STATISTICS AND PROBABILITY

Chapter 11

# Probability

11

SYLLABUS 1.10 4.5 4.6 4.11 4.13

Probability puts a number on how likely an event is. That step lets you measure risk, compare it and act on it, so that a decision can rest on more than a feeling about what might happen.

Here is a decision that depends on such a number. A health service screens everyone in an age group for a rare disease. The test is a good one. Of every 100 people who have the disease, it correctly identifies 99. Of every 100 people who do not have it, it correctly clears 95. About one person in a hundred has the disease.

Your result comes back positive. How likely is it that you have the disease?

A common answer is about 99%, and in published studies many doctors given a question of this kind have answered in the same way. The correct answer is about 17%. The test is not faulty and none of the numbers are unusual. The error is in the reasoning. That 99% says how the test behaves *when the disease is present*, and the quantity you actually want is a different one: how likely the disease is *now that the result is positive*. Those two numbers are not the same, and this chapter builds the rule that connects them.

So probability gives clear answers, and it is easy to apply a clear answer to the wrong question. Two habits guard against that, and both are practised throughout the chapter. The first is to state precisely which event you are asking about. The second is to keep track of where a probability came from: some are calculated from the symmetry of a fair die, others only measured by counting what happened in 200 trials, and the two are found, and can be trusted, in different ways.

We begin with a simple experiment, a single roll of a die. From there the chapter builds the language of events, the rules for combining them, conditional probability and independence, tree diagrams, and finally Bayes' theorem. By the end of this chapter you will be able to find probabilities from Venn diagrams, tree diagrams and counting, test events for independence, find conditional probabilities, and use Bayes' theorem to reverse a conditional probability. When we return to the test in the last section, you will be able to find its answer in one step.

 

# 11.1 The Language of Probability

Probability measures how likely something is, on a scale from 0 (impossible) to 1 (certain). To compute it we need to be precise about what "something" means.

Five words are used precisely from here on. The die is used to illustrate each one.

- **Experiment**: any process with an uncertain result. Rolling a die is an experiment.
- **Trial**: one repetition of the experiment. One roll is one trial.
- **Outcome**: the result of a single trial, such as a $3$.
- **Sample space** $U$: the set of all possible outcomes. For one die, $U = \{1, 2, 3, 4, 5, 6\}$.
- **Event**: any subset of the sample space, that is, a collection of outcomes you are interested in. "An even number" is the event $\{2, 4, 6\}$.

When every outcome is equally likely, probability is just counting: the fraction of outcomes that belong to the event.

KEY · IB

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

Here $n(A)$ is the number of outcomes in event $A$, and $n(U)$ the number in the whole sample space. Every probability obeys $0 \leq P(A) \leq 1$.

**Worked through.** When a fair die is rolled once, the sample space has $n(U) = 6$ equally likely outcomes, so the probabilities of an even number and of a number greater than 4 are

$$P(\text{even}) = \frac{n(\{2,4,6\})}{6} = \frac{3}{6} = \frac{1}{2}, \qquad P(>4) = \frac{n(\{5,6\})}{6} = \frac{2}{6} = \frac{1}{3}.$$

Counting works only because the die is fair. If the outcomes were not equally likely, this formula would not apply.

## 11.1.1 Relative frequency

When there is no symmetry to count with, the probability can be estimated by experiment. Repeat the trial many times and record how often the event occurs. The fraction of trials in which it happens is the **relative frequency** of the event, and it serves as an experimental estimate of the probability.

KEY After $n$ trials,

$$P(A) \approx \frac{\text{number of trials in which } A \text{ occurred}}{n}$$

The estimate is rough when $n$ is small, and it becomes more stable as $n$ grows. This is the difference between the two kinds of probability you will meet: a *theoretical* probability comes from the structure of the experiment (a fair die, by symmetry), while an *experimental* one comes from data, and is what you fall back on when there is no symmetry to use.

**Worked through.** A drawing pin is dropped 200 times and lands point-up 78 times. There is no symmetry argument for a drawing pin, so to estimate the probability that a dropped pin lands point-up, use the relative frequency:

$$P(\text{point-up}) \approx \frac{78}{200} = 0.39.$$

A further 200 drops would almost certainly not give exactly 78 again, but the estimate becomes more accurate as the number of trials increases.

## 11.1.2 The complement

Sometimes the event you want is awkward to count, but its opposite is easy. The **complement** of $A$, written $A'$, is the event that $A$ does *not* happen. Since $A$ either happens or it does not, their probabilities must fill the whole scale.

KEY · IB

$$P(A) + P(A') = 1 \qquad \Longrightarrow \qquad P(A') = 1 - P(A)$$

The complement is the route to take whenever the event itself is hard to count. In particular, when a question contains the phrase "at least one," the complement is usually the quicker route.

**Worked through.** To find the probability of getting at least one head when a fair coin is tossed three times, compare the direct count with the complement. Counting every way to get one, two, or three heads is possible but slow. The complement of "at least one head" is "no heads at all," a single outcome: TTT.

$$P(\text{at least one head}) = 1 - P(\text{TTT}) = 1 - \frac{1}{8} = \frac{7}{8}.$$

One subtraction replaces a count of seven outcomes. The saving becomes far greater as the number of tosses increases.

### 11.1.3 Expected number of occurrences

If an event has probability $p$ and you repeat the experiment $n$ times, you expect it to happen about $np$ times.

KEY If an event has probability $p$ and the experiment is repeated $n$ times, the **expected number of occurrences** is

$$\text{expected number} = np.$$

This is not a guarantee. It is a long-run average. If you repeated the whole set of $n$ trials many times and averaged the counts, that average would be close to $np$.

**Worked through.** To find how many sixes to expect when a fair die is rolled 60 times, note that each roll gives a six with probability $\frac{1}{6}$, so over 60 rolls:

$$\text{expected sixes} = 60 \times \frac{1}{6} = 10.$$

You will rarely get exactly 10, but the average over many sets of 60 rolls will be close to 10.

**YOUR TURN 11.1 The Language of Probability**

**Equally likely outcomes**

**1.**

(a) A fair die is rolled. Find the probability that the number is odd.
(b) A card is drawn from a deck of 52. Find the probability that it is a heart.
(c) A bag holds 4 red, 6 green and 5 blue counters. Find the probability that a counter drawn at random is green.
(d) A letter is chosen at random from the word STATISTICS. Find the probability that it is an S.
(e) A spinner numbered 1 to 8 is spun. Find the probability that it lands on a multiple of 3.

**2.** List the sample space in each case.

(a) Two fair coins are tossed. List the sample space, and find the probability of exactly one head.
(b) Three fair coins are tossed. List the sample space, and find the probability of exactly two heads.

**Relative frequency and expected number**

**3.** A spinner has five sectors numbered 1 to 5, but it is not fair. In 400 spins it lands on 3 exactly 96 times.

(a) Estimate the probability of landing on 3.
(b) Using that estimate, find the expected number of 3s in 250 spins.
(c) Explain why the probability in question 2(a) can be found without any experiment, while the one in part (a) of this question cannot.

**4.**

(a) A fair die is rolled 90 times. Find the expected number of 5s.
(b) Records show that a train was late on 45 of the last 300 days. Estimate the probability that it is late on a given day, and the expected number of late days in a month of 20 working days.

**The complement**

**5.**

(a) Given $P(A) = 0.35$, find $P(A')$.
(b) A letter is chosen at random from the word PROBABILITY. Find the probability that it is not a B.
(c) A fair coin is tossed four times. Find the probability of at least one tail.
(d) A fair die is rolled twice. Find the probability that at least one of the rolls is not a six.

**Concept check**

**6.** A fair coin has landed heads five times in a row. A student says: "*the next toss is more likely to be tails, because the results have to even out.*" Explain why the student is wrong.

**7.** A student writes: "*the probability of a six on one roll is* $\frac{1}{6}$, *so the probability of at least one six in six rolls is* $6 \times \frac{1}{6} = 1$." Identify the error, and find the correct probability.

## 11.2 Counting the Outcomes HL

The formula $P(A) = \dfrac{n(A)}{n(U)}$ is only as useful as your ability to find the two numbers in it. For a die you list the outcomes. For a committee of four chosen from twelve people you cannot: there are 495 of them. The counting tools of Ch. 2, §2.8 are what make the formula work at that scale, and this section uses them to find probabilities.

## 11.2.1 Which tool, and when

Three questions decide which tool to use: *how many objects are you choosing from, how many are you taking, and does the order matter?* The last one decides between the two formulas below.

Order matters when the positions are different from one another: first, second and third place; president, secretary and treasurer; the order of speakers at a meeting. Order does not matter when you end up with a set: a committee, a hand of cards, six lottery numbers. Choosing Ali then Ben gives the same committee as choosing Ben then Ali, and a count that treats those as different has counted the committee twice.

KEY · IB HL Choosing $r$ objects from $n$ distinct objects:

$$\text{order matters: } {}^nP_r = \frac{n!}{(n-r)!}, \qquad \text{order does not: } {}^nC_r = \frac{n!}{r!\,(n-r)!}.$$

*Every unordered selection corresponds to $r!$ ordered ones, which is the only difference between the two.*

CAUTION Decide this from the wording of the question. Read the question for a word that fixes an order: *arrange*, *rank*, *in a row*, *first and second*. Without one, it is usually a combination.

## 11.2.2 Probability by counting

Once both counts use the same tool, the ratio is the probability. Count the sample space with one method, count the event with the same method, and divide.

To count the event, split the objects into the groups the question names, choose the required number from each group with ${}^nC_r$, and multiply the counts, because the event needs one choice from each group. Many counting probability questions have this shape.

**Worked through.** To find the probability that a committee of 4, chosen at random from 7 girls and 5 boys, contains exactly 3 girls, note first that a committee is a set, so order does not matter and both counts are combinations. The sample space is every committee of 4 from the 12 students:

$$n(U) = {}^{12}C_4 = 495.$$

For the event, read the requirement as an *and*: 3 of the 7 girls *and* 1 of the 5 boys. And multiplies Ch. 2, §2.8, so the two counts are multiplied:

$$n(A) = {}^7C_3 \times {}^5C_1 = 35 \times 5 = 175.$$

So

$$P(\text{exactly 3 girls}) = \frac{175}{495} = \frac{35}{99} \approx 0.354.$$

Notice the shape of the event count: one combination for each group the question splits the people into, multiplied together.

**EXAMPLE 11.1**

Five cards are dealt from a standard pack of 52. Find the probability that exactly two of them are kings.

A hand is a set, so combinations again:

$$n(U) = {}^{52}C_5 = 2\,598\,960.$$

Split the pack into the two groups the question cares about: the 4 kings and the 48 non-kings. Take 2 from the first *and* 3 from the second, so the counts multiply:

$$n(A) = {}^{4}C_2 \times {}^{48}C_3 = 6 \times 17\,296 = 103\,776.$$

$$P(\text{exactly two kings}) = \frac{103\,776}{2\,598\,960} \approx 0.0399.$$

The 48 in the second factor matters: taking ${}^{4}C_2 \times {}^{50}C_3$ instead would let the other two kings into the hand, so it would count hands with three or four kings, some of them more than once.

TIP Write down $n(U)$ before anything else. It fixes what an outcome *is* for the rest of the question, and both counts must use that same definition of an outcome.

### 11.2.3 At least one, again

The complement of §11.1 is at its most useful here, because "at least one" spreads over many cases while "none" is a single count.

**EXAMPLE 11.2**

A box holds 20 components, of which 3 are defective. Four are drawn at random without replacement. Find the probability that at least one is defective.

Counting the cases directly means one defective, or two, or three: three separate counts. The complement is one count:

$$P(\text{none defective}) = \frac{{}^{17}C_4}{{}^{20}C_4} = \frac{2380}{4845} = \frac{28}{57}.$$

Therefore

$$P(\text{at least one defective}) = 1 - \frac{28}{57} = \frac{29}{57} \approx 0.509.$$

Three bad components in twenty, and a sample of four is already more likely than not to contain one. So even a small sample has a good chance of finding a defective component.

### 11.2.4 When order does matter

**Worked through.** Eight athletes run a final, and all orders of finishing are equally likely. To find the probability that a particular three of them take the first three places, in some order, note that here the positions are distinguishable, so the sample space is ordered:

$$n(U) = {}^{8}P_{3} = 336$$

ordered ways to fill the three medal positions. The three named athletes can fill them in $3! = 6$ orders, so

$$P = \frac{6}{336} = \frac{1}{56}.$$

The same answer comes from counting unordered: ${}^{8}C_{3} = 56$ possible medal-winning sets, one of which is the named three. Either tool works, provided the top and the bottom use the same one.

The counting principles, factorials, ${}^{n}P_{r}$ and ${}^{n}C_{r}$ are developed in Ch. 2, §2.8. The binomial distribution of Ch. 12 is this section's ${}^{n}C_{r}$ applied to repeated trials.

**YOUR TURN 11.2** Counting the Outcomes HL

**Choosing the counting tool**

**8.** In each part decide whether order matters, find $n(U)$ and then the probability.

(a) Three of nine books are chosen at random to take on holiday. Find the probability that one particular book is among them.

(b) Gold, silver and bronze medals are won by three of nine finalists, and all finishing orders are equally likely. Find the probability that Lena wins gold and her sister Mia wins silver.

(c) Five people sit in a row of five seats in a random order. Find the probability that Ana and Ben sit next to each other.

**Probability by counting groups**

**9.** A team of 5 is chosen at random from 6 women and 4 men.

(a) Find the number of possible teams.
(b) Find the probability that the team contains exactly 3 women.
(c) Find the probability that the team is all women.
(d) Find the probability that the team contains at least one man.

**10.** A box contains 15 light bulbs, of which 4 are faulty. Three are chosen at random without replacement.

(a) Find the probability that none is faulty.
(b) Find the probability that at least one is faulty.
(c) Find the probability that exactly two are faulty.

**When order matters**

**11.** A four-digit code is made from four different digits chosen from $1, 2, \dots, 9$, and every such code is equally likely.

(a) Find the probability that the digits of the code are in increasing order.
(b) Find the probability that the code starts with 1 and ends with 9.

**12.** Four of the 12 students in a club are chosen at random to be president, vice-president, secretary and treasurer. Find the probability that Maya and Leo are both given a post,

(a) counting ordered selections for both $n(U)$ and $n(A)$;
(b) counting unordered selections for both, and check that the answers agree.

**Concept check**

**13.** Three of nine books are chosen at random. To find the probability that one particular book is chosen, a student writes: "$n(A) = {}^8C_2 = 28$ *and* $n(U) = {}^9P_3 = 504$, *so the probability is* $\frac{28}{504} = \frac{1}{18}$." Identify the error, and find the correct probability.

# 11.3 Combining Events

Many questions involve more than one event. You want the probability that a student plays football *or* tennis, or that a card is a king *and* a heart. Two operations combine them: the **union** $A \cup B$ ("$A$ or $B$ or both") and the **intersection** $A \cap B$ ("$A$ and $B$ together").

The word *or* needs care, because everyday English and mathematics do not use it the same way. "Tea or coffee?" offers you one drink, not both. In mathematics $A \cup B$ always includes the case where both happen. This is called the **inclusive or**, and it is the reason the addition rule below needs a correction term. Questions that mean one but not both usually say so explicitly.

A **Venn diagram** makes the structure visible. The rectangle is the sample space, written $U$ and called the **universal set**: everything under discussion lies inside it. Each circle is an event, and every point of the diagram lies in exactly one of four regions. Combining an event with a complement names each region precisely.

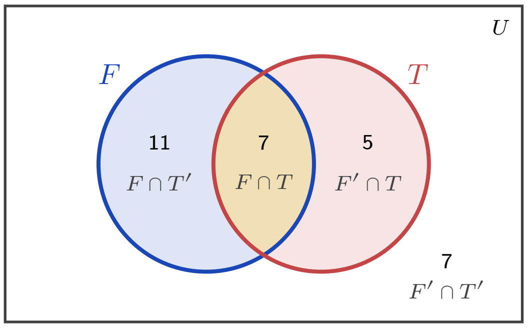


**Figure 11.1** The four regions of a two-event Venn diagram, with the class of 30 filled in. Football only is $F \cap T'$, both is $F \cap T$, tennis only is $F' \cap T$, and neither is $F' \cap T'$, the region outside both circles. The four counts add to 30.

Fill a Venn diagram from the middle outwards. Put the overlap in first, then subtract it from each circle total to get the "only" regions, then use the grand total to find the outside. Working the other way round double counts the overlap, which is the same mistake the addition rule is built to correct.

### 11.3.1 The addition rule

If you add $P(A)$ and $P(B)$ to find $P(A \cup B)$, you count the overlap twice: once in $A$, once in $B$. Subtracting it once corrects the double count.

KEY · IB

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

When two events cannot happen together, they are **mutually exclusive**: their intersection contains no outcomes at all, written $A \cap B = \emptyset$, where $\emptyset$ is the **empty set**. Then $P(A \cap B) = 0$ and the rule simplifies.

KEY · IB For mutually exclusive events:

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

**Worked through.** In a class of 30, 18 students play football, 12 play tennis, and 7 play both. For a student chosen at random, the probability of playing football or tennis comes from the addition rule directly:

$$P(F \cup T) = \frac{18}{30} + \frac{12}{30} - \frac{7}{30} = \frac{23}{30}.$$

The probability of playing neither is the complement of "football or tennis":

$$P(\text{neither}) = 1 - \frac{23}{30} = \frac{7}{30}.$$

The Venn diagram above shows the same thing: 11 play only football, 7 play both, 5 play only tennis, and 7 play neither, totalling 30.

**EXAMPLE 11.3**

A single card is drawn from a standard deck of 52. Find the probability that it is a king or a heart.

There are 4 kings and 13 hearts, but the king of hearts is both, so it is counted in each:

$$P(\text{king} \cup \text{heart}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}.$$

Forgetting to subtract the king of hearts gives $\frac{17}{52}$, which counts the king of hearts twice.

Three events need three circles, and the same instruction applies: start in the middle. The centre region belongs to all three events, and every other region is found by subtracting what has already been placed.

**EXAMPLE 11.4**

In a year group of 40 students, 20 study French, 18 study German and 15 study Spanish. Of these, 8 study French and German, 6 study German and Spanish, 7 study French and Spanish, and 3 study all three. A student is chosen at random. Find the probability that the student studies (a) exactly one of the three languages, (b) none of them.

Start with the 3 in the centre. The 8 who study French and German include those 3, so $8 - 3 = 5$ study those two and no other. In the same way $6 - 3 = 3$ study German and Spanish only, and $7 - 3 = 4$ study French and Spanish only.

Now the "only" regions. French has 20 altogether, of which $5 + 3 + 4 = 12$ are already placed, leaving 8 for French only. German gives $18 - (5 + 3 + 3) = 7$, and Spanish gives

$15-(4+3+3)=5$.

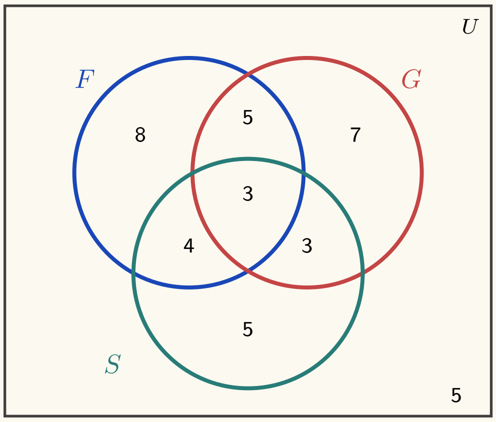


**Figure 11.2** The three-circle diagram, filled from the centre outwards. The seven inner counts total 35, so 5 of the 40 students lie outside all three circles.

(a) Exactly one language means the three outer regions: $8+7+5=20$, so

$$P(\text{exactly one})=\frac{20}{40}=\frac{1}{2}.$$

(b) The seven regions inside the circles total 35, so 5 students study none:

$$P(\text{none})=\frac{5}{40}=\frac{1}{8}.$$

No formula was used. Once the diagram is filled, every probability is a count divided by 40.

## 11.3.2 Sample space diagrams and two-way tables

When an experiment has two parts, drawing the whole sample space is often faster than any formula. A **sample space diagram** plots one part on each axis. For two dice, the 36 equally likely outcomes form a grid, and any event is just a set of grid points to count.

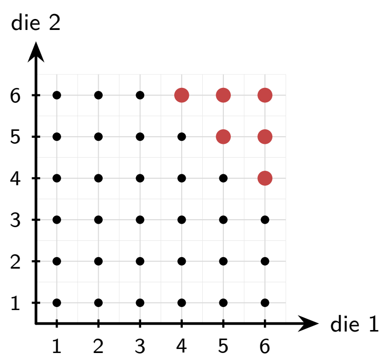


**Figure 11.3** The 36 equally likely outcomes of rolling two dice. The six larger scarlet points are the outcomes whose total is at least 10.

The six larger scarlet points are the outcomes with a total of at least 10. There are six of them, so

$$P(\text{total}\geq 10)=\frac{6}{36}=\frac{1}{6}.$$

No rule was needed, only a picture and a count.

A **two-way table** (a table of outcomes) does the same for grouped data: it splits a population by two characteristics at once, and every probability is a cell, row, or column total divided by the grand total.

**EXAMPLE 11.5**

The table shows the 100 students of a year group by age group and club membership.

| | Club | No club | Total |
|---|---|---|---|
| **Junior** | 14 | 26 | 40 |
| **Senior** | 21 | 39 | 60 |
| **Total** | 35 | 65 | 100 |

A student is chosen at random. Find the probability that the student (a) is a junior in a club, (b) is in a club or is a junior.

(a) Read the single cell where the two conditions meet:

$$P(J \cap C) = \frac{14}{100} = 0.14.$$

(b) Apply the addition rule with the row and column totals:

$$P(J \cup C) = \frac{40}{100} + \frac{35}{100} - \frac{14}{100} = \frac{61}{100} = 0.61.$$

The table already contains every count used in both calculations. We will return to this table in the next section, where it answers conditional questions just as directly.

**YOUR TURN 11.3** Combining Events

**The addition rule**

**14.**

(a) Given $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \cap B) = 0.2$, find $P(A \cup B)$.
(b) Events $A$ and $B$ are mutually exclusive, with $P(A) = 0.6$ and $P(B) = 0.3$. Find $P(A \cup B)$.
(c) Given $P(A \cup B) = 0.8$, $P(A) = 0.5$ and $P(B) = 0.4$, find $P(A \cap B)$.
(d) Given $P(A) = 0.3$, $P(B) = 0.4$ and $P(A \cup B) = 0.7$, determine whether $A$ and $B$ are mutually exclusive.

**15.**

(a) A card is drawn from a deck of 52. Find the probability that it is neither a heart nor a picture card (jack, queen or king).
(b) In a group of 40 people, 25 like coffee, 20 like tea and 10 like both. Find the probability that a person chosen at random likes coffee or tea.
(c) For the group in part (b), find the probability that the person likes neither.

**Venn diagrams**

**16.**

(a) Two events satisfy $n(A \text{ only}) = 8$, $n(B \text{ only}) = 5$, $n(A \cap B) = 4$ and $n(\text{neither}) = 3$. Find $P(A)$.
(b) Given $n(U) = 50$, $n(A) = 22$, $n(B) = 18$ and $n(A \cap B) = 7$, find $n(A \cap B')$ and $n(A' \cap B')$.

**17.** A survey of 50 households finds that 28 have a cat, 22 have a dog and 15 have a rabbit. Ten have a cat and a dog, 6 have a dog and a rabbit, 5 have a cat and a rabbit, and 3 have all three.

(a) Draw a Venn diagram showing the number of households in each region.
(b) A household is chosen at random. Find the probability that it has at least two of the three kinds of pet.
(c) Find the probability that it has none of the three.

**Sample space diagrams and two-way tables**

**18.** Two fair dice are rolled.

(a) Find the probability that the total is 7.
(b) Find the probability that the two numbers differ by exactly 1.
(c) Find the probability that the two dice show the same number.

**19.** The table shows 80 people classified by sex and by writing hand.

| | Left | Right | Total |
|---|---|---|---|
| **Male** | 7 | 33 | 40 |
| **Female** | 5 | 35 | 40 |
| **Total** | 12 | 68 | 80 |

One person is chosen at random.

(a) Find the probability that the person is left-handed.
(b) Find the probability that the person is female and right-handed.
(c) Find the probability that the person is female or left-handed.

**Concept check**

**20.** A card is drawn from a deck of 52. A student writes:
"$P(\textit{heart or king}) = \frac{13}{52} + \frac{4}{52} = \frac{17}{52}$." Identify the error, and find the correct probability.

## 11.4 Conditional Probability and Independence

Information changes probability. The chance that a random person has a particular illness is small, but the chance *given* that they tested positive is different. **Conditional probability** is the probability of one event once you know another has happened.

Knowing $B$ has occurred **reduces the sample space** to $B$ alone. Within that smaller set, you ask what fraction also lies in $A$.

KEY · IB

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) \neq 0$$

Read $P(A \mid B)$ as "the probability of $A$ given $B$." The denominator $P(B)$ is the new, restricted sample space; the numerator is the part of it where $A$ also holds.

**Worked through.** Using the class of 30 from before, suppose a student chosen at random is known to play football. To find the probability that they also play tennis, restrict to the 18 football players. Of those, 7 also play tennis:

$$P(T \mid F) = \frac{P(T \cap F)}{P(F)} = \frac{7/30}{18/30} = \frac{7}{18}.$$

Note this differs from $P(F \mid T) = \frac{7}{12}$. Conditioning is not symmetric: the order matters.

### 11.4.1 The multiplication rule

Rearranging the conditional formula gives a rule for the probability that two events both happen.

KEY

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

In words: the chance of both is the chance of the first times the chance of the second given the first. This is exactly how you move along the branches of a tree diagram, as the next section shows.

**Worked through.** A drawer holds nine pens, four of which are red. Two pens are taken at random, one after the other, without replacement. To find the probability that both are red, let $A$ be "the first pen is red" and $B$ "the second pen is red". Four of the nine pens are red, so $P(A) = \frac{4}{9}$.

For the second factor, assume $A$ has happened. Eight pens remain and three of them are red, so $P(B \mid A) = \frac{3}{8}$. The multiplication rule combines the two:

$$P(A \cap B) = P(A)\,P(B \mid A) = \frac{4}{9} \times \frac{3}{8} = \frac{12}{72} = \frac{1}{6}.$$

The second factor is $\frac{3}{8}$ and not $\frac{4}{9}$ because the first pen is not put back: removing a red pen changes the number of red pens and the total together, so both parts of the fraction drop by one.

### 11.4.2 Independence

Two events are **independent** when knowing one tells you nothing about the other, so $P(A \mid B) = P(A)$. Substituting this into the multiplication rule gives a cleaner, symmetric test.

KEY · IB $A$ and $B$ are **independent** if and only if

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

Provided $0 < P(B) < 1$, this is equivalent to $P(A \mid B) = P(A) = P(A \mid B')$, which is not itself printed in the booklet. To test independence, compute both sides of either form and compare. The conditional form states directly what independence means: whether $B$ happened or did not happen, the probability of $A$ is unchanged.

**EXAMPLE 11.6**

A fair coin and a fair die are tossed together. Let $H$ be "heads" and $S$ be "rolling a six." Are $H$ and $S$ independent?

Here $P(H) = \frac{1}{2}$ and $P(S) = \frac{1}{6}$. The outcome "heads and a six" is one of the 12 equally likely combined outcomes:

$$P(H \cap S) = \frac{1}{12} = \frac{1}{2} \times \frac{1}{6} = P(H)\,P(S).$$

The two sides match, so the events are independent, which fits intuition: a coin cannot influence a die.

The two-way table of §11.3 passes the same test. There, $P(C \mid J) = \frac{14}{40} = 0.35$ and $P(C) = \frac{35}{100} = 0.35$: knowing a student is a junior does not change the chance they are in a club, so age group and club membership are independent.

Exact independence like this is unusual in real data. Two characteristics recorded on the same group of people are often linked at least a little, and then the test fails. The calculation is identical when it comes out that way, and so is the shape the answer must take: the two numbers first, then the verdict.

**EXAMPLE 11.7**

The table classifies the 200 students of a year group by whether they take music and whether they take drama.

| | Drama | No drama | Total |
|---|---|---|---|
| **Music** | 30 | 50 | 80 |
| **No music** | 20 | 100 | 120 |
| **Total** | 50 | 150 | 200 |

A student is chosen at random. Determine whether the events $M$ (takes music) and $D$ (takes drama) are independent.

Read the two single-event probabilities from the totals, and the intersection from the cell where the two conditions meet:

$$P(M) = \frac{80}{200} = 0.4, \qquad P(D) = \frac{50}{200} = 0.25, \qquad P(M \cap D) = \frac{30}{200} = 0.15.$$

Compare the product with the intersection:

$$P(M)\,P(D) = 0.4 \times 0.25 = 0.10 \neq 0.15 = P(M \cap D),$$

so $M$ and $D$ are **not** independent.

The conditional form says the same thing and adds the direction. Restricting to the 80 music students gives $P(D \mid M) = \frac{30}{80} = 0.375$, against $P(D) = 0.25$ for the year group as a whole, so knowing a student takes music raises the chance of drama from a quarter to three eighths. Independence is an exact condition, so any difference between the two numbers, however small, means the events are not independent. Write both numbers down and then the verdict; a bare "not independent" with only one side computed shows nothing.

CAUTION Mutually exclusive and independent are different ideas. If two events with non-zero probability are mutually exclusive, then one happening prevents the other completely, so $P(A \mid B) = 0 \neq P(A)$ and the events cannot be independent. Test independence with $P(A \cap B) = P(A)\,P(B)$; do not assume it.

## YOUR TURN 11.4 Conditional Probability and Independence

### Conditional probability

**21.**

(a) Given $P(A \cap B) = 0.2$ and $P(B) = 0.5$, find $P(A \mid B)$.

(b) Given $P(A) = 0.6$ and $P(B \mid A) = 0.3$, find $P(A \cap B)$.

(c) Given $P(A) = 0.5$, $P(B) = 0.4$ and $P(A \cup B) = 0.7$, find $P(A \mid B)$.

(d) Two cards are drawn from a deck of 52 without replacement. Find the probability that both are kings.

(e) A card is drawn from a deck of 52. Given that it is a picture card, find the probability that it is a king.

(f) Two fair dice are rolled. Given that the first die shows an even number, find the probability that the total is 8.

**22.** The table classifies 120 patients by the treatment they received and whether their condition improved.

| | Improved | Not improved | Total |
|---|---|---|---|
| **New treatment** | 36 | 24 | 60 |
| **Standard treatment** | 30 | 30 | 60 |
| **Total** | 66 | 54 | 120 |

One patient is chosen at random. Let $I$ be "the condition improved" and $N$ be "received the new treatment".

(a) Find $P(I \mid N)$.
(b) Find $P(N \mid I)$.
(c) Find $P(I \mid N')$, and use it to determine whether $I$ and $N$ are independent.

**Independence**

**23.**

(a) Events $A$ and $B$ are independent, with $P(A) = 0.5$ and $P(B) = 0.4$. Find $P(A \cap B)$.
(b) Given $P(A) = 0.3$, $P(B) = 0.5$ and $P(A \cap B) = 0.2$, determine whether $A$ and $B$ are independent.
(c) Events $A$ and $B$ have $P(A) = 0.3$, $P(B) = 0.2$ and $P(A \cup B) = 0.5$. Show that $A$ and $B$ are mutually exclusive, and determine whether they are independent.

**24.** In a class of 25 students, 15 study Spanish, 10 study art and 6 study both.

(a) Given that a student studies Spanish, find the probability that they also study art.
(b) Given that a student studies art, find the probability that they also study Spanish.
(c) Determine whether studying Spanish and studying art are independent.

**25.** Events $A$ and $B$ have $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \mid B) = 0.4$. Find $P(A \mid B')$, and state whether $A$ and $B$ are independent.

**Concept check**

**26.** Events $A$ and $B$ have $P(A) = 0.3$ and $P(B) = 0.4$, and are mutually exclusive. A student says: "*mutually exclusive events do not affect each other, so $A$ and $B$ are independent.*" Explain the difference between *mutually exclusive* and *independent*, and show that $A$ and $B$ are not independent.

# 11.5 Tree Diagrams

When events happen in sequence, a **tree diagram** organises every path. Each branch carries the probability of that step, and two rules then do the whole calculation: **multiply along the branches** of a single path (the multiplication rule), and **add across paths** that satisfy the condition (the paths are mutually exclusive).

When each draw is made **with replacement**, the object goes back before the next draw, every stage repeats the same probabilities, and the stages are independent. Care is needed with sequences **without replacement**, where the first outcome changes the probabilities on the second branches.

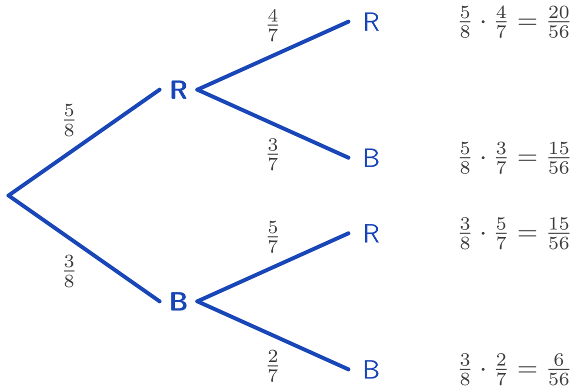


**Figure 11.4** A tree diagram for drawing two marbles from 5 red and 3 blue without replacement. The probabilities on the second stage differ from the first, because one marble has already been removed. Multiply along a path; add across paths.

**Worked through.** For the bag in the figure, with 5 red and 3 blue marbles and two drawn without replacement, the second branches change because after drawing one red, only 4 reds remain out of 7 marbles. For both red, multiply along the top path:

$$P(\text{both red}) = \frac{5}{8} \times \frac{4}{7} = \frac{20}{56} = \frac{5}{14}.$$

"One of each" covers two paths, red-then-blue and blue-then-red; add them:

$$P(\text{one of each}) = \frac{5}{8} \cdot \frac{3}{7} + \frac{3}{8} \cdot \frac{5}{7} = \frac{15+15}{56} = \frac{30}{56} = \frac{15}{28}.$$

For "at least one blue," use the complement, which is "both red":

$$P(\text{at least one blue}) = 1 - \frac{5}{14} = \frac{9}{14}.$$

All three rules appear here: multiply along a path, add across alternatives, and take the complement for "at least one."

Putting the marble back changes every second-stage branch. The bag returns to 5 red and 3 blue, so both stages read $\frac{5}{8}$ and $\frac{3}{8}$, and the two draws are independent.

**EXAMPLE 11.8**

The same bag holds 5 red and 3 blue marbles, but now the first marble is replaced before the second is drawn. Find the probability that both are red, and compare with the answer without replacement.

Every branch keeps its original probability, so

$$P(\text{both red}) = \frac{5}{8} \times \frac{5}{8} = \frac{25}{64} \approx 0.391.$$

Without replacement the answer was $\frac{5}{14} \approx 0.357$, which is smaller. Removing a red marble leaves fewer reds for the second draw, so the second red becomes less likely. Replacing it removes that effect.

Independent repeated trials, such as draws with replacement, give a useful shortcut. When the same trial is repeated $n$ times independently, "at least one success" has a complement that is a single product, because "no successes at all" means failure every time.

KEY For $n$ independent trials in which the event has probability $p$ each time,

$$P(\text{at least one}) = 1 - (1-p)^n.$$

**EXAMPLE 11.9**

A machine part fails on any given day with probability $0.02$, independently of other days. Find the probability that it fails at least once during a 30-day month.

Counting the ways to fail once, twice, three times and so on would need a separate calculation for each number of failures. The complement is a single product: the part survives all 30 days, each day with probability $0.98$.

$$P(\text{at least one failure}) = 1 - (0.98)^{30} = 1 - 0.545 = 0.455.$$

A part that is safe on 98 days out of 100 still has close to an even chance of failing within a month. A small daily probability, repeated over many days, gives a much larger probability of at least one failure.

### 11.5.1 Three stages without replacement

A third draw adds a third stage and nothing else. Each branch still carries the probability of that step given everything before it, and without replacement the denominator falls by one at every draw.

What does change is the counting. An event such as "exactly two of the three" now corresponds to three paths, not one or two, because the odd draw out can come first, second or last. Identify those paths before computing anything. Each of these paths uses the same numerators and the same denominators in a different order, so all of them have the same probability. Evaluate one path and multiply by the number of orders.

**Worked through.** A bag contains 7 green and 3 yellow counters. Three are drawn at random without replacement. To find the probability that all three are green, follow one path, with the bag shrinking under it:

$$P(\mathsf{GGG}) = \frac{7}{10} \times \frac{6}{9} \times \frac{5}{8} = \frac{210}{720} = \frac{7}{24}.$$

For the probability that exactly two are green, note that exactly two green means one yellow, and the yellow may be drawn first, second or third. The three paths are YGG, GYG and GGY:

$$P(\mathsf{YGG}) = \frac{3}{10} \cdot \frac{7}{9} \cdot \frac{6}{8}, \qquad P(\mathsf{GYG}) = \frac{7}{10} \cdot \frac{3}{9} \cdot \frac{6}{8}, \qquad P(\mathsf{GGY}) = \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{3}{8}.$$

Every one of them is $\frac{126}{720}$, so adding across the three paths gives

$$P(\text{exactly two green}) = 3 \times \frac{126}{720} = \frac{378}{720} = \frac{21}{40} = 0.525.$$

The three products agree because each is built from the same numerators, $7$, $6$ and $3$, over the same denominators $10 \times 9 \times 8$; only the order of the factors differs. Count the orders first, because a missed path gives an answer that is too small.

**YOUR TURN 11.5** Tree Diagrams

**Two draws, with and without replacement**

**27.**

(a) A bag holds 3 white and 2 black balls. Two are drawn with replacement. Find the probability that both are white.

(b) Repeat part (a) for two balls drawn without replacement.

(c) A box holds 6 working and 4 defective bulbs. Two are drawn without replacement. Find the probability that exactly one is defective.

**Different probabilities on the second branches**

**28.** The probability that it rains on a given morning is $0.3$. If it rains, the probability that Sam is late for school is $0.6$; if it does not rain, the probability is $0.1$.

(a) Find the probability that Sam is late.

(b) Find the probability that it rains and Sam is not late.

**Repeated independent trials**

**29.**

(a) A machine works on any given day with probability $0.9$, independently of other days. Find the probability that it works on two consecutive days.

(b) Ali and Bea take turns to roll a fair die, Ali first, and the first to roll a six wins. Find the probability that Ali wins with his second roll.

(c) Using the machine in part (a), find the probability that it fails on at least one of five consecutive days.

**30.** A component works with probability $0.95$, independently of other components. A circuit contains four of them.

(a) Find the probability that all four components work.

(b) Find the probability that exactly one of the four components fails.

**Three draws without replacement**

**31.** A bag holds 4 red and 6 blue counters. Three are drawn without replacement.

(a) Find the probability that all three are blue.

(b) Find the probability that exactly two are red.

**Concept check**

**32.** A bag holds 3 red and 2 blue counters. Two are drawn without replacement. A student writes: "$P(\textit{one of each colour}) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$." Identify the error, and find the correct probability.

# 11.6 Bayes' Theorem HL

We can now return to the screening test from the start of the chapter. About one person in a hundred has the disease. The test correctly identifies $99\%$ of those who have it, and wrongly reports a positive for $5\%$ of those who do not. A positive result has just arrived.

The difficulty is that the test gives you $P(\text{positive} \mid \text{disease})$, but what you actually want is the reverse, $P(\text{disease} \mid \text{positive})$. Conditional probability is not symmetric, so these are different numbers. **Bayes' theorem** is the tool that reverses the condition.

It is built from two facts you already have. The conditional formula gives $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$. The multiplication rule rewrites the numerator as $P(A \cap B) = P(B)\,P(A \mid B)$, and the denominator can be split according to whether $B$ happened or not: $P(A) = P(B)P(A \mid B) + P(B')P(A \mid B')$. Putting these together:

KEY · IB HL

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

Read the letters the way the booklet intends them: $B$ is the thing you want to know about, and $A$ is the evidence that has arrived. In the screening test, $B$ is having the disease and $A$ is the positive result.

The numerator is the single path through $B$; the denominator is every path that leads to $A$. You are asking what fraction of all the ways to observe $A$ passed through $B$.

That description also gives the method. Written out in full the formula is long, but a tree diagram carries the same information in a form that is quicker to set up. Draw the first stage, $B$ and $B'$. Draw the second stage, $A$ given each of them. Multiply along every path. Then divide the path that passes through $B$ by the total of all the paths that end in $A$.

TIP Do not memorise the Bayes formula as a string of symbols. Draw the tree, multiply along the paths, and divide the path you want by the total of the paths that match the evidence. It gives the same answer, it shows every step, and a correct method earns the same marks.

The probability of $B$ before any evidence arrives, $P(B)$, is called the **base rate**. When the base rate is small, the few true positives can be outnumbered by false positives from the much larger group without the condition, so $P(B \mid A)$ can be small even for an accurate test.

**Worked through.** For the screening test, let $D$ be "has the disease" and $+$ be "tests positive." We know $P(D) = 0.01$, $P(+ \mid D) = 0.99$, and $P(+ \mid D') = 0.05$. Draw the tree first, then multiply along every path.

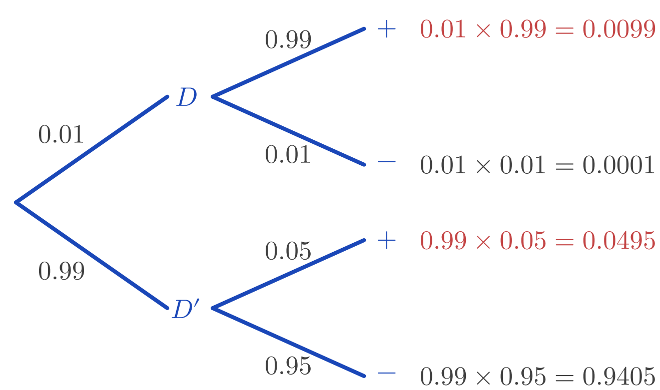


**Figure 11.5** The screening test as a tree. Two paths end in a positive result, shown in scarlet. Bayes' theorem divides the scarlet path through $D$ by the total of both scarlet paths.

Only the two scarlet paths match the evidence, so they are the only ones that count. The path through $D$ divided by the total of both gives

$$P(D \mid +) = \frac{(0.01)(0.99)}{(0.01)(0.99) + (0.99)(0.05)} = \frac{0.0099}{0.0099 + 0.0495} = \frac{0.0099}{0.0594} \approx 0.167.$$

That is the formula and the tree saying the same thing: the numerator is one path, the denominator is every path that ends in a positive.

So a person who tests positive has only about a 17% chance of having the disease. The reason is the base rate. The disease is so rare that the few true positives are greatly outnumbered by the false positives, which come from the very large healthy majority. The test is good; the disease is simply very rare.

**EXAMPLE 11.10**

A city runs two taxi companies. $85\%$ of its taxis are Green and $15\%$ are Blue. A taxi is involved in a night-time accident and a witness says it was Blue. Tested under the same conditions, the witness identifies each colour correctly $80\%$ of the time. Find the probability that the taxi really was Blue.

The evidence is the witness's statement, not the colour of the taxi, so put the colour on the first stage and the statement on the second.

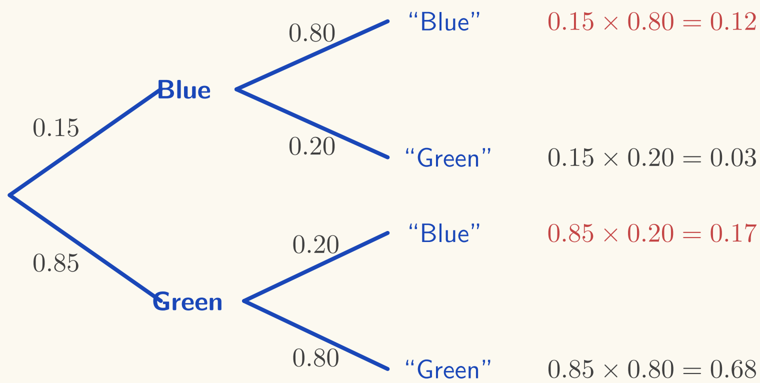


**Figure 11.6** The witness problem. The two scarlet paths are the ones on which the witness says "Blue". Note that the larger of the two comes from a Green taxi being misreported.

The witness said "Blue", so only the two scarlet paths remain:

$$P(\text{Blue} \mid \text{"Blue"}) = \frac{0.12}{0.12 + 0.17} = \frac{0.12}{0.29} \approx 0.414.$$

A witness who is right $80\%$ of the time reports "Blue", and the taxi is still more likely to have been Green. There are so many more Green taxis that the $20\%$ of them misreported as Blue $(0.17)$ outnumbers the $80\%$ of Blue taxis reported correctly $(0.12)$.

This is the same base-rate effect as the screening test, in a setting with no medicine in it. An answer of $80\%$ is tempting, because the witness's reliability is vivid and the proportion of Blue taxis is not. The tree keeps both in view.

### 11.6.1 Three alternatives

When the first stage has three possibilities instead of two, the denominator simply sums over all three paths. The structure is identical, and the booklet prints this version too.

KEY · IB HL

$$P(B_i \mid A) = \frac{P(B_i)\,P(A \mid B_i)}{P(B_1)\,P(A \mid B_1) + P(B_2)\,P(A \mid B_2) + P(B_3)\,P(A \mid B_3)}$$

Here $B_1$, $B_2$ and $B_3$ are the three possibilities at the first stage, and $B_i$ is whichever one you are asking about. Nothing has changed except the number of terms underneath.

**EXAMPLE 11.11**

Three machines produce all the parts in a factory: machine A makes 50%, B makes 30%, C makes 20%. Their defect rates are 1%, 2%, and 3% respectively. A part is found to be defective. Find the probability it came from machine C.

Weight each machine's defect rate by its share of production, then take C's share of the total:

$$\begin{aligned} P(C \mid \text{def}) &= \frac{(0.20)(0.03)}{(0.50)(0.01) + (0.30)(0.02) + (0.20)(0.03)} \\ &= \frac{0.006}{0.005 + 0.006 + 0.006} = \frac{0.006}{0.017} \approx 0.353. \end{aligned}$$

Machine C makes only a fifth of the parts but, because it is the least reliable, accounts for over a third of the defects.

**YOUR TURN 11.6** Bayes' Theorem HL

**Two alternatives**

**33.**

(a) The probability of rain on a given morning is $0.3$. If it rains, the probability of being late is $0.6$; if it does not, the probability is $0.1$. A person is late. Find the probability that it rained.

(b) A disease affects 2% of a population. A test is positive for 95% of those who have it and for 10% of those who do not. Given a positive result, find the probability of having the disease.

(c) Factory A makes 60% of all items, of which 2% are defective; factory B makes the other 40%, of which 5% are defective. Given that an item is defective, find the probability it came from factory B.

**34.**

(a) In a mailbox, 30% of messages are spam. The word "free" appears in 90% of spam messages and in 5% of the others. A message contains the word "free". Find the probability that it is spam.

(b) A box contains two coins: one fair, and one with heads on both sides. A coin is chosen at random and tossed once, showing heads. Find the probability that it is the two-headed coin.

(c) Of the students in a year group, 70% revise for a test. Of those who revise, 80% pass; of those who do not, 30% pass. A student passes. Find the probability that the student revised.

**Three alternatives**

**35.** Three suppliers provide the components used by a factory. Supplier A provides 40%, supplier B 35% and supplier C 25%. Of their components, 1%, 3% and 4% respectively are defective.

(a) Find the probability that a component chosen at random is defective.

(b) Given that a component is defective, find the probability that it came from supplier B.

(c) Given that a component is defective, find the probability that it came from supplier A.

(d) Supplier A provides the most components but accounts for the fewest defective ones. Explain why.

**Concept check**

**36.** A disease affects 1% of a population. A test is positive for 95% of people who have the disease and for 5% of people who do not. A student says: "*the test is 95% accurate, so a person who tests positive has a 95% chance of having the disease.*"

(a) Explain the difference between $P(\text{positive} \mid \text{disease})$ and $P(\text{disease} \mid \text{positive})$.

(b) Find $P(\text{disease} \mid \text{positive})$, and explain why it is so much smaller than $0.95$.

Chapter 11 · Probability

# Reflections

TOK Bayes' theorem describes how a belief should be updated when new evidence arrives: start from a *prior*, the probability believed before the evidence, weigh the evidence, and arrive at a *posterior*, the probability believed afterwards. Some philosophers argue this is not just a formula but a model of knowledge itself. If that is right, the screening-test result of §11.6 is not special to medicine: evidence means little until you account for how rare the thing you are testing for is. Is Bayes' theorem a description of how people reason, or a rule for how they ought to?

TOK We say a fair die has probability $\frac{1}{6}$ because of its symmetry, and we say a drawing pin lands point-up with probability $0.39$ because that is what 200 drops showed. The first number was never measured; the second was never proved.

Are these two the same kind of claim? And when a measured value disagrees with a calculated one, which of the two should be revised?

IM Probability theory is usually dated to 1654. That year the French writer Antoine Gombaud, the Chevalier de Méré, put several gambling problems to Blaise Pascal, among them how to divide the stakes fairly when a game is stopped before it ends. Pascal's letters with Pierre de Fermat on these problems set out many of the methods of probability theory, although Gerolamo Cardano had written about dice about a century earlier. A subject now central to physics, finance, and medicine grew out of questions about games of chance.

RW Ignoring the base rate has real costs. Courts have been misled when a "one-in-a-million" DNA match probability was presented as if it were the probability that the defendant was innocent, without asking how many people in the population would also match by chance. Mass screening programmes for rare cancers can produce more false alarms than true detections. Understanding $P(D \mid +)$ versus $P(+ \mid D)$ is not only an examination topic. It changes verdicts and medical decisions.

# End-of-Chapter Problems
# Chapter 11: Probability

**1.** At a school, 60% of students study French and 25% study both French and Spanish. Find the probability that a student studies Spanish, given that they study French. [3]

**2.** Events $A$ and $B$ are independent. Given $P(A) = 0.4$ and $P(A \cup B) = 0.7$, find $P(B)$. [4]

**3.** A biased four-sided spinner numbered 1 to 4 was spun 250 times. The results were:

| **Number** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Frequency** | 95 | 60 | 55 | 40 |

(a) Estimate the probability that the spinner lands on 1. [2]
(b) Estimate the probability that the spinner lands on an odd number. [2]
(c) The spinner is spun a further 60 times. Find the expected number of times it lands on 4. [2]

**4.** Events $A$ and $B$ satisfy $P(A) = 0.5$ and $P(B) = 0.3$.

(a) Find $P(A \cup B)$ if $A$ and $B$ are independent. [3]
(b) Find $P(A \cup B)$ if $A$ and $B$ are mutually exclusive. [2]
(c) Explain why $A$ and $B$ cannot be both independent and mutually exclusive here. [2]

**5.** In a class of 50 students, 28 study Mathematics ($M$), 20 study Art ($A$), and 10 study neither.

(a) Find the number of students who study both Mathematics and Art. [3]
(b) Find the probability that a randomly chosen student studies only Art. [2]
(c) Find $P(A \mid M)$. [2]

**6.** The two-way table shows whether 100 people exercise regularly.

| | **Exercises** | **Does not** | **Total** |
|---|---|---|---|
| **Male** | 30 | 20 | 50 |
| **Female** | 24 | 26 | 50 |
| **Total** | 54 | 46 | 100 |

(a) Find the probability that a person is male and exercises. [2]
(b) Find the probability that a person exercises, given that they are female. [2]
(c) Determine whether the events "male" and "exercises" are independent. [3]

**7.** In a class of 40 students, the numbers studying two activities $A$ and $B$ are: only $A$ is $2x$, both is $x$, only $B$ is $x+4$, and neither is 8.

(a) Find the value of $x$. [3]
(b) Find $P(A \cap B)$. [1]
(c) Find $P(A \cup B)$. [2]

**8.** Events $A$ and $B$ satisfy $P(A) = 0.45$, $P(B) = 0.4$, and $P((A \cup B)') = 0.25$.

(a) Find $P(A \cap B)$. [2]
(b) Find the probability that $A$ occurs but $B$ does not. [1]
(c) Find $P(A \mid B)$. [2]
(d) Determine whether $A$ and $B$ are independent. [2]

**9.** Two fair dice are rolled and the scores noted.

(a) Find the probability that the sum is a prime number. [2]
(b) Find the probability that at least one die shows a 6. [2]
(c) Given that at least one die shows a 4, find the probability that the sum is 7. [3]

**10.** A bag contains 7 red and 3 green sweets. Two are drawn without replacement.

(a) Draw a tree diagram for the two draws. [3]
(b) Find the probability that both sweets are red. [2]
(c) Find the probability that both sweets are the same colour. [3]
(d) Find the probability that at least one sweet is green. [2]

**11.** Events $A$, $B$, and $C$ are mutually exclusive and exhaustive, with $P(A) = 2k$, $P(B) = 3k$, and $P(C) = k + 0.1$.

(a) Find the value of $k$. [2]
(b) Find $P(A \cup B)$. [2]
(c) Find $P(A \mid B')$. [3]

**12.** A fair four-sided die (numbered 1–4) and a fair six-sided die (numbered 1–6) are rolled together, and the product of the two scores is recorded.

(a) Represent the outcomes in a sample space diagram. [2]
(b) Find the probability that the product is 12. [2]
(c) Find the probability that the product is odd. [2]

(d) Given that the product is even, find the probability that the four-sided die shows a 4. [3]

**13.** A football team's key striker is fit for a match with probability 0.85. The team wins with probability 0.7 if she is fit, and with probability 0.4 if she is not.

(a) Draw a tree diagram to represent this information. [2]
(b) Find the probability that the team wins the match. [2]
(c) Given that the team won, find the probability that the striker was fit. [3]

**14.** The 120 members of a sports club are classified by age group and by whether they wear glasses: 18 of the 50 adults and 12 of the 70 juniors wear glasses.

(a) Display this information in a two-way table, including totals. [2]
(b) Find the probability that a randomly chosen member is a junior who does not wear glasses. [1]
(c) Given that a member wears glasses, find the probability that the member is an adult. [2]
(d) Determine, using conditional probability, whether the events "adult" and "wears glasses" are independent. [3]

**15.** An airline routes 50% of its flights through airport A, 30% through airport B, and 20% through airport C. The probabilities that a flight is delayed are 0.06, 0.03, and 0.10 respectively.

(a) Find the probability that a randomly chosen flight is delayed. [3]
(b) Given that a flight was delayed, find the probability that it was routed through airport A. [3]

**16.** Three suppliers provide components: X supplies 25%, Y supplies 45%, and Z supplies 30%. The probabilities that a component from X, Y and Z is defective are $0.04$, $0.02$ and $p$ respectively. Overall, $3.4\%$ of all components are defective.

(a) Find the value of $p$. [3]
(b) Given that a component is defective, find the probability that it came from X. [3]

**17.** A condition affects 5% of a population. A screening test is positive for 90% of those who have it, and for 15% of those who do not.

(a) Find the probability that a randomly chosen person tests positive. [3]
(b) Given a positive result, find the probability the person has the condition. [3]
(c) Comment on what your answer says about screening for rare conditions. [2]

**18.** A box contains 4 red and 2 green counters. Three counters are drawn at random without replacement.

(a) Find the probability that all three counters are red. [2]
(b) Given that at least one counter is green, find the probability that exactly one is green. [3]
(c) Find the probability that the third counter drawn is the second green one. [2]

**19.** Events $A$ and $B$ satisfy $P(A) = 0.6$, $P(A \mid B) = 0.6$, and $P(A \cup B) = 0.84$.

(a) Explain why $A$ and $B$ are independent. [1]
(b) Find $P(B)$. [3]
(c) Write down $P(B \mid A')$, justifying your answer. [2]

**20.** A family has two children. Assume each child is equally likely to be a boy or a girl, independently.

(a) List the sample space. [1]
(b) Given that at least one of the children is a boy, find the probability that both are boys. [3]
(c) Given instead that the *elder* child is a boy, find the probability that both are boys. [2]
(d) Explain why the answers to (b) and (c) differ. [1]

**21.** In a survey of 100 people, 50 like tea, 60 like coffee, and 35 like juice. Also, 30 like tea and coffee, 20 like tea and juice, 25 like coffee and juice, and 10 like all three.

(a) Find the number who like at least one of the three drinks. [3]
(b) Find the number who like none. [1]
(c) Find the number who like exactly one drink. [3]

**22.** GDC An archer hits the target with probability 0.9 on each shot, independently of all other shots.

(a) Find the probability that her first miss is on her third shot. [2]
(b) Find the probability that she hits at least four of her first five shots. [2]
(c) Find the least number of shots for which the probability of at least one miss exceeds 0.95. [3]

**23.** GDC A system contains three components $A$, $B$ and $C$. Components $A$ and $B$ are connected in parallel, and this pair is connected in series with $C$, so the system works when $C$ works and at least one of $A$ and $B$ works. Each component works with probability $p$, independently of the others.

(a) Show that the probability that the system works is $p(1 - (1-p)^2)$. [2]
(b) Find the probability that the system works when $p = 0.9$. [1]
(c) Find the value of $p$ for which the probability that the system works is $0.95$. [2]
(d) A fourth component $D$, which also works with probability $p$ independently, is connected in parallel with $C$. Find the probability that the system now works when $p = 0.9$. [2]

**24.** GDC A disease affects 0.5% of a population. A test is positive for 98% of those with the disease and for 3% of those without. A person tests positive *twice*, the two tests being independent given disease status.

(a) Find the probability the person has the disease after one positive test. [3]
(b) Find the probability the person has the disease after two positive tests. [4]
(c) Comment on the effect of the second test. [1]

**25.** Events $A$ and $B$ are independent. Prove that $A$ and $B'$ are also independent, that is, show $P(A \cap B') = P(A)\,P(B')$. [5]

**26.** GDC Ana and Ben take turns to roll a fair six-sided die, with Ana rolling first. The first player to roll a six wins the game.

(a) Write down the probability that Ana wins with her first roll. [1]
(b) Find the probability that Ana wins with her third roll. [2]
(c) Show that the probability that Ana wins the game is $\dfrac{6}{11}$. [3]
(d) Find the probability that neither player has rolled a six after each has rolled ten times. [2]

**27.** Four cards are dealt at random from a standard pack of 52. Find the probability that the four cards are all from different suits. [5]

**28.** A committee of 4 is chosen at random from 5 teachers and 4 students.

(a) Find the number of possible committees. [2]
(b) Find the probability that the committee has exactly 2 teachers and 2 students. [3]
(c) Find the probability that the committee contains at least one teacher and at least one student. [4]
(d) Explain why part (c) is quicker through its complement than by adding the cases directly. [2]

**29.** GDC A lottery draws 5 numbers at random from 40. A ticket also carries 5 numbers.

(a) Find the number of possible draws. [2]
(b) Find the probability that a ticket matches all 5 numbers. [2]
(c) Find the probability that a ticket matches exactly 3 of the 5. [4]
(d) The organisers propose drawing 5 from 45 instead. Find the probability that a ticket matches all five under this rule, and compare it with your answer to part (b). [3]

**30.** EXACT The letters of the word PARALLEL are arranged in a row, all arrangements being equally likely.

(a) Explain why the number of distinct arrangements is $\dfrac{8!}{2!\,3!}$, and evaluate it. [3]
(b) Find the probability that an arrangement begins with the letter P. [4]

(c) Find the probability that the three Ls are together. [4]

**31.** Of the $200$ students at a school, $130$ cycle to school $(C)$, $90$ travel by train $(T)$, and $40$ do both. The remaining students walk.

(a) Draw a Venn diagram showing the number of students in each of the four regions. [3]
(b) Find the probability that a student chosen at random walks to school. [2]
(c) Find $P(C \mid T)$. [3]
(d) Determine whether $C$ and $T$ are independent, justifying your answer. [3]
(e) Two students are chosen at random, without replacement. Find the probability that both of them walk. [3]
(f) State whether your answer to part (e) would be larger or smaller if the two students were chosen with replacement, and explain why. [2]

**32.** GDC A machine produces components, of which $6\%$ are defective. Each component is inspected. The inspection rejects a defective component with probability $0.95$, and wrongly rejects a component that is not defective with probability $0.02$.

(a) Draw a tree diagram to represent this information. [3]
(b) Find the probability that a randomly chosen component is rejected. [3]
(c) Find the probability that a rejected component was in fact not defective. [3]
(d) Find the probability that a component which passes inspection is defective. [3]
(e) In a batch of $5000$ components, find the expected number that are defective but pass inspection. [2]
(f) Using your answers to parts (d) and (e), comment on how well the inspection protects a customer. [2]

**33.** A box contains $10$ batteries, of which $3$ are flat. Batteries are taken from the box one at a time, without replacement, and tested until a working one is found. Let $N$ be the number of batteries tested.

(a) Find $P(N = 1)$. [2]
(b) Find $P(N = 2)$. [3]
(c) Find $P(N \leq 3)$. [4]
(d) Explain why $N$ can never be greater than $4$. [2]
(e) Find $P(N = 3)$ and $P(N = 4)$, and verify that the four probabilities sum to $1$. [4]

# Challenge Questions

**34.** GDC **The birthday problem.** In a group of $n$ people, assume that every birthday is equally likely among 365 days, that different people's birthdays are independent, and ignore leap years.

(a) Explain why, for $n = 3$, the probability that all three birthdays differ is $\dfrac{364}{365} \times \dfrac{363}{365}$, and write down an expression for the probability that all $n$ birthdays differ. [4]

(b) Hence find, to three decimal places, the probability that at least two people share a birthday when $n = 10$, $n = 23$ and $n = 50$. [3]

(c) Find the smallest $n$ for which the probability of a shared birthday exceeds $\frac{1}{2}$. [2]

(d) Two tempting guesses give much larger answers. The first is $183$ people, about half the number of days. The second is the number of people needed for the probability that someone shares *your* birthday to exceed $\frac{1}{2}$. Show that the second guess gives $254$ people (you and $253$ others). Explain why each guess answers a different question from part (c), and why the count that matters is the number of *pairs*, ${}^nC_2$. Evaluate ${}^nC_2$ for $n = 23$. [5]

(e) Each pair matches with probability $\frac{1}{365}$. Find the expected number of matching pairs in a group of $23$, and comment on your answer. [2]

(f) A school issues every student a random four-digit code, so there are $10\,000$ equally likely codes. Find the smallest number of students for which two students sharing a code is more likely than not. [3]

(g) Now let there be $N$ equally likely values. Using $1 - x \approx e^{-x}$ for small $x$, show that the probability that all $n$ values differ is approximately $e^{-n(n-1)/(2N)}$. Hence show that, for large $N$, the smallest $n$ for which a match is more likely than not is about $\sqrt{2N \ln 2}$. Evaluate this for $N = 365$ and $N = 10\,000$, and compare with parts (c) and (f). [4]

(h) Real birthdays are not spread evenly across the year. Investigate, using published birth data or a model of your own, whether unequal frequencies make a shared birthday more or less likely. Alternatively, investigate how large a group must be for *three* people to share a birthday with probability above $\frac{1}{2}$. [2]

**35.** **Three doors.** A prize is behind one of three closed doors, each door equally likely. You choose a door. The host knows where the prize is, and always opens one of the other two doors that does not hold the prize; if both of those doors are empty, because you chose the prize, he chooses between them at random. You may then keep your door or switch to the last closed door.

(a) Find the probability of winning if you always switch. Treat the case where your first choice is correct separately from the case where it is not. [5]

(b) State the probability of winning if you always stay, and say which strategy is better. [1]

(c) Now use Bayes' theorem. You chose door 1 and the host opened door 3. Write down $P(\text{host opens } 3 \mid \text{prize behind } 1)$, $P(\text{host opens } 3 \mid \text{prize behind } 2)$ and $P(\text{host opens } 3 \mid \text{prize behind } 3)$, and hence find $P(\text{prize behind } 2 \mid \text{host opens } 3)$. [5]

(d) A second host knows nothing. He opens one of the two doors you did not choose at random, and it happens to be empty. Repeat the calculation of part (c) for this host, and show that switching now wins with probability $\frac{1}{2}$. [4]

(e) The two hosts performed the same visible action: they opened door 3 and it was empty. Explain what makes the answers differ, and what it is that the first host's action tells you. [3]

(f) The game is now played with $n \geq 3$ doors, one prize and each door equally likely. After your choice, the knowing host opens $n-2$ of the other doors, all of them empty, leaving your door and one other closed; if your door holds the prize, he chooses the other door to leave closed at random. Show that switching wins with probability $\dfrac{n-1}{n}$. [3]

(g) Return to three doors and a knowing host. When you have chosen the prize, this host opens door 3 with probability $q$ and door 2 with probability $1-q$, where $0 \leq q \leq 1$. You chose door 1 and he opened door 3. Show that
$P(\text{prize behind } 2 \mid \text{host opens } 3) = \dfrac{1}{1+q}$. [2]

(h) Deduce that switching is never worse than staying, whatever the value of $q$, and describe what the host's choice of door tells you when $q=0$ and when $q=1$. Suggest a further change to the host's rules that would be worth investigating. [2]

**36. Independent in pairs, but not all together.** Two events $A$ and $B$ are independent when $P(A \cap B) = P(A)P(B)$. Three events $A$, $B$ and $C$ are called *mutually independent* when each of the three pairs is independent and, in addition, $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Two fair coins are tossed. Let $A$ be the event "the first coin shows heads", $B$ the event "the second coin shows heads", and $C$ the event "the two coins show different faces".

(a) Find $P(A)$, $P(B)$, $P(C)$, $P(A \cap B)$, $P(A \cap C)$ and $P(B \cap C)$, and show that each pair of the events $A$, $B$, $C$ is independent. [4]

(b) Find $P(A \cap B \cap C)$, and show that $A$, $B$ and $C$ are not mutually independent. Explain in words why knowing whether $A$ and $B$ happened tells you whether $C$ happened. [3]

(c) Let $A$, $B$ and $C$ be any three events that are independent in pairs, with $P(A \cap B) > 0$. Prove that $P(C \mid A \cap B) = P(C)$ if and only if $P(A \cap B \cap C) = P(A)P(B)P(C)$. [3]

(d) A fair eight-sided die, with faces numbered $1$ to $8$, is rolled once. Let $A = \{1,2,3,4\}$, $B = \{1,2,3,5\}$ and $C = \{1,6,7,8\}$. Show that $P(A \cap B \cap C) = P(A)P(B)P(C)$, but that no two of these events are independent. [3]

(e) Explain what parts (b) and (d) together show about the definition of mutual independence. Mutual independence of four events requires the product rule for every collection of two or more of them. Show that this gives $11$ equations, and find a formula for the number of equations needed for $n$ events. [3]

(f) Now toss $n$ fair coins, and for each pair of coins $i$ and $j$ let $E_{ij}$ be the event that coins $i$

and $j$ show the same face. Investigate which collections of the events $E_{ij}$ are mutually independent, starting with $E_{12}$, $E_{13}$ and $E_{23}$ for three coins, and state a conjecture. [2]