Determining the Probability of a Good Day for Sowing Cotton

Task 3 – Calculating probability and conditional probability

Calculating Probabilities using a Venn Diagram and a Two-way Table

Probabilities are proportions (i.e. fractions). They indicate what proportion of times a particular outcome is expected to occur.

Venn diagrams and two-way tables can be used to calculate a probability and/or a conditional probability of something occurring.

The thing that occurs is called the **outcome**.

The number of times that an outcome occurs is its **frequency.**

The **relative frequency** of an outcome is its frequency expressed as a fraction of the largest possible number of times it could have occurred.

The **probability** of an outcome occurring in the future can be considered to be its relative frequency (the proportion of times it occurred in the past), assuming that nothing changes.

The relative frequency of an outcome is also called its **experimental probability**.

It is calculated as:

$$Experimental\ Probability\ of\ an\ outcome = \frac{The\ frequency\ of the\ outcome}{The\ sample\ space}$$

The **sample space** is the number of times the "experiment" was done (i.e. the greatest number of times that the outcome *could* have occurred).

A farmer at Trangie wants to know the relative frequency (probability) of days in a planting season being good for sowing cotton.

There are two conditions which, if satisfied, result in days with a favourable outcome for sowing.

Condition A: Soil temperature at 10 cm depth above 14°C at 9am (AEST)

Condition B: Forecast average temperatures for the week following planting on a rising plane.

A day in the cotton planting season may or may not meet these conditions.

In the language of probability, each of these conditions is an **event**.

Each event has 2 possible **outcomes** so altogether, the 2 events result in 4 possible outcomes:

- If conditions A and B are *both* satisfied, it is a **green light** day, meaning that the farmer should go ahead and sow cotton that day.
- If *neither* condition is satisfied, it is a **red light** day, meaning that the farmer should stop and not sow cotton that day.
- If Condition A is satisfied and Condition B is *not* satisfied it is a **brown light** day, meaning that the farmer should be cautious about sowing cotton that day.

- If Condition B is satisfied and Condition A is *not* satisfied it is an **orange light** day, meaning that the farmer should be cautious about sowing cotton that day.

Probabilities

Probabilities

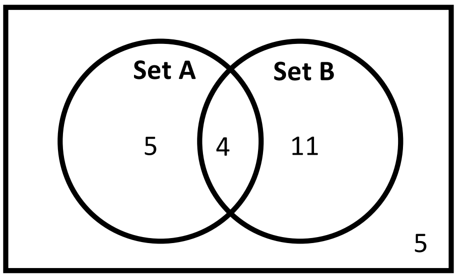
Events are **mutually exclusive** if when one event occurs, it is impossible for another event to also occur. They cannot occur together (mutually).

- Are the two events (a day when soil temperature is greater than 14°C and a day when average temperatures for the following week are on a rising plane) mutually exclusive? Why or why not?

Outcomes are mutually exclusive if when one outcome of an event occurs, it is impossible for another outcome of that event to also occur.

- For Event A (whether soil temperature is greater than 14°C) what are the two possible outcomes? Are these outcomes, mutually exclusive? Why or why not?

Below is the Venn diagram from Task 2 showing the numbers of the different types of days in the 1992 season (the coldest planting season of the climatic data record).
Set A is the group of days when Condition A was satisfied and Set B is the group of days when Condition B was satisfied.



For each day (a day being an element in the universal set), one of 4 possible **outcomes** occurs.
The number written in each region of the Venn diagram is the frequency of each outcome.
The number of elements in the universal set (the total of the frequencies) is the sample space.

- How many days are in the sample space?

The notation for the frequency (i.e. the number) of elements in Set A is $\mathrm{n}(\mathrm{A})$.
The notation for the probability of an element being in Set A is $\mathrm{P}(\mathrm{A})$.

$$\mathrm{P}(\mathrm{A}) = \frac{\mathrm{n}(\mathrm{A})}{\textit{The sample space}} = \frac{9}{25}$$

The probability of a **green light** day is $\mathrm{P}(\mathrm{A} \cap \mathrm{B})$.

$$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \frac{\mathrm{n}(\mathrm{A} \cap \mathrm{B})}{\textit{The sample space}} = \frac{4}{25}$$

Determining the Probability of a Good Day for Sowing Cotton

- Shade the Venn diagram above to help you find out which colour of day is represented by each of the following probabilities. Then calculate the probabilities.
  - $P(A' \cap B')$
  - $P(A \cap B')$
  - $P(A' \cap B)$

Conditional Probabilities

The same data that is in the Venn diagram, is in the two-way table below.

1992	**Set A**	**Set A'**	**Marginal Totals**
Set B	4	11	**15**
Set B'	5	5	**10**
Marginal Totals	**9**	**16**	**25**

The shaded cell in the table gives the number of elements in the sample space (the universal set).

Look at the column of the table headed Set A.
The marginal total for Set A tells you that there were 9 days for which Condition A was satisfied.
Of these 9 days, 4 also satisfied Condition B. This is $\frac{4}{9}$ of the days in Set A.
The denominator of the fraction (the sample space) is now *restricted* to the **marginal total**, 9.

Look at the column of the table for Set A'.
The marginal total for Set A' tells you that there were 16 days for which Condition A was *not* satisfied. Of these 16 days, 11 did *not* satisfy Condition B. This is $\frac{11}{16}$ of the days in Set A'.

Event A and Event B are **dependent events** because the outcome of Event A (i.e. whether the day is in Set A or Set A') affects the probabilities of the outcomes of Event B (whether the temperatures for the following week are on a rising plane).
If the day is in Set A, $P(B) = \frac{4}{9}$, whereas if the day is in Set A', $P(B) = \frac{11}{16}$.

Determining the Probability of a Good Day for Sowing Cotton

Set A Set B

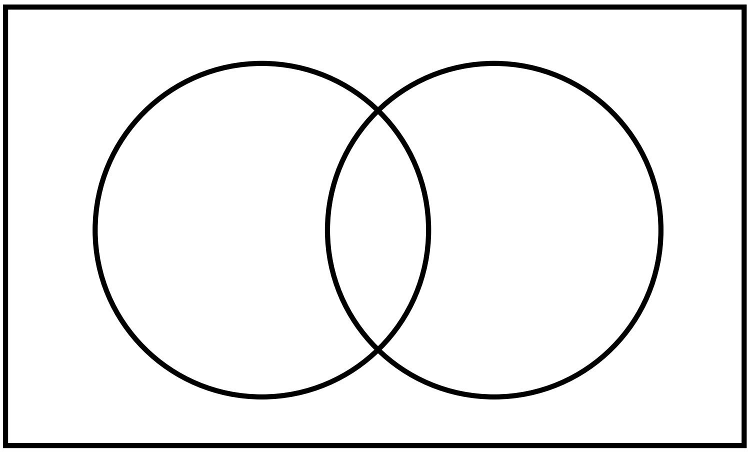
5 4 11

The probability that a day is in Set B, given that day is in Set A is called a **conditional probability**, where the day being in Set A is the **initial condition**.

5

The notation used for P(B given A) is $\mathbf{P(B|A)}$.

Looking at Venn diagram below, $\mathrm{P(B|A)} = \frac{\mathrm{P(A \cap B)}}{\mathrm{P(B)}} = \frac{4}{4+5} = \frac{4}{9}$



The probability that a day is in Set B, given that day is in Set A, is $\frac{4}{9}$.
This can be expressed in different ways.

- With a partner, decide whether the following sentences are true or false, i.e. whether they mean the same as *The probability that a day is in Set B, given that day is in Set A, is* $\frac{4}{9}$.

  Then share your answers with another pair of people. Convince them that your answers are correct, or be convinced by them that you are incorrect and change your answers.

  1) If a day is in Set B, then the probability that it is in Set A is $\frac{4}{9}$

  2) If a day is in Set A, then the probability that it is in Set B is $\frac{4}{9}$

  3) Given that a day is in Set B, the probability that it is in Set A is $\frac{4}{9}$

  4) Knowing that a day is in Set B, the probability that it is also in Set A is $\frac{4}{9}$

  5) If you already know that a day is in Set A, the probability that it is in Set B is $\frac{4}{9}$

- What is meant by $\mathrm{P(B|A')} = \frac{11}{16}$?
- What is meant by $\mathrm{P(A|B)}$? What is the value of $\mathrm{P(A|B)}$ in the 1992 season?
- What is meant by $\mathrm{P(A|B')}$? What is the value of $\mathrm{P(A|B')}$ in the 1992 season?

If events A and B are **independent**, whether an element is in Set A or not in Set A, has no effect on the probability of it being in Set B or *not* in Set B.

Determining the Probability of a Good Day for Sowing Cotton

In this case, $\mathrm{P(B|A) = P(B|A')}$
so $\mathrm{P(B|A)}$ is simply the same as $\mathrm{P(B)}$. i.e. $\mathrm{P(B|A) = P(B)}$
Also, $\mathrm{P(A|B) = P(A|B')}$ so $\mathrm{P(A|B) = P(A)}$

Independent events are *not* the same as mutually exclusive events.
Two events are mutually exclusive if the occurrence of one event *excludes* the occurrence of the other. Mutually exclusive events *cannot* happen at the same time. In other words an element cannot be in both sets, so for mutually exclusive events $\mathrm{P(A \cap B) = 0}$.

When events are independent, the conditional probability of Event B occurring given that Event A has occurred, is simply the probability of Event B occurring i.e. P(B|A) = P(B).

- With a partner, decide whether the pairs of events below are dependent or independent.
  Read information at http://cottonaustralia.com.au/australian-cotton/basics/cotton-facts.
  Then share your answers with others so you can convince them or be convinced by them.

  (i) Whether a cotton crop is genetically modified, and the quantity of insecticide used on it.

  (ii) Whether an area is sown to cotton or left fallow, and the emission of greenhouse gases.

  (iii) The weight of a bale of cotton, and what that bale of cotton is used to make.

  (iv) Rain occurring on a day at Trangie, and rain occurring on the same day at Narromine.

Probability Trees
Probabilities can be calculated from three different representations:
- Venn diagrams
- Two-way tables
- Probability trees

Probability trees show the possible outcomes by creating a **branch** for each one.
For planting cotton, there are 2 events: whether Condition A is satisfied (Event A) and whether Condition B is satisfied (Event B).

To begin with, the tree diagram for the 1992 season has 2 branches representing the 2 possible outcomes (Yes and No) for Event A.
$\mathrm{P(A)} = \frac{9}{25}$ $\quad \mathrm{P(A')} = \frac{16}{25}$
These probabilities are written on the initial two branches of the probability tree.

Each of these branches then splits into 2, representing the 2 possible outcomes (Yes and No) for Event B, given the outcome for Event A.
$\mathrm{P(B|A)} = \frac{4}{9}$ $\quad \mathrm{P(B'|A)} = \frac{5}{9}$

Determining the Probability of a Good Day for Sowing Cotton

Event A **Event B** **Outcome**

$\frac{9}{25}$ Yes $\frac{4}{9}$ Yes **Green light** day

$\frac{5}{9}$ No **Brown light** day

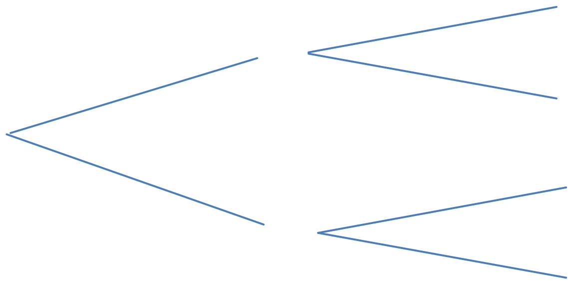
$\frac{16}{25}$ No Yes **Orange light** day

No **Red light** day

These probabilities are written on the two branches that split from the initial top branch.

- What probabilities should be written on the branches that split from the initial bottom branch? Calculate their values. Then write them on these two branches.

  The probabilities on the bottom two branches are *not* the same as the probabilities on the top two branches because Event A and Event B are *not* independent events.



The probability of the final outcome (the colour of the day) depends on all the probabilities on the branches that lead to that outcome.

For independent events, $\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B})$

For dependent events, $\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B}|\mathrm{A})$ so $\mathrm{P}(\mathrm{B}|\mathrm{A}) = \frac{\mathrm{P}(\mathrm{A} \cap \mathrm{B})}{\mathrm{P}(\mathrm{A})}$ (as seen on page 3)

- The branches of the **green light** day outcome are $\mathrm{P}(\mathrm{A}) = \frac{9}{25}$ and $\mathrm{P}(\mathrm{B}|\mathrm{A}) = \frac{4}{9}$

  Calculate the probability of a **green light** day using these two values.

- Using the tree diagram, write the calculations needed to find:
  - the probability of a **brown light** day in the 1992 season
  - the probability of an **orange light** day in the 1992 season
  - the probability of a **red light** day in the 1992 season.

Event A **Event B** **Event C**

Yes Yes Yes No

No Yes No

Yes Yes No

No Yes No

The two-way table of frequencies in 2000 (a fairly typical planting season) is shown below.

- Draw a probability tree and calculate the probability of each of the four outcomes.
- For this season, could you have drawn a probability tree with fewer branches?

2000	Set A	Set A'	Marginal Totals
Set B	13	1	**14**
Set B'	10	0	**10**

Marginal Totals	**23**	**1**	**24**

The two-way table of frequencies in 2006 (the warmest of planting season) is shown below.

- Draw a probability tree and calculate the probability of each of the four outcomes.
- For this season, could you have drawn a probability tree with fewer branches?

2006	**Set A**	**Set A'**	**Marginal Totals**
Set B	12	0	**12**
Set B'	11	0	**11**
Marginal Totals	**23**	**0**	**23**

- A farmer with a small property is interested in the probability of a **green light** day occurring.
  Does this probability vary much between seasons?

- A farmer with a large property is interested in the probability of a **green light** *or* a **brown light** *or* an **orange light** day occurring as he will sow on any of these days.
  Calculate the probability of this occurring on a day in the 1992 season, the 2000 season and the 2006 season. Does this probability vary much between seasons?

- If the average soil temperature and the average air temperature over a planting season at Trangie increase in the future,
  - would you expect the probability of **green light** days to increase or decrease? Why?
  - would you expect the probability of **red light** days to increase or decrease? Why?