Modelling with probability

S3.3 · Statistics · revision notes

Using probability to model real situations: stating the assumptions a model relies on, criticising them in context and explaining how more realistic assumptions would change the answers. Usually a short ‘state’, ‘explain’ or ‘comment’ part within a longer probability or distribution question.

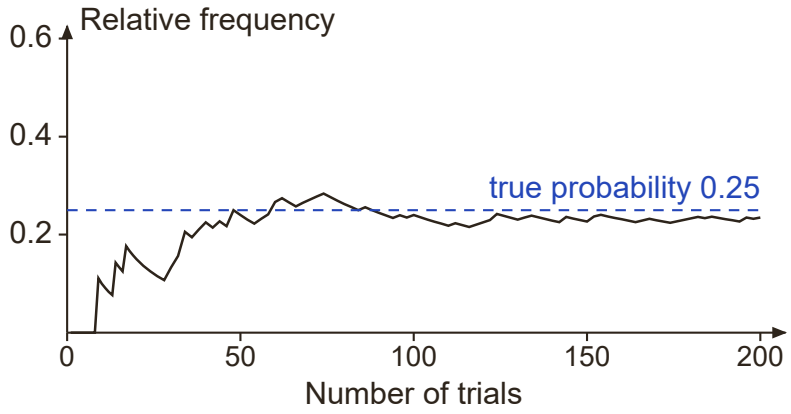
Key facts

- Expected frequency $= n \times P(\text{event})$
- More trials give a better estimate of a probability
- Refine a model by changing an assumption and checking it against data
- Relative frequency = frequency ÷ number of trials
- Common assumptions: fair, equally likely, independent, constant probability

Notes

Probability models

- A probability model lists the possible outcomes and gives each one a probability, e.g. for a fair six-sided die each score has probability $\frac{1}{6}$
- Probabilities can come from symmetry (equally likely outcomes), from past data (relative frequency = frequency ÷ number of trials) or from a theory.
- The more trials there are, the closer the relative frequency tends to be to the true probability.



Relative frequency settles down towards the true probability as the number of trials grows

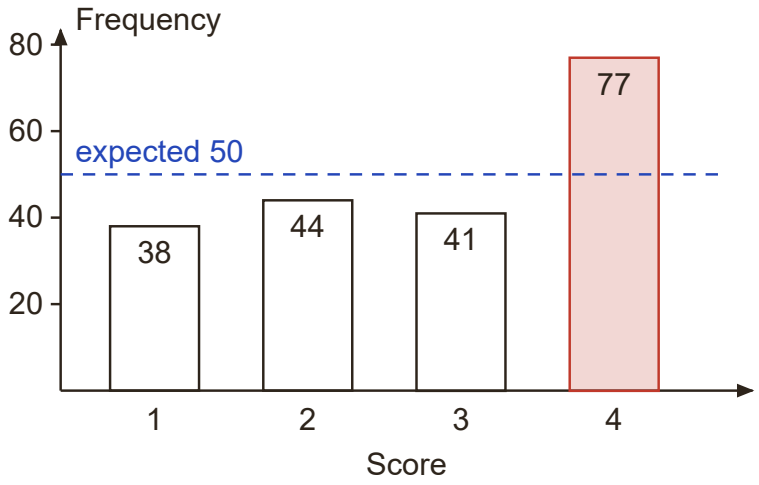
- Expected frequency = number of trials × probability. Observed frequencies vary around this by chance.

Assumptions

- Typical assumptions: the coin or die is fair; the outcomes are equally likely; events are independent (one result does not affect another); the probability stays the same each time.
- State assumptions in context, e.g. ‘whether it rains on one day does not affect whether it rains the next day’.

Criticising and refining a model

- Compare the model's expected frequencies with observed data. Large differences suggest the model is not appropriate (formal tests come in hypothesis testing).



Observed against expected: 77 fours is far above 50, so the spinner looks biased

- Criticise an assumption by explaining why it may fail here: weather comes in spells, so wet days are not independent; the chance of rain varies from month to month, so it is not constant.
- Refine the model: use relative frequencies from data, different probabilities for different conditions, or conditional probabilities such as $P(\text{rain} \mid \text{rain the day before})$
- Say how the answer would change: if wet days tend to follow each other, a run of wet days is **more** likely than an independent model predicts.

How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own.

Expected frequencies and judging a model

3 marks | Grade C

1. Multiply the number of trials by each probability.
2. Compare observed with expected: is any difference large compared with the expected value?
3. If the model looks wrong, use the observed relative frequency as a better estimate.

Example. A spinner has four equal sections labelled 1, 2, 3 and 4. It is spun 200 times, with these results.

Score	1	2	3	4
Frequency	38	44	41	77

(a) Assuming the spinner is fair, find the expected frequency of each score.
(b) Comment on whether the spinner is fair.
(c) Suggest a better estimate of the probability of scoring 4.

Model answer.

(a) $200 \times \frac{1}{4} = 50$ for each score (B1)

(b) 77 is much larger than 50 (and the other scores are all below 50), so the spinner appears to be biased towards 4 (B1)

(c) $\dfrac{77}{200} = 0.385$ (B1)

Criticise an assumption and describe its effect

2 to 3 marks | Grade B

1. Name the assumption and say why it may fail, in context.
2. Calculate the probability the model gives, if asked.
3. State whether a more realistic model would give a higher or lower answer, and why.

Example. A student models the weather at a UK station in July by assuming that each day has probability 0.3 of rain, independently of other days.
(a) Using this model, find the probability that it rains on each of three given days in a row.
(b) Explain why the assumption of independence may not be realistic.
(c) Describe the likely effect on your answer to (a) of using a more realistic model.

Model answer.

(a) $0.3^3 = 0.027$ (B1)

(b) Weather tends to come in spells: if it rains one day, rain is more likely the next day, so the days are not independent (B1)

(c) Because wet days tend to follow each other, three wet days in a row is likely to be more probable than 0.027 (B1)

Compare two models

3 to 4 marks | Grade A

1. Work out the probability under each model.
2. For a refined model, use conditional probabilities along a tree.
3. Compare the answers and say which assumption explains the difference.

Example. Model 1: a football team wins each match with probability 0.6, independently of other matches. Model 2: the probability of a win is 0.7 if the team won its previous match and 0.45 if it did not. The team won its last match.
(a) Find the probability that the team wins both of its next two matches under each model.
(b) Explain the difference between the two answers.

Model answer.

(a) Model 1: $0.6 \times 0.6 = 0.36$ (B1). Model 2: 0.7×0.7 (M1) $= 0.49$ (A1)

(b) Model 2 does not assume that matches are independent: a win makes another win more likely (e.g. through confidence or form), so it gives a higher probability of two wins in a row (B1)

Don't lose marks

- Stating an assumption with no reference to the context.
- Deciding a model is wrong just because observed and expected frequencies are not exactly equal; some variation is natural.
- Not saying whether a more realistic model would give a higher or a lower probability.

More tips

Memory tricks

- Assumption checklist: is it fair? Are the trials independent? Is the probability the same every time?
- Assumptions need context: not just 'independent', but what is independent of what.
- Direction of the effect: outcomes that cluster together make runs more likely than an independent model predicts.

Exam technique

- 'Criticise' or 'evaluate' means give a specific reason why the assumption may fail in this situation.
- When describing a refinement, say exactly what you would change (e.g. use a different probability of rain for each month).

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade E — **Calculate expected frequencies from probabilities**
Expected frequency = number of trials × probability.

Grade D — **State an assumption made by a model**
E.g. the die is fair, outcomes are equally likely, or trials are independent, in context.

Grade C

Compare a model's predictions with observed data

Decide whether observed frequencies are close to the expected ones, and what that says about the model.

Grade B

Criticise an assumption in context

Give a specific reason why an assumption may not hold in the situation described.

Grade A

Refine a model with more realistic assumptions

Use relative frequencies or conditional probabilities instead of equal or constant probabilities.

Grade A*

Explain how changed assumptions affect probabilities

Say whether a more realistic model gives a higher or lower probability, and why.

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.