Edexcel International AS Maths: Statistics 1

Your notes

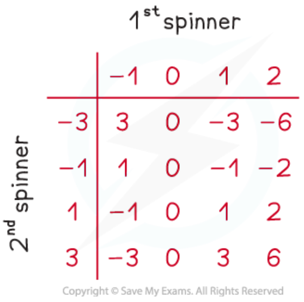
Basic Probability

Contents

Calculating probabilities & events

What language is used in probability?

- The language used in probability can be confusing so here are some definitions of commonly misunderstood terms
    - An **experiment** is a repeatable activity that has a result that can be observed or recorded; it is what is happening in a question
    - An **outcome** is the result of an experiment
    - All possible outcomes can be shown in a **sample space** – this may be a list or a table and is particularly useful when it is difficult to envisage all possible outcomes in your head
- e.g. The sample space below is for two **fair** four-sided spinners whose outcomes are the product of the sides showing when spun.



1st spinner

2nd spinner

	−1	0	1	2
−3	3	0	−3	−6
−1	1	0	−1	−2
1	−1	0	1	2
3	−3	0	3	6

- An **event** is an outcome or a collection of outcomes; it is what we are interested in happening
    - Do note how this could be more than one outcome e.g. For the spinners above, the event "the product is -2" has one outcome but the event "the product is negative" has 6 outcomes
- **Terminology** - be careful with the words **'not'**, **'and'** and **'or'**
    - ***A* and *B*** means both the events *A* and *B* happen at the same time
        - (You will have seen this written as $A \cap B$ at GCSE)
    - ***A* or *B*** means event *A* happens, or event *B* happens, **or both** happen
        - (You will have seen this written as $A \cup B$ at GCSE)
    - **not *A*** means the event *A* does not happen
        - (You will have seen this written as A' at GCSE)
- **Notation** – the way probabilities are written is formal and consistent at A-level

  - $P(A)=0.6$ “the probability of **event** A happening is 0.6”
  - $P(A')=0.4$ “the probability of **event** A **not** happening equals 0.4”
  - (This is sometimes written as $P(\overline{A})$)
- $P(X \le 4)=0.4$ “the probability of X being less than four is 0.4”

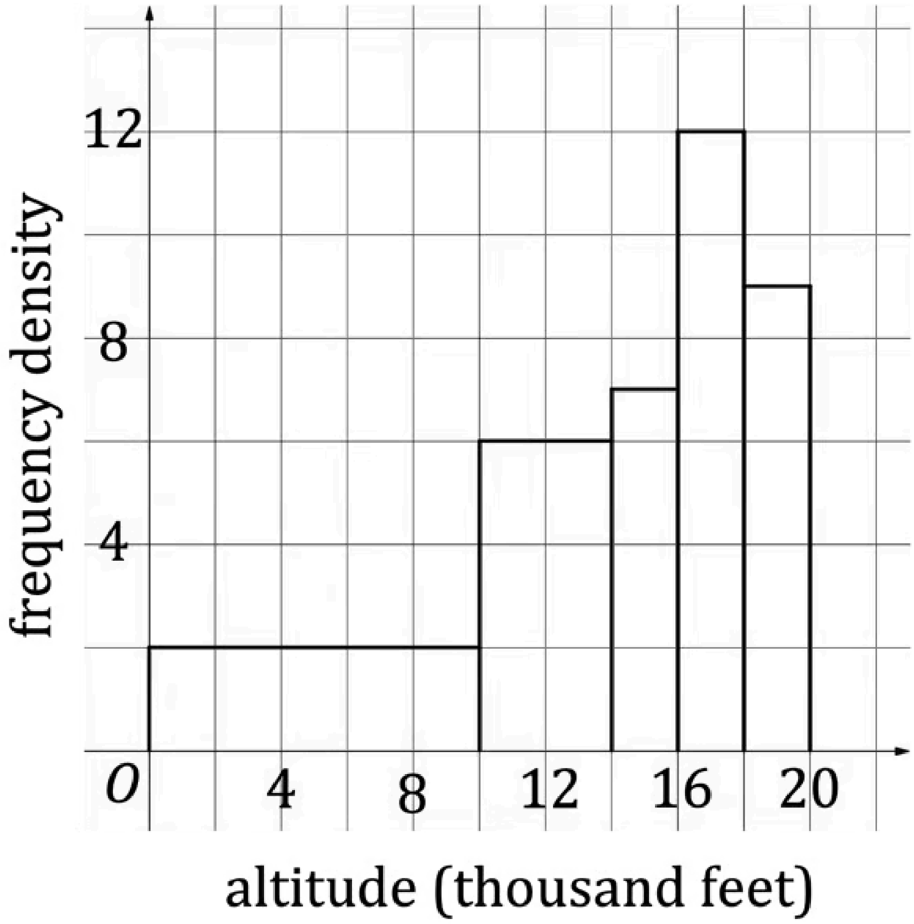
How do I find probabilities?

- The big difference with probability at A level is the language and the notation used
- Recall basic results of probability
  - $\mathrm{P}(\text{" success "}) = \dfrac{\text{number of ways to get " success "}}{\text{total number of outcomes}}$
    - It is important to understand that the above only applies if all outcomes are **equally likely**
  - $\mathrm{P}(\text{not A}) = 1 - \mathrm{P}(\mathrm{A})$
    - The probability of “$not\ A$” is the **complement** of the probability of “*A*”
    - One of the easiest results in probability to understand, one of the hardest results to spot!
- Be aware of whether you are using **theoretical probabilities** or probabilities based on the results of several experiments (**relative frequency**). You may have to compare the two and make a judgement as to whether there is bias in the experiment.
- e.g. The outcomes from rolling a fair dice have theoretical probabilities but the outcomes from a football match would be based on previous results between the two teams
- Ensure you can interpret common ways of displaying data – from frequency tables, histograms, box plots and other ways to illustrate data
  - See Revision Notes
  - 2.1.2 Frequency Tables
  - 2.2.1 Data Presentation
  - 2.2.2 Box Plots & Cumulative Frequency
  - 2.2.3 Histograms
  - Be particularly careful when using **histograms**
    - These use **frequency density**, not frequency
    - Using *parts* of bars may be required due to where class boundaries fall so values will be **estimates** (using the proportion of the bar needed, sometimes called *interpolation*)

Your notes

Worked Example

100 skydivers took part in an all-day charity event, with the altitude of the aeroplane at which they jumped from summarised in the histogram below.



(a) Use the histogram to find the probability that a randomly chosen skydiver jumped from the aeroplane at an altitude

(i) between 14 000 and 16 000 feet,

(ii) between 16 000 and 20 000 feet.

(b) Estimate the probability that a randomly chosen skydiver jumped from the aeroplane at an altitude between 13 000 and 15 000 feet.

Answer:

(Note the use of the word 'event' is used in the context of the question, not for its statistical meaning.)

a) (i) Notice:
- the altitude scale is given in thousands of feet.
- the histogram uses frequency density (fd)

14 − 16, fd = 7 class width (cw) = 16 − 14 = 2

$\therefore f = 7 \times 2 = 14$

$\therefore$ P (altitude is between 14 000 and 16 000 feet) $= \dfrac{14}{100} \left(= \dfrac{7}{50}\right)$

(ii) 16 − 20, Two groups but work out frequencies separately

16 − 18, fd = 12 cw = 18 − 16 = 2

$\therefore f = 12 \times 2 = 24$

18 − 20, fd = 9 cw = 20 − 18 = 2

$\therefore f = 9 \times 2 = 18$

$\therefore$ 16 − 20, f = 24 + 18 = 42

$\therefore$ P (altitude is between 16 000 and 20 000 feet) $= \dfrac{42}{100} \left(= \dfrac{21}{50}\right)$

b) Parts of two groups needed, use interpolation on both.

10 − 14, fd = 6 cw = 14 − 10 = 4

$\therefore f = 6 \times 4 = 24$

For 13 − 14, $f = \dfrac{1}{4} \times 24 = 6$

14 − 16, fd = 7 cw = 16 − 14 = 2 This was also used in part a)

$\therefore f = 7 \times 2 = 14$

For 14 − 15, $f = \dfrac{1}{2} \times 14 = 7$

$\therefore$ 13 − 15, f = 6 + 7 = 13 Remember this is an estimate

$\therefore$ P (altitude is between 13 000 and 15 000 feet) $\approx \dfrac{13}{100}$

Examiner Tips and Tricks

- Most probability questions are in context so can be long and wordy; go back and re-read the question, several times, whenever you need to
- Try to get immersed in the context of the question to help understand a problem

Independent & mutually exclusive events

What are independent events?

- **Independent events** do not affect each other

- For two independent events, the probability of one event happening is unaffected by the outcome of the other event
- e.g. The events “rolling a 6 on a dice” and “flipping heads on a coin” are independent - the outcome “rolling a 6” does not affect the probability of the outcome “heads” (and vice versa)
- For two **independent** events, *A* and *B*

$$\mathrm{P}(A\ \textbf{AND}\ B) = \mathrm{P}(A) \times \mathrm{P}(B)$$

- e.g. $\mathrm{P}(\text{"6 on a dice" } \textbf{AND} \text{ "heads on a coin"}) = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$
- Independent events could refer to events from different experiments

What are mutually exclusive events?

- **Mutually exclusive events** cannot occur simultaneously
  - $\mathrm{P}(A \text{ AND } B) = 0$
- For two mutually exclusive events, the outcome of one event means the other event cannot occur e.g. The events “rolling a 5 on a die” and “rolling a 6 on a die” are mutually exclusive
- For two **mutually exclusive** events, *A* and *B*

$$\mathrm{P}(A\ \textbf{OR}\ B) = \mathrm{P}(A) + \mathrm{P}(B)$$

- e.g. $\mathrm{P}(\text{"6 on a dice" } \textbf{OR} \text{ "5 on a dice"}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \quad \left(\frac{1}{3}\right)$
- **Mutually exclusive** events generally refer to events from the same (single trial of an) experiment
- Mutually exclusive events cannot be independent; the outcome of one event means the probability of the other event is zero

How do I solve problems involving independent and mutually exclusive events?

- Make sure you know the statistical terms – **independent** and **mutually exclusive**
- Remember
  - **independence** is **AND** and is $\times$
  - **mutual exclusivity** is **OR** and is $+$
- Solving problems will require interpreting the information given and the application of the appropriate formula
  - Information may be explained in words or by diagram(s)
  - (including Venn diagrams – see Revision Note 3.1.2 Venn Diagrams)

- **Showing** or **determining** whether two events are independent or mutually exclusive are also common
  - To do this you would show the relevant formula is true
- **Just for fun ...**
- A well-known sports TV broadcaster used to advertise their football matches as either "Live and exclusive" or "Exclusively live" – can you tell the difference?
- "Live and exclusive" meant that the broadcaster was airing the football match live and was the only broadcaster allowed to air any of the match at any time. "Exclusively live" meant that the broadcaster was the only one airing the football match live, but other broadcasters would be able to air any of the match afterwards.

Worked Example

(a) Two events, Q and R are such that $\mathrm{P}(Q) = 0.8$ and $\mathrm{P}(Q \text{ and } R) = 0.1$. Given that Q and R are independent, find $\mathrm{P}(R)$

(b) Two events, S and T are such that $\mathrm{P}(S) = 2\mathrm{P}(T)$. Given that S and T are mutually exclusive and that $\mathrm{P}(S \text{ or } T) = 0.6$ find $\mathrm{P}(S)$ and $\mathrm{P}(T)$.

(c) A fair five-sided spinner has sides labelled 2, 3, 5, 7, 11. Find the probability that the spinner lands on a number greater than 5.

Answer:

Your notes

a) For independent events. "P (A AND B) = P (A) × P (B)"

∴ P (Q AND R) = P (Q) × P (R)

$0.1 = 0.8 \times P(R)$

$P(R) = \frac{0.1}{0.8} = 0.125$

$P(R) = 0.125$

b) For mutually exclusive events. "P (A OR B) = P (A) + P (B)"

∴ P (S OR T) = P (S) + P (T)

$0.6 = 2P(T) + P(T)$ Since P (S) = 2P (T)

∴ $3P(T) = 0.6$

$P(T) = 0.2$

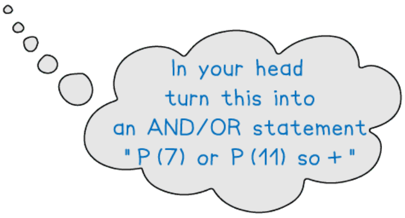
∴ $P(S) = 2 \times 0.2 = 0.4$

$P(S) = 0.4$

$P(T) = 0.2$

c) You should be able to do this question using common sense/logic, but you should now be aware P (7) and P (11) are mutually exclusive

P (greater than 5) = P (7) + P (11) = 0.2 + 0.2 = 0.4



P (greater than 5) = 0.4

Examiner Tips and Tricks

- Try to rephrase questions in your head in terms of **AND** and/or **OR** ! e.g. A fair six-sided die is rolled and a fair coin is flipped. "Find the probability of obtaining a prime number with heads."

  would be

  "Find the probability of rolling a 2 **OR** a 3 **OR** a 5 **AND** heads."

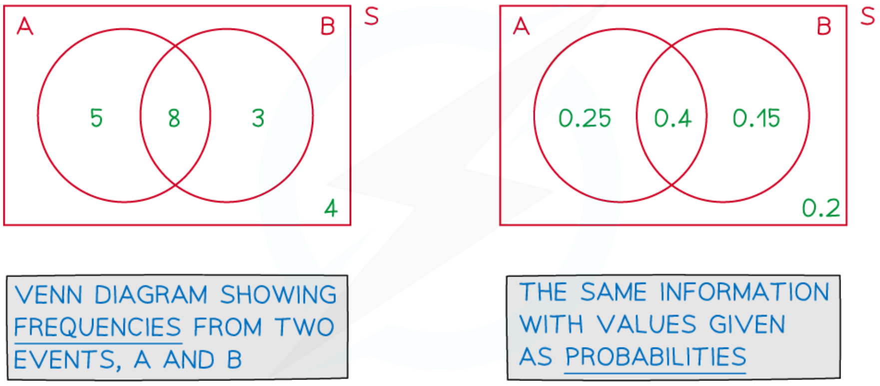
Venn diagrams

What is a Venn diagram?

- A Venn diagram is a way to illustrate **events** from an **experiment** and are particularly useful when there is an overlap (or lack of) between possible **outcomes**
- A Venn diagram consists of
  - a **rectangle** representing the **sample space**
  - a **bubble** (usually drawn as a circle/ellipse/oval) for each **event**
    - Bubbles may or may not overlap depending on which **outcomes** are shared between **events**
    - Bubble(s) is not a technical term but we like it!

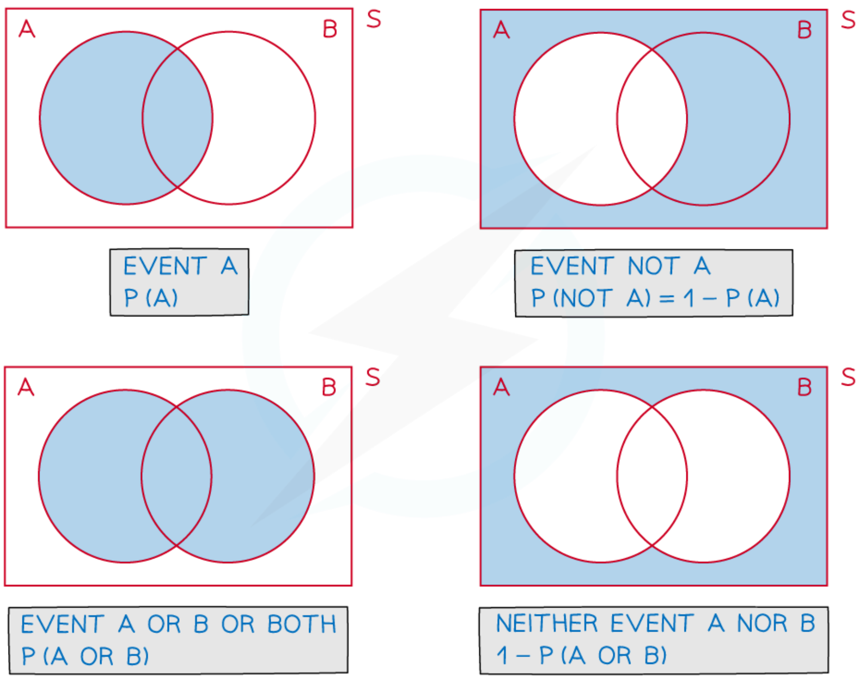
How do I interpret a Venn diagram?

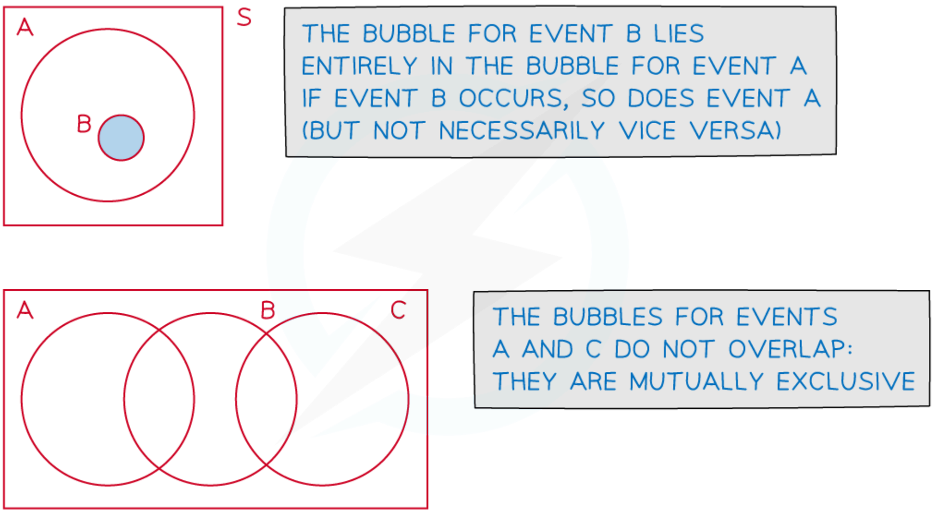
- The rectangle is usually labelled with an S – as it represents the **sample space** (all possible **outcomes** from the **experiment)**
  - It is often referred to as the Universal Set and is commonly labelled with $S,\ U,\ \xi$ (the Greek lower case letter Xi) or $\mathscr{E}$ (Kunstler script font) There is no standardised symbol used for this purpose
- Bubbles are labelled with their event name (*A*, *B*, *etc)*
- The numbers inside a Venn diagram (there should be one in each region) will represent either a **frequency** or a **probability**
  - In the case of probabilities being shown, all values should total 1



What do the different regions mean on a Venn diagram?

- This will depend on how many events there are and how the outcomes overlap
- Venn diagrams show '**AND**' and '**OR**' statements easily
- Venn diagrams also instantly show mutually exclusive events
    - **Independence** can be deduced from the probabilities involved





THE BUBBLE FOR EVENT B LIES ENTIRELY IN THE BUBBLE FOR EVENT A IF EVENT B OCCURS, SO DOES EVENT A (BUT NOT NECESSARILY VICE VERSA)

THE BUBBLES FOR EVENTS A AND C DO NOT OVERLAP: THEY ARE MUTUALLY EXCLUSIVE

How do I find probabilities from Venn diagrams?

- Draw, or add to a given Venn diagram, filling in as many values as possible from the information provided in the question
- It is usually helpful to work from the centre outwards
  - i.e. fill in **intersections** (overlaps) first
  - This is particularly crucial with Venn diagrams with three events
    - the **intersection** of events A and B will include the intersection of events *A*, *B* and C
    - a question would make it clear if a given frequency or probability is only for events *A* and *B* , and ***not*** *C*
- Any **frequencies** or **probabilities not given** may be able to be **calculated** from those that are
  - Use the results from Basic Probability to deduce missing frequencies or probabilities and answer questions
    - $\mathrm{P}(\text{not } A) = 1 - P(A)$
      - For **independent** events, $\mathrm{P}(A \text{ and } B) = \mathrm{P}(A) \times \mathrm{P}(B)$
      - For **mutually exclusive** events, $\mathrm{P}(A \text{ or } B) = \mathrm{P}(A) + \mathrm{P}(B)$
- Check a completed Venn diagram that the **frequencies sum** to the **total** involved or that **probabilities sum** to **1**

Worked Example

40 people were surveyed regarding which games consoles they owned.

8 people said they owned a Playstation 5 (P) and an Xbox Series X (X); 11 people said they owned a Playstation 5 (P) and a Nintendo Switch (N); 7 people said they owned an Xbox Series X (X) and a Nintendo Switch (N).

4 people said they owned none of these consoles whilst 2 people said they owned all 3.

Of those people that owned only one games console, twice as many owned a Nintendo Switch as a Playstation 5 and half as many owned an Xbox Series X as a Playstation 5.

(a) Draw a complete Venn diagram to illustrate the information given above.

(b) One of the 40 people is chosen at random. Find the probability that this person

(i) owns all three consoles,

(ii) owns exactly two consoles,

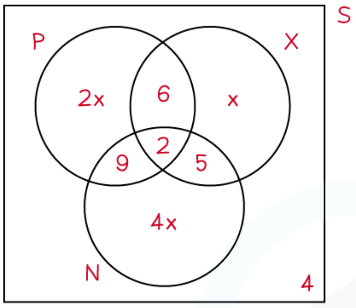
(iii) doesn't own a Playstation 5.

(c) Determine if the events X and N are independent.

Answer:

(a) Let the number of people who own an X Box Series X only be x.
Then Playstation 5 only would be 2x. Nintendo Switch 4x

Draw an initial Venn diagram



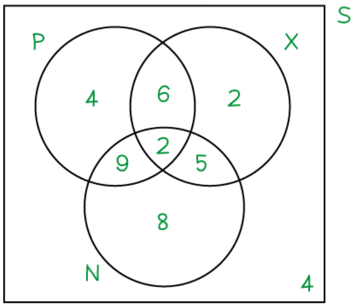
Set up an equation to find x

$2x + 6 + x + 9 + 2 + 5 + 4x + 4 = 40$

$7x + 26 = 40$

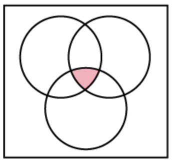
$x = 2$

Now draw the complete Venn diagram



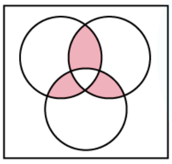
(b) Use "mini–Venn" diagrams to shade the parts required

(i)



$$P\,(P \text{ and } X \text{ and } N) = \frac{2}{40} \left(= \frac{1}{20}\right)$$

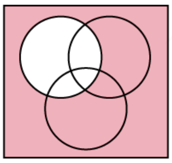
(ii)



P (P and X or P and N or X and N)

$$= \frac{6}{40} + \frac{9}{40} + \frac{5}{40} = \frac{20}{40} \left(= \frac{1}{2}\right)$$

(iii)



$$P\,(\text{not } P) = 1 - P\,(P) = 1 - \frac{4+6+9+2}{40} = \frac{19}{40}$$

(c) For independence, "P (A and B) = P (A) × P (B)"

From Venn diagram $P\,(X) = \frac{6+2+2+5}{40} = \frac{15}{40}$

$$P\,(N) = \frac{9+2+5+8}{40} = \frac{24}{40}$$

$$P\,(X \text{ and } N) = \frac{2+5}{40} = \frac{7}{40}$$

$$P\,(X) \times P\,(N) = \frac{15}{40} \times \frac{24}{40} = \frac{9}{40}$$

P (X and N) ≠ P (X) × P (N). Therefore the events X and N are not independent

Examiner Tips and Tricks

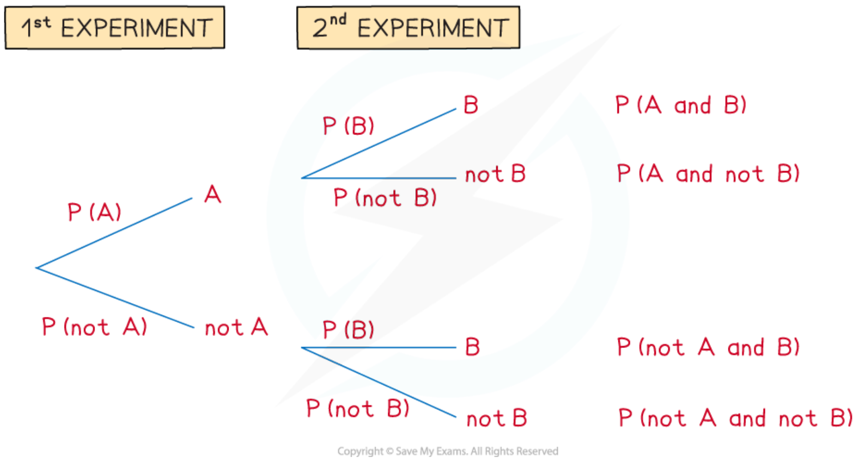
- Always draw the box in a Venn diagram; it represents all possible outcomes of the experiment so is a crucial part of the diagram, the bubbles merely represent the events we are particularly interested in
- In complicated problems it can be helpful to draw a “mini-Venn” diagram for part of a question and shade/label the regions of the diagram that are relevant to that part

Tree diagrams

What is a tree diagram?

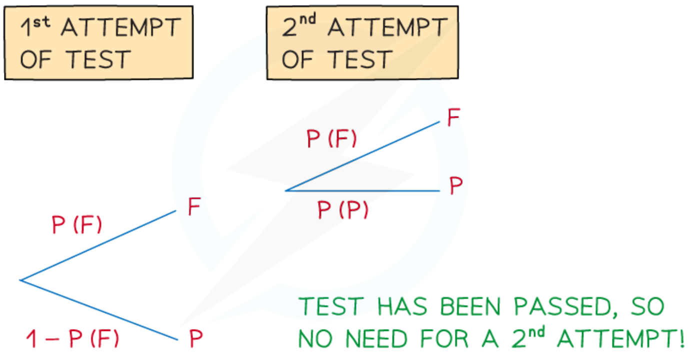
- A **tree diagram** is used to
  - show the (combined) **outcomes** of more than one **event** that happen one after the other
  - help calculate probabilities when **AND** and/or **OR**'s are involved
- **Tree diagrams** are mostly used when there are only two **mutually exclusive outcomes** of interest
  - e.g. "*Rolling a 6 on a die*" and "*Not rolling a 6 on a die*"
- More than three outcomes per event can be shown on a tree diagram but they soon become difficult to draw and so lose their effectiveness
- **Tree diagrams** are very helpful when **probabilities** for a second event change **depending** on the first event

How do I draw and label a tree diagram?



- In the second experiment, P(B) may be different on the top set of branches than the bottom set
  - this is because the top set of branches follow on from **event A** but the bottom set of branches follow on from **event** "$\textbf{\textit{not A}}$"
  - e.g. This is most commonly seen in drawing one item at random, not replacing it, then drawing another
- Sometimes a second branch may not be needed following a first event

- e.g. In aiming to pass a test (**experiment**) the event **fail** on the first attempt would require a second attempt but the event **pass** on the first attempt would not



How do I find probabilities from tree diagrams?

- Interpret questions in terms of **AND** and/or **OR** (See 1 Basic Probability)
- Draw, or complete a given, tree diagram Determine any **missing** probabilities; often using $1-\mathrm{P}(A)$ and considering if probabilities change depending on the outcome from the 1st experiment
- Write down the (final) outcome of the combined events and work out their probabilities – these are **AND** statements $\mathrm{P}(A \text{ AND } B)=\mathrm{P}(A)\times\mathrm{P}(B)$ (*"Multiply along branches"*)
  - Do not simplify fractions yet – it'll be easier to calculate with them later
    - you can of course use your calculator
- If more than one (final) outcome is required to answer a question then add their probabilities – these are **OR** statements $\mathrm{P}(AB \ OR \ \text{"}not\ A\text{"}\ \text{"}not\ B\text{"})=\mathrm{P}(AB)+\mathrm{P}(\text{"}not\ A\text{"}\ \text{"}not\ B\text{"})$ (*"Add outcomes"*)
  - This applies since all the (final) outcomes are **mutually exclusive**
  - Note that $AB,\ \text{"}not\ A\text{"}\ \text{"}not\ B\text{"}$ are implied **AND** statements (for example *AB* means *A* and *B*)
- When you are confident with tree diagrams you can just pull out the (final) outcome(s) you need to answer a question rather than routinely list all of them

Worked Example

A contestant on a game show has three attempts to hit a target in a shooting game. They have a maximum of three attempts to hit the target in order to win the star prize – a speedboat. If they do not hit the target within three attempts, they do not win anything.

The probability of them hitting the target first time is 0.2. With each successive attempt the probability of them failing to hit the target is halved.

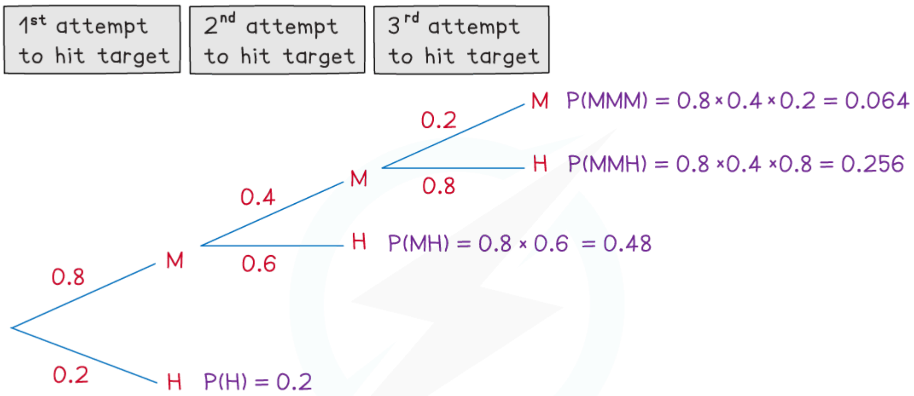
Find the probability that a contestant wins the star prize of a speedboat.

Answer:

Draw a tree diagram and determine missing probabilities

H is the event "hits target"

M is the event "misses target"



P(MMM) = 0.8 × 0.4 × 0.2 = 0.064

P(MMH) = 0.8 × 0.4 × 0.8 = 0.256

P(MH) = 0.8 × 0.6 = 0.48

P(H) = 0.2

Read along branches write down the outcome and work out its probability

Answer the question

P (wins speedboat) = 0.2 + 0.48 + 0.256 = 0.936

Alrenatively, in STEP 2, only calculate P (MMM), then,

P (wins speedboat) = 1 − 0.064 = 0.936

Examiner Tips and Tricks

- It can be tricky to get a tree diagram looking neat and clear first attempt – it can be worth drawing a rough one first, especially if there are more than two outcomes or more than two events; do keep an eye on the exam clock though!
- Always worth another mention – tree diagrams make particularly frequent use of the result $\mathrm{P}(\text{not } A) = 1 - P(A)$
- Tree diagrams have built-in checks
    - the probabilities for each pair of branches should add up to 1
    - the probabilities for each outcome of combined events should add up to 1