Lesson - Conditional probability

Year: Year 11

This is a Higher tier topic and it often appears near the end of a paper. The word "given" (or phrases like "knowing that" or "of those who") tells you the sample space has become smaller: you only look at the part of the diagram where the condition is true.

Starter activity

A bag contains 3 red and 2 blue sweets. You take one without looking and eat it; it is red. What is the probability that the next one is red as well? How does this differ from the probability you would have given before the first sweet was eaten? Write down both answers and explain why the second probability depends on what happened first.

Conditional probability and notation

Definition (Higher): conditional probability

A **conditional probability** is the probability of an event happening *given that* another event has already happened. $P(A \mid B)$ means "the probability of A, given that B has happened". The vertical line is read as "given".

Rule (Higher): the conditional probability formula

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

You do not need to memorise this for most GCSE questions, because a table, tree or Venn diagram lets you work it out by counting. The denominator is always the total for the condition.

Using two-way tables

Worked example 1 (Higher): a two-way table

100 students: 45 are boys and 55 are girls. 20 boys and 15 girls walk to school. Find $P(\text{walks} \mid \text{boy})$ and $P(\text{girl} \mid \text{walks})$.

$P(\text{walks} \mid \text{boy}) = \frac{20}{45} = \frac{4}{9}$ (the condition "boy" limits us to the 45 boys).

$P(\text{girl} \mid \text{walks}) = \frac{15}{35} = \frac{3}{7}$ (the condition "walks" limits us to the 35 who walk).

$P(\text{boy} \mid \text{walks}) = \frac{20}{35}$ is different from $P(\text{walks} \mid \text{boy}) = \frac{20}{45}$: the order matters.

Using Venn diagrams

Worked example 2 (Higher): a Venn diagram

50 people: 30 like tea, 25 like coffee, 10 like both. Find $P(\text{tea} \mid \text{coffee})$, $P(\text{coffee} \mid \text{tea})$ and $P(\text{not coffee} \mid \text{tea})$.

Tea only $= 20$, coffee only $= 15$, both $= 10$, neither $= 50 - 45 = 5$. Then $P(\text{tea} \mid \text{coffee}) = \frac{10}{25} = \frac{2}{5}$, $P(\text{coffee} \mid \text{tea}) = \frac{10}{30} = \frac{1}{3}$ and $P(\text{not coffee} \mid \text{tea}) = \frac{20}{30} = \frac{2}{3}$.

Using tree diagrams

Key fact (Higher): trees without replacement

Tree diagrams without replacement are naturally conditional: the second set of branches shows the probability given what happened first. A bag has 4 red and 6 blue counters and two are taken without replacement: $P(\text{second red} \mid \text{first red}) = \frac{3}{9} = \frac{1}{3}$ and $P(\text{second blue} \mid \text{first blue}) = \frac{5}{9}$.

Method (Higher): conditional probability from a tree

1. Find the probability of each route to the condition.
2. Add them to get the denominator (write it first, so you identify the allowed routes).
3. Divide the route (or routes) you want by that total.

Worked example 3 (Higher): at least one red

Find $P(\text{both red} \mid \text{at least one red})$ for the bag above.

$P(RR) = \frac{4}{10} \times \frac{3}{9} = \frac{2}{15}$. $P(\text{at least one red}) = 1 - \frac{6}{10} \times \frac{5}{9} = 1 - \frac{1}{3} = \frac{2}{3}$. So the answer is $\frac{2/15}{2/3} = \frac{2}{15} \times \frac{3}{2} = \frac{1}{5}$.

Worked example 4 (Higher): working backwards

The probability of rain is 0.3. If it rains, the probability Ali is late is 0.5; if not, 0.1. Find $P(\text{rain} \mid \text{late})$.

$P(\text{rain and late}) = 0.3 \times 0.5 = 0.15$. $P(\text{no rain and late}) = 0.7 \times 0.1 = 0.07$.

$P(\text{late}) = 0.15 + 0.07 = 0.22$. $P(\text{rain} \mid \text{late}) = \dfrac{0.15}{0.22} = \dfrac{15}{22}$.

Worked example 5 (Higher): a medical test

2% of people have a disease. The test is positive for 90% of people with the disease and for 5% of people without it. Find $P(\text{disease} \mid \text{positive})$.

$P(\text{positive}) = 0.02 \times 0.9 + 0.98 \times 0.05 = 0.018 + 0.049 = 0.067$.

$P(\text{disease} \mid \text{positive}) = \dfrac{0.018}{0.067} = \dfrac{18}{67} \approx 0.27$. Even after a positive test the chance of having the disease is only about 27%, because far more healthy people are tested than ill people. A good answer to a "comment on this result" question refers to the numbers in the calculation.

Independence

Definition (Higher): independent events

Events A and B are **independent** if $P(A \mid B) = P(A)$, which is the same as $P(A \cap B) = P(A) \times P(B)$.

Worked example 6 (Higher): testing independence

Are A and B independent if (a) $P(A) = 0.4$, $P(B) = 0.5$, $P(A \cap B) = 0.2$; (b) $P(A) = 0.5$, $P(B) = 0.4$, $P(A \cap B) = 0.3$?

(a) $0.4 \times 0.5 = 0.2$, so they are independent. (b) $P(A \mid B) = 0.75 \neq 0.5$, so they are not independent.

Worked example 7 (Higher): a sweet problem

A bag has 7 mint and 5 toffee sweets. Two are taken without replacement. Find the probability that the second is toffee.

Mint then toffee: $\dfrac{7}{12} \times \dfrac{5}{11} = \dfrac{35}{132}$. Toffee then toffee: $\dfrac{5}{12} \times \dfrac{4}{11} = \dfrac{20}{132}$. Add: $\dfrac{55}{132} = \dfrac{5}{12}$, the same as the probability that the first sweet is toffee. By symmetry, when nothing is known about the first draw, the second has the same chance as the first.

Choosing a method

Note: choosing a method

Information given	Best method

Counts of people in categories	Two-way table or Venn diagram: divide by the condition total
Stages with probabilities on each stage	Tree diagram: multiply along, add routes, then divide
Probability of "A and B" and of B	Formula: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

With a tree, the final answer is a fraction whose numerator is one route (or a sum of routes) and whose denominator is the sum of all routes leading to the condition. The routes in the top must be included in those in the bottom, so if your answer is greater than 1 something is wrong.

Watch out: common mistakes

- Using the wrong total as the denominator: it must be the total for the condition.
- Mixing up $P(A \mid B)$ with $P(B \mid A)$.
- Forgetting to adjust the probabilities in a tree without replacement.
- Not adding all the routes that lead to the condition.

Worked example 8: Conditional probability from a two-way table

In a class of 30 pupils, 18 study French and 12 study German; 5 study both. A pupil who studies French is chosen. What is the probability that they also study German?

Only the French students matter now, so the denominator is 18, not 30. Of these, 5 study both languages. So $P(\text{German} \mid \text{French}) = \dfrac{5}{18}$. Notice that $P(\text{German}) = \dfrac{12}{30} = 0.4$ is different, so the two events are not independent.

Worked example 9: A tree diagram with a conditional step

The probability that a student revises is 0.7. If they revise, the probability of passing is 0.9. If not, the probability is 0.4. Find the probability that a student who passed had revised.

$P(\text{revise and pass}) = 0.7 \times 0.9 = 0.63$. $P(\text{not revise and pass}) = 0.3 \times 0.4 = 0.12$. $P(\text{pass}) = 0.63 + 0.12 = 0.75$. So $P(\text{revised} \mid \text{passed}) = \dfrac{0.63}{0.75} = 0.84$.

Worked example 10: Testing independence

$P(A) = 0.3$, $P(B) = 0.5$ and $P(A \cup B) = 0.65$. Are A and B independent?

Use $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.3 + 0.5 - 0.65 = 0.15$. If the events are independent, $P(A) \times P(B) = 0.3 \times 0.5 = 0.15$. These are equal, so A and B **are**

independent.

Worked example 11: Without replacement with three branches

A drawer has 3 black, 2 white and 1 grey sock. Two are drawn without replacement. Find the probability that they match.

Black pair: $\frac{3}{6} \times \frac{2}{5} = \frac{6}{30}$. White pair: $\frac{2}{6} \times \frac{1}{5} = \frac{2}{30}$. Grey cannot form a pair because there is only one. Total $= \frac{8}{30} = \frac{4}{15}$.

Worked example 12: Conditional probability on a Venn diagram

In a Venn diagram of 50 people, A only is 12, both is 8, B only is 15 and neither is 15. Find $P(A \mid B)$ and $P(B \mid A)$.

B has $8 + 15 = 23$ people, of whom 8 are also in A: $P(A \mid B) = \frac{8}{23}$. A has $12 + 8 = 20$ people, of whom 8 are also in B: $P(B \mid A) = \frac{8}{20} = \frac{2}{5}$. The two conditional probabilities are different because the group you are restricting to is different.

Worked example 13: Expected frequencies from conditional probability

In a town, $P(\text{rain}) = 0.2$ and $P(\text{umbrella} \mid \text{rain}) = 0.8$. On how many of 150 days do you expect it to rain and a person to carry an umbrella?

$P(\text{rain and umbrella}) = 0.2 \times 0.8 = 0.16$. Expected days $= 150 \times 0.16 = 24$. Multiplying the conditional probability by the probability of the condition gives the joint probability, which you can scale up by the number of trials.

Worked example 14: Reading both conditional probabilities from one table

In a survey of 200 adults, 80 own a bike, 120 own a car and 50 own both. Find $P(\text{bike} \mid \text{car})$ and $P(\text{car} \mid \text{bike})$ and explain why they are different.

Among the 120 car owners, 50 also own a bike, so $P(\text{bike} \mid \text{car}) = \frac{50}{120} = \frac{5}{12}$. Among the 80 bike owners, 50 also own a car, so $P(\text{car} \mid \text{bike}) = \frac{50}{80} = \frac{5}{8}$. They are different because the groups you restrict to have different sizes: there are more car owners than bike owners, so a smaller fraction of car owners also own a bike.

Worked example 15: A two-stage problem with a final decision

Box A has 2 red and 3 green sweets; box B has 4 red and 1 green. A die is rolled: a six chooses box A, anything else chooses box B. A sweet is taken. Find the probability that it is green, and that it came from box A given that

it is green.

$P(\text{A}) = \frac{1}{6}$ and $P(\text{B}) = \frac{5}{6}$. $P(\text{green}) = \frac{1}{6} \times \frac{3}{5} + \frac{5}{6} \times \frac{1}{5} = \frac{3}{30} + \frac{5}{30} = \frac{8}{30} = \frac{4}{15}$.

Then $P(\text{A} \mid \text{green}) = \frac{3/30}{8/30} = \frac{3}{8}$. Even though box A has more green sweets, box B is chosen far more often, so a green sweet is less likely to come from A.

Note: choosing between tree, table and Venn diagram

Use a **tree diagram** when events happen one after another and the second probability depends on the first. Use a **two-way table** when you have counts of two categories. Use a **Venn diagram** when the question gives the number in each set and in the overlap, or uses the words "both", "either" or "neither". Whichever you choose, label every branch or region, and write a final sentence that answers the question in words.

Method: a conditional probability question

1. Identify the condition. The words "given that" mean you restrict to that group.
2. Find the probability of *both* events together, $P(A \cap B)$, from the tree, table or Venn diagram.
3. Find the probability of the condition, $P(B)$, by adding every branch that gives B.
4. Divide: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ and simplify.

Watch out: conditional probability mistakes

- $P(A \mid B)$ and $P(B \mid A)$ are usually **different**. Check which one is being asked.
- On a tree without replacement, remember that the second branch probabilities change because the total decreases.
- When you divide, the denominator is the probability of the **given** event, not 1 and not the total.
- Do not assume events are independent. Test by checking whether $P(A \cap B) = P(A)P(B)$.

Did you know?

The Reverend Thomas Bayes, an English minister and mathematician who died in 1761, found a rule for working backwards from an effect to a cause. His result, now called Bayes\' theorem, was published after his death. Today it is used in medical tests, spam filters in email and in the way computers decide which videos to suggest to you.

Otto's tip

Read the word after "given" and circle it. That is your new total. The top of the fraction is the part of that group that also satisfies the other event.