832

(f) (i) 0.0562 (ii) 0.776

2. (a) (i) $\frac{5}{32}$ (ii) $\frac{1}{32}$

(b) (i) $\frac{31}{32}$ (ii) $\frac{6}{32}$

(c) (i) $\frac{1}{32}$ (ii) $\frac{26}{32}$

3. (a) (i) $E(X)=10, \sigma(X)=3$

(ii) $E(X)=8,\ \sigma(X)=2$

(b) (i) $E(X)=4.5, \sigma(X)=1.77$

(ii) $E(X)=7, \sigma(X)=2.13$

(c) (i) $E(X)=\frac{n-1}{n}, \sigma(X)=\left(\frac{n-1}{n}\right)$

(ii) $E(X)=2,\ \sigma(X)=\sqrt{\frac{2(n-2)}{n}}$

4. (a) 0.238
(b) 0.181
(c) 0.5

5. (a) Probability is not constant (effectively sampling without replacement)
(b) 0.103

6. (a) 2700 (b) 30

7. (a) 0.0231 (b) 2
(c) 1.22 (d) 0.0273

8. (a) 0.108
(b) 0.0267
(c) The probability that a person going to the doctor has the virus is the same as for the whole country.
OR
Doctors' patients have colds independently of their chance of visiting.

9. (a) 4.8 (b) 5

10. The second one (0.0165>0.0154)

11. (a) $\frac{2}{9}(n^2 - n) \times 0.6^n$
(b) 10

12. $p = 0.56, n = 30$

13. 0.14 or 0.20

14. $p = \frac{2}{3}$, $n = 18$

15. 0.560 or 0.891

16. $\frac{1}{5}$ or $\frac{4}{5}$

Exercise 23D

1. (a) (i) Po(36) (ii) Po(9)
(b) (i) Po(2.4) (ii) Po(120)
(c) (i) B(10,0.04) (ii) B(20,0.96)
(d) (i) Po(3.6) (ii) Po(48)

2. (a) (i) 0.180 (ii) 0.271
(b) (i) 0.946 (ii) 0.592
(c) (i) 0.201 (ii) 0.729
(d) (i) 0.933 (ii) 0.981
(e) (i) 0.215 (ii) 0.525

3.

x	0	1	2	3	4	>4
$P(X = x)$	0.183	0.311	0.264	0.150	0.0636	0.0296

4. 0.0116

5. (a) 0.0595 (b) 0.0548

6. (a) 2.5 (b) 0.0821

7. (a) 0.298 (b) 0.973

8. (a) 0.616 (b) 0.351
(c) 0.825 (d) 0.570

9. (a) 0.189 (b) 0.372

10. (a) 0.0537 (b) 0.321

11. (a) 0.511 (b) 0.0935

12. (a) $m = 4.5914$ (b) 0.270

13. 0.430

14. (a) 0.161,0.0614 (b) 2.76×10^{-6}
(c) There are other ways to get 42 emails in a week than 6 every day.

15. (a) 0.475 (b) 1
(c) 0.00413

16. (a) 0.273 (b) 0.143
(c) 201

17. (a) 0.219 (b) 4
(c) 2.8 (d) 7

18. $1 + \sqrt{3}$

19. (b) 7

Mixed examination practice 23

Short questions

1. 0.261

2. 55

3. (a) 4.8 (b) 0.0834

4. (a) 0.175 (b) 0.0195

5. (a) 0.232 (b) 0.894
(c) 0.096

6. (a) 0.0412 (b) 2.41×10^{-5}

7. (a) 0.0296 (b) 5.02×10^{-4}

8. (a) 1.19 (b) 0.259, 2.54
(c) $w = \lambda - 1$

9. (a) 15 (b) 0.273

10. (a) 0.804

Long questions

1. (a) 2.5 (b) 0.0162
(c) 0.756 (d) 2
(e) It is constant, and equal to 0.25.

2. (a) 0.362 (b) 0.279
(c) 0.659 (d) 0.0464
(e) 1.68×10^6 litres (f) 0.247
(g) 0.641

3. (a) 0.460 (b) $1-(0.95)^n$
(c) 32

4. (a) (i) $\frac{5}{36}$ (ii) $\frac{25}{216}$
(iii) $\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{2n-2}$

(c) $\frac{5}{11}$ (d) 0.432

5. (a) (i) 0.547 (ii) 0.0894
(b) 3

6. (a) (i) $\frac{1}{9}$ (ii) $\frac{1}{81}$
(b) (i) 0.113 (ii) 0.444
(c) (ii)

x	1	2	3	4	5	6
$P(x = x)$	$\frac{1}{1296}$	$\frac{15}{1296}$	$\frac{65}{1296}$	$\frac{175}{1296}$	$\frac{369}{1296}$	$\frac{671}{1296}$

(iii) 5.24

7. (a) (i) 0.214 (ii) 0.190
(b) (i) 0.209 (ii) 0.818
(c) (i) 0.322

Chapter24

Exercise 24A

1. (a) (i) $\frac{4}{65}$ (ii) $\frac{3}{14}$
(b) (i) $-\frac{2}{3}$ (ii) -1.3
(c) (i) -0.797 (ii) 0.210, 1.38
(d) (i) $\frac{-1 \pm \sqrt{5}}{2}$ (ii) 0.582
(e) (i) $\sqrt{2}$ (ii) $\frac{1 \pm \sqrt{5}}{2}$
(f) (i) -0.418 (ii) 0.754
(g) (i) $3^{\frac{1}{4}}$ (ii) $\sqrt{\frac{2}{3}}$
(h) (i) 51 100 (ii) 0.277
(i) (i) No such k (ii) 0

2. (a) (i) 0.48 (ii) 0.75
(b) (i) $\frac{\sqrt{3} - \sqrt{2}}{2}$ (ii) 0.5
(c) (i) 0.301 (ii) 0.477

3. (a) (i) 0.632 (ii) 0.949
(b) (i) 1.26 (ii) 2.83
(c) (i) 0.4 (ii) 3.5

4. (a) 0.0968
(b) 370

5. e^{-6}

6. 0.399

7. (a) $b = e^k$ (b) $a = \frac{2(e^k - 1)}{e^k + 1}$

8. 0.560

Exercise 24B

1.

	$E(X)$	Median	Mode	$Var(X)$
(a) (i)	$\frac{1}{3}$	0.293	0	$\frac{1}{18}$
(a) (ii)	$\frac{8}{3}$	$\sqrt{8}$	8	$\frac{8}{9}$
(b) (i)	3.91	$\sqrt{10}$	1	6.22
(b) (ii)	1.39	$\frac{4}{3}$	1	0.0782
(c) (i)	0.571	$\frac{\pi}{6}$	0	0.0142
(c) (ii)	0.386	0.405	ln 2	0.0391
(d) (i)	1.5	1.26	1	0.75
(d) (ii)	1.33	1.19	1	0.222

2. (a) (i) 1.21 (ii) 11
(b) (i) 2.82 (ii) 4.18

3. (a) 1.44 (b) 2

4. (a) $\frac{3}{973}$ (b) 7.65

5. (b) $\frac{1}{k}$

6. (b) $\frac{5}{\sqrt{3}}$

7. 0

Exercise 24C

1. (a) (i) 0.885 (ii) 0.212
(b) (i) 0.401 (ii) 0.878
(c) (i) 0.743 (ii) 0.191
(d) (i) 0.807 (ii) 0.748
(e) (i) 0.997 (ii) 0.055

2. (a) (i) 0.5 (ii) 1
(b) (i) -1.67 (ii) -0.4

3. (a) (i) $P(Z < 1.6)$ (ii) $P(Z < 1.28)$
(b) (i) $P(Z \geq -0.68)$ (ii) $P(Z \geq -2.96)$