

MATHS4U / STUDENT REVISION NOTES

Statistics foundations

The bridge to Further Mathematics 9231

Twelve lessons covering the probability, distributions and inference ideas that support Further Statistics. Each lesson contains clear explanations, fully worked examples, practice with solutions, and a readiness checklist.

How to use these notes

Read a short section, cover the worked solution and attempt the example. Then try the practice questions independently. Compare methods as well as final answers, and revisit any item on the checklist that you cannot yet explain.

Quick route into continuous random variables: 01 $\rightarrow$ 02 $\rightarrow$ 03 $\rightarrow$ 06. Definite integration is needed for lesson 06. Continue with the remaining lessons as required by your Further Statistics teaching sequence.

Three connected stages

01-03 Data, probability and random variables

04-08 Distributions and probability models

09-12 Combining variables, sampling and inference

Original Maths4U teaching material. This is a prerequisite course, followed by the new methods taught in 9231. All probability models state their assumptions; numerical answers retain extra precision before final rounding.

The twelve lessons

01 Data summaries and variation	3
02 Probability and combinations	9
03 Discrete random variables	16
04 Binomial distribution	23
05 Geometric distribution	30
06 Continuous random variables	37
07 Normal distribution	43
08 Poisson distribution	50
09 Linear combinations of random variables	57
10 Sampling and estimation	64
11 Hypothesis testing foundations	71
12 Confidence intervals foundations	77

Use the contents links or the PDF bookmarks to jump to a lesson. Each lesson is also available as a separate printable file.

LESSON 01 9709 TO 9231

Data summaries and variation

Describe a dataset accurately and prepare the notation needed for probability distributions and estimation.

By the end of this lesson

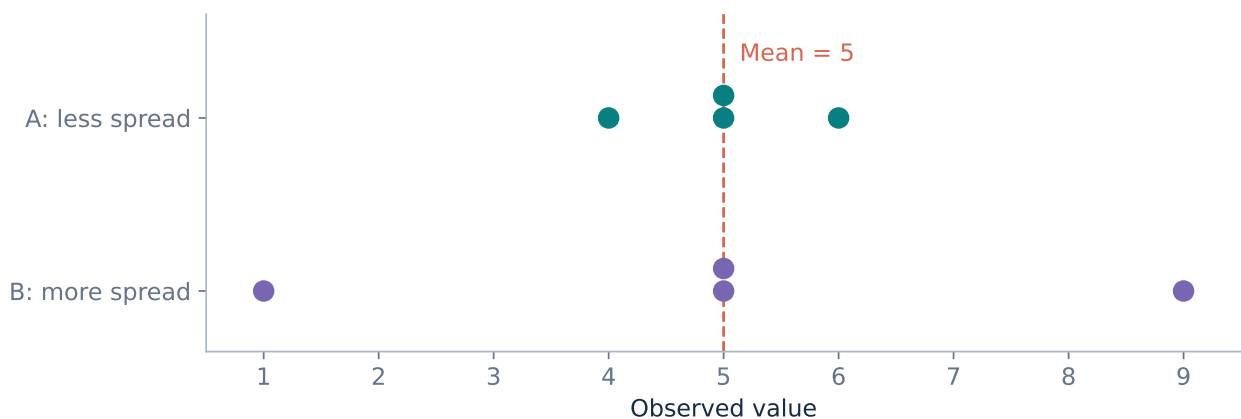
- Calculate and interpret a mean, median, variance and standard deviation.
- Use raw values, frequency tables and summary totals.
- Track the effect of coding and distinguish description from estimation.

Before you begin: Arithmetic with fractions, squares and square roots; Substitution into a formula.

1. What does a summary describe?

A dataset contains observations of a variable, such as journey time, height or the number of books borrowed. Record the unit and the number of observations, n . Before calculating, ask what one observation represents. A summary is useful only when its meaning remains attached to the numbers.

A measure of location describes where the data are centred. A measure of spread describes how much the observations differ. Two groups can have the same mean but very different consistency. Neither a mean alone nor a spread alone gives the whole description.



Datasets $A = (4, 5, 5, 6)$ and $B = (1, 5, 5, 9)$ both have mean 5. Their descriptive variances are 0.5 and 8, respectively.

2. Mean and median

The notation $\sum x$ means the sum of all observed values. The arithmetic mean, written $\bar{x}$, shares that total equally between the n observations. It uses every value, so a single unusually large value can move it substantially.

KEY FORMULA Arithmetic mean

$$\bar{x} = \frac{\sum x}{n}$$

For $n > 0$ observations. The mean has the same unit as the data.

To find the median, put the observations in ascending order. For odd n , select the value at position $(n+1)/2$. For even n , average the values at positions $n/2$ and $n/2 + 1$. The median identifies the middle of the ordered data and is less affected by an extreme value.

WORKED EXAMPLE Journey times

Six journey times are 12, 14, 15, 15, 17, 23 minutes. Find the mean and median.

1. Add all six times, then divide by six.

$$\sum x = 96, \quad \bar{x} = \frac{96}{6} = 16$$

2. The data are already ordered. Average the third and fourth values.

$$\text{Median} = \frac{15 + 15}{2} = 15$$

Answer. The mean is 16 minutes and the median is 15 minutes. The 23-minute journey pulls the mean above the median.

COMMON MISTAKE A position is not a data value

For six observations, the middle positions are 3 and 4. The median comes from the values in those positions, not from averaging the position numbers.

3. Variance and standard deviation

A deviation is an observation minus the mean. Positive and negative deviations cancel when added, so their sum cannot measure spread. Squaring makes every contribution nonnegative and gives larger deviations more influence.

KEY FORMULA Descriptive variance

$$v = \frac{\sum (x - \bar{x})^2}{n} = \frac{\sum x^2}{n} - \bar{x}^2$$

Here v describes these n observations. It is not yet an unbiased estimate of a population variance.

KEY FORMULA Standard deviation

$$SD = \sqrt{v}$$

Variance has squared units. Standard deviation returns to the original unit.

For the values 2, 4, 4, 6, the mean is 4. Their squared deviations are 4, 0, 0, 4, so $v = 8/4 = 2$ and the standard deviation is $\sqrt{2} \approx 1.41$. A variance of zero means every observation has the same value.

COMMON MISTAKE Two different sums

$\sum x^2$ means square each observation and then add. $(\sum x)^2$ means add first and then square the total. For 2, 4, 4, 6, these are 72 and 256.

METHOD TIP Keep precision until the end

Use an exact mean or retain calculator precision when finding variance. Premature rounding can noticeably affect the subtraction in the shortcut formula.

4. Frequencies and summary totals

A frequency f tells you how many times a value occurs. Multiply each value by its frequency when calculating the total. For the square total, use fx^2 : square the observation, then multiply by its frequency.

KEY FORMULA Frequency-table calculations

$$\bar{x} = \frac{\sum fx}{\sum f}, \quad v = \frac{\sum fx^2}{\sum f} - \bar{x}^2$$

The number of observations is $n = \sum f$, not the number of distinct values.

WORKED EXAMPLE Books borrowed

The values $x = 0, 1, 2, 3$ have frequencies $f = 1, 3, 4, 2$. Find the mean and standard deviation.

1. Build the two weighted totals.

$$\sum f = 10, \quad \sum fx = 17, \quad \sum fx^2 = 37$$

2. Apply the mean and variance formulas.

$$\bar{x} = \frac{17}{10} = 1.7, \quad v = \frac{37}{10} - 1.7^2 = 0.81$$

3. Take the square root.

$$SD = \sqrt{0.81} = 0.9$$

Answer. The mean is 1.7 books and the standard deviation is 0.9 books. A mean need not be a possible individual observation.

WORKED EXAMPLE Only the totals are known

Eight measurements have $\sum x = 120$ and $\sum x^2 = 1848$. Find their mean and standard deviation.

1. The mean needs only the total and the number of observations.

$$\bar{x} = \frac{120}{8} = 15$$

2. Use the square total to calculate variance.

$$v = \frac{1848}{8} - 15^2 = 231 - 225 = 6$$

3. Return to the measurement unit.

$$\text{SD} = \sqrt{6} \approx 2.45$$

Answer. The mean is 15 and the standard deviation is approximately 2.45, in the original measurement unit.

5. Combining and coding data

To combine groups, add their totals and their sizes. A group of four observations with mean 10 contributes 40 to the total. A group of six with mean 15 contributes 90. Their combined mean is $(40+90)/(4+6) = 13$. Simply averaging the two means would give equal weight to unequal groups.

Coding replaces inconvenient values with simpler ones. If $x = a + by$, the shift a changes location and the scale factor b changes spread. A negative scale reverses the order of values, but variance and standard deviation remain nonnegative.

KEY FORMULA Undoing a linear code

$$\bar{x} = a + b\bar{y}, \quad v_x = b^2 v_y, \quad \text{SD}_x = |b| \text{SD}_y$$

The same transformation must apply to every observation.

For example, if $y = x - 50$, $n = 5$, $\sum y = 10$ and $\sum y^2 = 30$, then $\bar{y} = 2$ and $v_y = 6 - 4 = 2$. Adding 50 back gives $\bar{x} = 52$, while $v_x = 2$. A shift leaves all distances from the mean unchanged.

6. Description and estimation use different denominators

This topic describes the observations you have, using division by n . Later, a random sample will help estimate the variance of a larger population. For independent observations from the same population with finite variance, the usual unbiased variance estimate uses $n - 1$, where $n > 1$.

KEY FORMULA Unbiased variance estimate

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = \frac{\sum x^2 - (\sum x)^2/n}{n - 1}$$

Use this when estimating population variance, as requested in sampling and estimation. Do not switch denominators without checking the purpose.

The observed mean $\bar{x}$ is a number calculated from data. In later topics, $\bar{X}$ will denote the random sample mean before the observations are collected. Keeping this distinction clear helps connect descriptive calculations to inference.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

For 3, 5, 7, 9, find the mean, median, descriptive variance and standard deviation.

1. Calculate the total and identify the middle two values.

$$\bar{x} = \frac{24}{4} = 6, \quad \text{Median} = \frac{5 + 7}{2} = 6$$

2. Square each observation and add.

$$\sum x^2 = 9 + 25 + 49 + 81 = 164$$

3. Calculate variance, then take its square root.

$$v = \frac{164}{4} - 6^2 = 5, \quad \text{SD} = \sqrt{5} \approx 2.24$$

Answer. Mean 6, median 6, variance 5 and standard deviation approximately 2.24.

PRACTICE Question 2

Ten observations have $\sum x = 80$ and $\sum x^2 = 670$. Find their mean and descriptive variance.

1. Use the total to find the mean.

$$\bar{x} = \frac{80}{10} = 8$$

2. Subtract the square of the mean from the mean square.

$$v = \frac{670}{10} - 8^2 = 67 - 64 = 3$$

Answer. The mean is 8 and the descriptive variance is 3.

PRACTICE Question 3

Values 0, 1, 2 have frequencies 2, 4, 4. Find the mean and standard deviation.

1. Use frequencies as weights.

$$\sum f = 10, \quad \sum fx = 12, \quad \sum fx^2 = 20$$

2. Find the mean and descriptive variance.

$$\bar{x} = 1.2, \quad v = \frac{20}{10} - 1.2^2 = 0.56$$

3. Take the square root.

$$\text{SD} = \sqrt{0.56} \approx 0.748$$

Answer. Mean 1.2 and standard deviation approximately 0.748.

PRACTICE Question 4

Five values are coded using $y = (x - 50)/2$. Given $\sum y = 10$ and $\sum y^2 = 30$, find the mean and descriptive variance of x .

1. Calculate summaries of the coded values.

$$\bar{y} = \frac{10}{5} = 2, \quad v_y = \frac{30}{5} - 2^2 = 2$$

2. Rewrite the code as $x = 50 + 2y$, then undo the shift and scale.

$$\bar{x} = 50 + 2(2) = 54, \quad v_x = 2^2(2) = 8$$

Answer. The original mean is 54 and the original variance is 8. The added 50 does not affect variance.

Ready to move on?

- ☐ I can explain what a mean and a standard deviation describe.
- ☐ I can distinguish a square total from the square of a total.
- ☐ I can use frequencies and summary totals without reconstructing the data.
- ☐ I can undo a linear code and combine group means correctly.
- ☐ I can choose between descriptive variance and an unbiased variance estimate.

Next: Probability and combinations. Probability supplies the weights that replace observed frequencies in a random-variable model.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

Cambridge 9709 syllabus, sections 5.1 and 6.4 · Cambridge 9231 prior-learning guidance

LESSON 02 9709 TO 9231

Probability and combinations

Translate chance statements into events and calculate probabilities for simple selections and repeated trials.

By the end of this lesson

- Use complements, unions, intersections and conditional probabilities.
- Distinguish mutually exclusive events from independent events.
- Calculate with probability trees and combinations.

Before you begin: Fractions, decimals and percentages; Basic set notation is helpful but introduced here.

1. Outcomes, events and complements

A random experiment has an uncertain outcome, such as the result of rolling a die. Its sample space is the set of all possible outcomes. An event is a subset of that space. For a fair six-sided die, the event $A = \text{“an even result”}$ is $\{2, 4, 6\}$.

Notation	Meaning
$A \cap B$	Both A and B occur
$A \cup B$	A or B or both occur
A'	A does not occur

KEY FORMULA Equally likely outcomes

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in the sample space}}$$

Use counting only when all outcomes in the chosen sample space are equally likely.

KEY FORMULA The complement rule

$$P(A') = 1 - P(A)$$

An event and its complement cover every outcome and cannot occur together.

Every probability lies between 0 and 1. Impossible events have probability 0 and certain events have probability 1. A complement often simplifies an “at least one” question: no successes may be easier to calculate than one, two or more successes separately.

2. Adding probabilities without double counting

Adding $P(A)$ and $P(B)$ counts the overlap twice. Subtract its probability once to obtain the probability of the union. The word “or” in this rule includes outcomes where both events occur.

KEY FORMULA General addition rule

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

This rule applies whether or not the events are independent.

For a fair die, let $A = \text{“even”}$ and $B = \text{“greater than 3”}$. Their intersection is $\{4, 6\}$ and their union is $\{2, 4, 5, 6\}$. Therefore $P(A \cup B) = 3/6 + 3/6 - 2/6 = 2/3$.

Mutually exclusive events cannot occur together, so $P(A \cap B) = 0$. Only in that case does the addition rule reduce to $P(A \cup B) = P(A) + P(B)$. Rolling a 1 and rolling a 6 on one die are mutually exclusive.

3. Conditional probability changes the reference group

$P(A | B)$ means the probability of A given that B has occurred. Once B is known, only outcomes inside B remain relevant. The part of this smaller reference group that also belongs to A is the intersection.

KEY FORMULA Conditional probability

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

The condition requires $P(B) > 0$. Reversing A and B usually changes the answer.

WORKED EXAMPLE Music and sport

Among 60 students, 24 study music, 30 play a sport and 18 do both. A student is chosen uniformly at random. Find the probability that they do either activity, then find the probability that they study music given that they play a sport.

- Count the union, removing the duplicated overlap.

$$P(M \cup S) = \frac{24 + 30 - 18}{60} = \frac{36}{60} = 0.6$$

- Given sport, the reference group is the 30 students who play a sport.

$$P(M | S) = \frac{18}{30} = 0.6$$

- For comparison, reversing the condition produces a different denominator.

$$P(S | M) = \frac{18}{24} = 0.75$$

Answer. The probability of either activity is 0.6. The conditional probability of music given sport is also 0.6, while the reverse conditional probability is 0.75.

METHOD TIP Read the condition first

In $P(A \mid B)$, identify the B group before choosing a denominator. A two-way table or a labelled Venn diagram can help organise the counts.

4. Repeated events and probability trees

A tree records successive choices. Probabilities leaving any one branch point sum to 1. Multiply along a complete path because later branch probabilities are conditional on the earlier outcomes. Add the probabilities of separate complete paths that lead to the requested event.

KEY FORMULA Multiplication along a path

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

This form requires $P(A) > 0$. Multiplying the two unconditional probabilities is justified when A and B are independent.

WORKED EXAMPLE Drawing without replacement

A bag contains 3 red and 2 blue tokens. Two tokens are drawn at random without replacement. Find the probability of different colours.

1. Different colours can occur through either path RB or BR.

$$P(RB) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$$

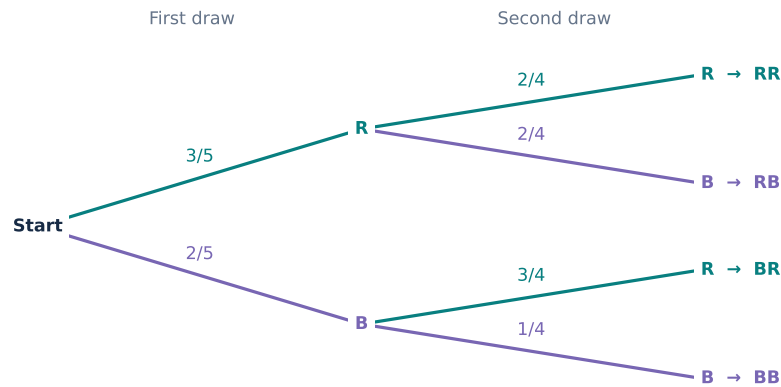
2. After a blue token is drawn, three red tokens remain among four tokens.

$$P(BR) = \frac{2}{5} \times \frac{3}{4} = \frac{3}{10}$$

3. These complete paths cannot both occur, so add them.

$$P(\text{different}) = \frac{3}{10} + \frac{3}{10} = \frac{3}{5}$$

Answer. The probability of different colours is $3/5 = 0.6$.



Two draws without replacement from 3 red and 2 blue tokens. The second-draw probabilities depend on the first colour.

COMMON MISTAKE Update the bag

Without replacement, the second denominator is 4. Its numerator also changes when the first token has the colour being counted.

5. Independence and mutual exclusivity

Two events are independent when learning that one has occurred does not change the probability of the other. If $P(B) > 0$, this means $P(A | B) = P(A)$. A useful test that also handles events of probability zero is the product rule below.

KEY FORMULA Independence test

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

This equality characterises independence of two events.

In the student survey, $P(M) = 24/60 = 0.4$ but $P(M | S) = 0.6$, so the events are dependent. The calculation identifies an association in this group; it does not establish that one activity causes the other.

Two mutually exclusive events with positive probabilities are dependent. If one occurs, the other becomes impossible. Their intersection probability is 0, while the product of their positive probabilities is greater than 0.

Replacing a token restores the original composition. With fresh random draws from the restored bag, the draw colours are independent, so $P(RR) = (3/5)^2 = 9/25$. Replacement alone should not replace a check that the drawing process is random.

6. Counting selections with combinations

The factorial $n!$ is the product $n(n-1) \cdots 2 \times 1$, with $0! = 1$. A combination counts an unordered selection of distinct objects. Selecting A and B produces the same pair as selecting B and A.

KEY FORMULA Selecting r objects from n

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

For integers $0 \leq r \leq n$, selecting without replacement from n distinct objects. Order does not matter.

WORKED EXAMPLE Choosing a pair

A group contains 5 boys and 4 girls. Every pair of students is equally likely to be selected. Find the probability that the pair contains one boy and one girl.

1. Count all unordered pairs from the nine students.

$$\binom{9}{2} = 36$$

2. Choose one student from each group.

$$\binom{5}{1} \binom{4}{1} = 5 \times 4 = 20$$

3. Divide the favourable count by the total count.

$$P(\text{one of each}) = \frac{20}{36} = \frac{5}{9}$$

Answer. The probability is $5/9$. Each pair is counted once in both numerator and denominator.

METHOD TIP Use one counting convention

Keep selections unordered throughout a combination calculation. Do not compare an ordered favourable count with an unordered total count.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

$P(A) = 0.6$, $P(B) = 0.4$ and $P(A \cap B) = 0.25$. Find $P(A \cup B)$ and $P(A | B)$. Are the events independent?

1. Remove the overlap when adding.

$$P(A \cup B) = 0.6 + 0.4 - 0.25 = 0.75$$

2. Divide by the probability of the condition.

$$P(A | B) = \frac{0.25}{0.4} = 0.625$$

3. Compare the intersection with the product.

$$P(A)P(B) = 0.24 \neq 0.25$$

Answer. The probabilities are 0.75 and 0.625. The events are dependent.

PRACTICE Question 2

Two tokens are drawn without replacement from 4 green and 3 yellow tokens. Find the probabilities of (a) two green tokens and (b) different colours.

1. For two green tokens, update both numerator and denominator after the first draw.

$$P(GG) = \frac{4}{7} \times \frac{3}{6} = \frac{2}{7}$$

2. Add the two different-colour paths.

$$P(GY) + P(YG) = \frac{4}{7} \times \frac{3}{6} + \frac{3}{7} \times \frac{4}{6} = \frac{4}{7}$$

Answer. (a) $2/7$. (b) $4/7$.

PRACTICE Question 3

A group of seven students contains four who study chemistry and three who do not. Three students are selected uniformly at random. Find the probability that exactly two study chemistry.

1. Count all possible groups of three.

$$\binom{7}{3} = 35$$

2. Choose two from the chemistry group and one from the other group.

$$\binom{4}{2} \binom{3}{1} = 6 \times 3 = 18$$

3. Use the ratio of the two counts.

$$P(\text{exactly two}) = \frac{18}{35}$$

Answer. The probability is $18/35 \approx 0.514$.

PRACTICE Question 4

Three independent trials each have success probability 0.3. Find the probability of at least one success.

1. The complementary event is no successes. Each failure has probability 0.7.

$$P(\text{no successes}) = 0.7^3 = 0.343$$

2. Subtract from one.

$$P(\text{at least one}) = 1 - 0.343 = 0.657$$

Answer. The probability is 0.657. Independence justifies multiplying the three failure probabilities.

Ready to move on?

- ☐ I can translate “and”, “or” and “not” into event notation.
- ☐ I can choose the correct denominator for a conditional probability.
- ☐ I can update probabilities after a draw without replacement.
- ☐ I can explain why independence and mutual exclusivity are different.
- ☐ I can count unordered selections and use a complement efficiently.

Next: Discrete random variables. A probability distribution assigns probabilities to numerical outcomes so that we can calculate a theoretical mean and variance.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

Cambridge 9709 syllabus, sections 5.2-5.3 · Cambridge 9231 prior-learning guidance

LESSON 03 9709 TO 9231

Discrete random variables

Build a numerical probability model and calculate its expectation, variance and probabilities of events.

By the end of this lesson

- Identify the support of a discrete random variable and check a probability distribution.
- Calculate event probabilities, expectations and variances.
- Find expectations of functions and work with a discrete uniform model.

Before you begin: Complements, addition and simple conditional probabilities; Weighted means, squares and square roots.

1. A random variable assigns numbers to outcomes

A random variable is a numerical rule applied to the outcome of a random experiment. We use a capital letter, such as X , for the variable and a lower-case letter, such as x , for a particular possible value. Before the experiment, X is uncertain; afterwards, an observed value is a number.

For two independent fair coin tosses, define X as the number of heads. The outcomes TT, HT, TH and HH give values 0, 1, 1, 2. Different outcomes can produce the same value of a random variable.

A discrete random variable has a finite or countably infinite set of possible values. The support consists of the values assigned positive probability. Counts are common examples, but a discrete variable can also take negative values or non-integer values, such as a net profit measured in pounds.

METHOD TIP Define the variable before calculating

“Let X be the number of heads in two tosses” is precise. “Let X be the coin” does not identify a numerical quantity or its support.

2. A valid probability distribution

A probability distribution lists the possible values together with their probabilities. The probability mass function is $p(x) = P(X = x)$. Probabilities are weights attached to values, so every weight must be nonnegative and the total weight must be exactly one.

KEY FORMULA Validity conditions

$$p(x) \geq 0 \quad \text{for every } x, \quad \sum_x p(x) = 1$$

The sum covers all possible values. Values outside the support have probability zero.

x	0	1	2
$P(X = x)$	1/4	1/2	1/4

This is the distribution for the number of heads in two fair independent tosses. The probability of one head is $1/2$ because two of the four equally likely coin outcomes produce that value. The three numerical values of X are therefore not equally likely.

COMMON MISTAKE Possible values and probabilities are different

A negative value of X is allowed, but a negative probability is not. A table can contain values such as $-1, 0, 2$ provided its probability weights satisfy the validity conditions.

3. Event probabilities and unknown constants

To find the probability of an event, add the probabilities of all values that satisfy it. Pay attention to endpoints. For integer-valued X , the events $X > 1$ and $X \geq 2$ are the same, but $X \geq 1$ also includes the value 1.

KEY FORMULA Adding the relevant masses

$$P(X \in A) = \sum_{x \in A} P(X = x)$$

A is any specified set of values. Include only values allowed by the distribution.

The identity $P(X \geq 1) = 1 - P(X = 0)$ applies when X is a non-negative integer-valued variable. If negative values are possible, the complementary event is $X < 1$, which may contain more than just zero.

WORKED EXAMPLE Finding a normalising constant

A variable has values 0, 1, 2, 3 with probabilities $k, 2k, 3k, 4k$, respectively. Find k and $P(X \geq 2)$.

1. The four probabilities must sum to one.

$$k + 2k + 3k + 4k = 10k = 1$$

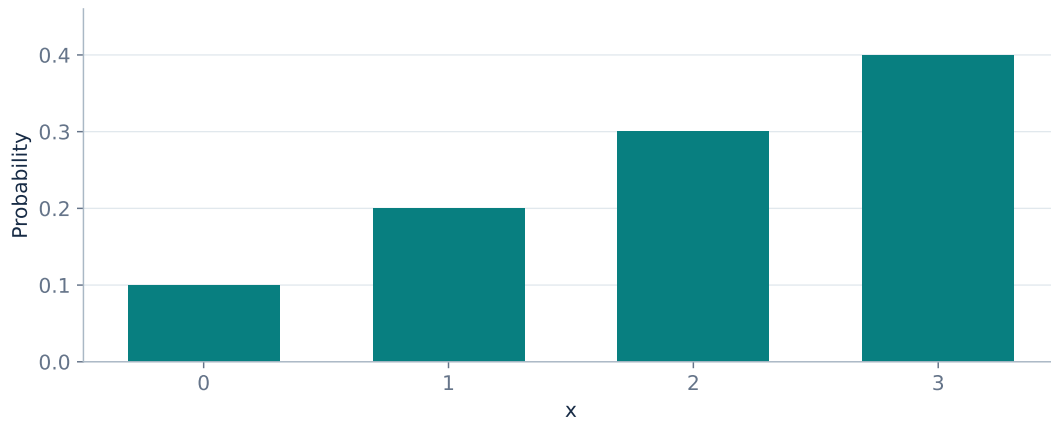
2. Solve for the constant and check that every resulting probability is nonnegative.

$$k = 0.1$$

3. The event includes the values 2 and 3.

$$P(X \geq 2) = 3k + 4k = 0.7$$

Answer. $k = 0.1$ and $P(X \geq 2) = 0.7$. The full distribution is 0.1, 0.2, 0.3, 0.4.



The probability mass function at $x = 0, 1, 2, 3$ has heights 0.1, 0.2, 0.3, 0.4. The four probabilities add to 1.

4. Expectation is a probability-weighted mean

The expectation $E(X)$ describes the theoretical mean of the model. Multiply each possible value by its probability, then add. This has the same structure as a frequency-weighted mean, but the probability weights already sum to one.

KEY FORMULA Expected value

$$\mu = E(X) = \sum_x xP(X = x)$$

For a finite distribution, sum all its values. For an infinite distribution, the required expectation must exist.

In repeated independent trials of the same model, the average observed value tends towards the expectation under the usual finite-mean conditions. The expectation need not occur in any single trial. A fair die has expectation 3.5, although a die never displays 3.5.

The most likely value is the mode, not necessarily the expectation. In the distribution 0, 1, 2, 3 with probabilities 0.1, 0.2, 0.3, 0.4, the mode is 3, while the expectation is 2.

COMMON MISTAKE No extra division

Once you have added the products $xP(X = x)$, the calculation is complete. Dividing by the number of possible values would incorrectly average weights that already sum to one.

5. The second moment and variance

The second moment, $E(X^2)$, is the weighted mean of the squared values. Square each value before multiplying by its probability. It is generally different from the square of the expectation.

KEY FORMULA Variance and standard deviation

$$\text{Var}(X) = E(X^2) - [E(X)]^2, \quad \text{SD}(X) = \sqrt{\text{Var}(X)}$$

Assume the second moment is finite. Variance is nonnegative and has squared units; standard deviation has the same unit as X .

WORKED EXAMPLE Calculating two moments

For $x = 0, 1, 2, 3$ with probabilities $0.1, 0.2, 0.3, 0.4$, find the expectation, variance and standard deviation.

1. Weight each value by its probability.

$$E(X) = 0(0.1) + 1(0.2) + 2(0.3) + 3(0.4) = 2$$

2. Weight each squared value by the same probability.

$$E(X^2) = 0^2(0.1) + 1^2(0.2) + 2^2(0.3) + 3^2(0.4) = 5$$

3. Subtract the square of the first moment, then take a square root.

$$\text{Var}(X) = 5 - 2^2 = 1, \quad \text{SD}(X) = 1$$

Answer. The mean is 2, the variance is 1 and the standard deviation is 1.

Variance measures spread around the expected value. It can also be written as $E[(X - \mu)^2]$. If your calculation gives a negative variance, check the probabilities, the squared values and the order of subtraction.

6. Uniform models and functions of a variable

A discrete uniform variable gives equal probability to each value in a finite support. If there are m values, each has probability $1/m$. “Uniform” refers to the equal probabilities; the possible numerical values still need to be stated.

WORKED EXAMPLE A fair four-score spinner

A fair spinner has scores 1, 2, 3, 4. Find the mean and variance of its score S .

1. Each score has probability one quarter.

$$E(S) = \frac{1 + 2 + 3 + 4}{4} = 2.5$$

2. Calculate the second moment.

$$E(S^2) = \frac{1 + 4 + 9 + 16}{4} = 7.5$$

3. Subtract the square of the expectation.

$$\text{Var}(S) = 7.5 - 2.5^2 = 1.25$$

Answer. The mean is 2.5 and the variance is 1.25.

Sometimes the quantity of interest is a function $g(X)$, such as a payoff or a squared distance. Apply the function to each possible value first, then average using the original probabilities.

KEY FORMULA Expectation of a function

$$E[g(X)] = \sum_x g(x)P(X = x)$$

The expectation must exist. In general, this is different from $g[E(X)]$.

For the two-coin variable, let $g(X) = (X - 1)^2$. Its transformed values are 1, 0, 1, so $E[g(X)] = 1(1/4) + 0(1/2) + 1(1/4) = 1/2$. But $E(X) = 1$ gives $g[E(X)] = 0$. Nonlinear transformations usually prevent interchanging the function and the expectation.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

X takes values $-1, 0, 2$ with probabilities $0.2, 0.5, 0.3$. Find its expectation, variance and standard deviation.

1. Calculate the two moments.

$$E(X) = -1(0.2) + 0(0.5) + 2(0.3) = 0.4$$

2. The negative value becomes positive when squared.

$$E(X^2) = 1(0.2) + 0(0.5) + 4(0.3) = 1.4$$

3. Find variance and its square root.

$$\text{Var}(X) = 1.4 - 0.4^2 = 1.24, \quad \text{SD}(X) \approx 1.11$$

Answer. Expectation 0.4, variance 1.24 and standard deviation approximately 1.11.

PRACTICE Question 2

Y takes values 0, 1, 2 with probabilities 0.2, 0.3, k . Find k , $E(Y)$ and $\text{Var}(Y)$.

1. Use the total-probability condition.

$$k = 1 - 0.2 - 0.3 = 0.5$$

2. Calculate the first and second moments.

$$E(Y) = 0.3 + 2(0.5) = 1.3, \quad E(Y^2) = 0.3 + 4(0.5) = 2.3$$

3. Calculate variance.

$$\text{Var}(Y) = 2.3 - 1.3^2 = 0.61$$

Answer. $k = 0.5$, expectation 1.3 and variance 0.61.

PRACTICE Question 3

Z is uniform on $\{1, 2, 3, 4, 5, 6\}$. Find $P(Z \geq 5)$, its expectation and its variance.

1. Two of the six equally likely values satisfy the event.

$$P(Z \geq 5) = \frac{2}{6} = \frac{1}{3}$$

2. Average the values and their squares.

$$E(Z) = \frac{21}{6} = 3.5, \quad E(Z^2) = \frac{91}{6}$$

3. Keep the fraction exact until the final subtraction.

$$\text{Var}(Z) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Answer. Probability $1/3$, expectation 3.5 and variance $35/12 \approx 2.92$.

PRACTICE Question 4

A game has score $X = 0, 1, 2$ with probabilities $0.5, 0.3, 0.2$. Its net payoff in pounds is $Y = 3X - 2$. Find the expected net payoff and the probability of making a profit.

1. Transform each score to obtain net payoffs of $-2, 1$ and 4 pounds.

$$E(Y) = (-2)(0.5) + 1(0.3) + 4(0.2) = 0.1$$

2. A profit occurs when the net payoff is positive, which corresponds to scores 1 and 2 .

$$P(Y > 0) = 0.3 + 0.2 = 0.5$$

Answer. The expected net payoff is £0.10 per game and the probability of a profit is 0.5 . An expected profit is not a guaranteed gain in one play.

Ready to move on?

- ☐ I can identify the support and check the probability weights.
- ☐ I can handle inequalities by adding the correct masses.
- ☐ I can distinguish $E(X^2)$ from $[E(X)]^2$.
- ☐ I can calculate and interpret expectation, variance and standard deviation.
- ☐ I can average transformed values to find $E[g(X)]$.

Next: The binomial distribution. The binomial model gives a formula for the distribution of a count of successes in independent repeated trials.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

Cambridge 9709 syllabus, section 5.4 · Cambridge 9231 prior-learning guidance

LESSON 04 9709 TO 9231

Binomial distribution

Model a fixed number of independent trials and calculate the number of successes.

By the end of this lesson

- Recognise when a binomial model is appropriate.
- Calculate exact, cumulative and complementary probabilities.
- Use the binomial mean and variance and interpret the results.
- Translate probability wording into the correct integer event.

Before you begin: Independent events and complementary probabilities; Combinations and factorial notation; Mean and variance of a discrete random variable.

1. Recognising a binomial experiment

A binomial variable counts successes in a fixed number of trials. A success is simply the outcome you decide to count: a faulty component can be a success in the mathematical model. Define the variable in words before writing a distribution. For example, X might count faulty components among 20 independently produced components.

- The number of trials, n , is fixed in advance.
- Each trial has two categories: success and failure.
- The trials are independent: one outcome does not change another trial's probability.
- The success probability, p , is constant on every trial.

KEY FORMULA Model and possible values

$$X \sim B(n, p), \quad X \in \{0, 1, \dots, n\}$$

n is a positive integer and $0 \leq p \leq 1$. Write $q = 1 - p$ for the failure probability.

Check these assumptions against the situation. Sampling without replacement from a small finite group usually changes the success probability and creates dependence. An exact binomial model then fails. A large population can sometimes justify an approximation, but that needs a reason; the word “random” alone does not establish independence.

2. The probability of exactly r successes

One particular sequence containing r successes and $n-r$ failures has probability $p^r(1-p)^{n-r}$. There are $\binom{n}{r}$ different positions for those successes. Add the probabilities of these mutually exclusive sequences by multiplying by the combinations factor.

KEY FORMULA Binomial probability

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}, \quad \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Use integer r with $0 \leq r \leq n$ and a valid binomial model. Remember $0! = 1$.

WORKED EXAMPLE Worked example 1: exactly two wins

An electronic game has independent rounds, each with win probability 0.4. Six rounds are played. Find the probability of exactly two wins.

1. Let X count wins. There are six independent trials with constant success probability, so $X \sim B(6, 0.4)$.
2. Substitute the number of trials, wins and failures.

$$P(X = 2) = \binom{6}{2} (0.4)^2 (0.6)^4$$

3. The 15 combinations count all possible positions of the two wins.

$$P(X = 2) = 15(0.16)(0.1296) = 0.31104$$

Answer. The probability is 0.31104, or approximately 0.311.

COMMON MISTAKE Do not calculate just one ordering

The expression $(0.4)^2(0.6)^4$ represents one specified ordering. “Exactly two wins” allows the wins in any two of the six positions, so the factor $\binom{6}{2}$ is essential.

3. Cumulative probabilities and complements

A cumulative probability adds probabilities for several possible counts. Because the values are integers, wording such as “more than” and “at least” changes the first included value. Write that value down before choosing a sum or calculator command.

Wording	Event	Useful equivalent
At most k	$X \leq k$	$\sum_{r=0}^k P(X = r)$
Fewer than k	$X < k$	$X \leq k - 1$
At least k	$X \geq k$	$1 - P(X \leq k - 1)$
More than k	$X > k$	$1 - P(X \leq k)$

WORKED EXAMPLE Worked example 2: at least two late deliveries

In an illustrative model, eight deliveries are independent and each is late with probability 0.25. Find the probability that at least two are late.

1. Let $X \sim B(8, 0.25)$. The complement of “at least two” is “zero or one”.

2. Calculate the shorter complementary sum.

$$P(X \geq 2) = 1 - [(0.75)^8 + 8(0.25)(0.75)^7]$$

3. Evaluate without rounding the two terms early.

$$P(X \geq 2) = 0.6329193115 \dots$$

Answer. The probability is approximately 0.633.

METHOD TIP Show the event before using technology

If a calculator gives $P(X \leq k)$, enter $k - 1$ when finding the complement for $P(X \geq k)$. Record your model and probability statement so the calculation remains understandable.

4. Mean, variance and standard deviation

The mean is the long-run average success count over many repetitions of the entire experiment. It need not be a possible individual count. Variance measures spread around that mean; standard deviation expresses the spread in the same units as X .

KEY FORMULA Binomial moments

$$E(X) = np, \quad \text{Var}(X) = np(1 - p), \quad \text{SD}(X) = \sqrt{np(1 - p)}$$

These formulas apply when $X \sim B(n, p)$.

WORKED EXAMPLE Worked example 3: interpreting the spread

Each of 20 seeds independently fails to germinate with probability 0.1. Let F count failures. Find its mean, variance and standard deviation.

1. Define failure to germinate as a success in the probability model: $F \sim B(20, 0.1)$.

2. Apply the moment formulas.

$$E(F) = 20(0.1) = 2, \quad \text{Var}(F) = 20(0.1)(0.9) = 1.8$$

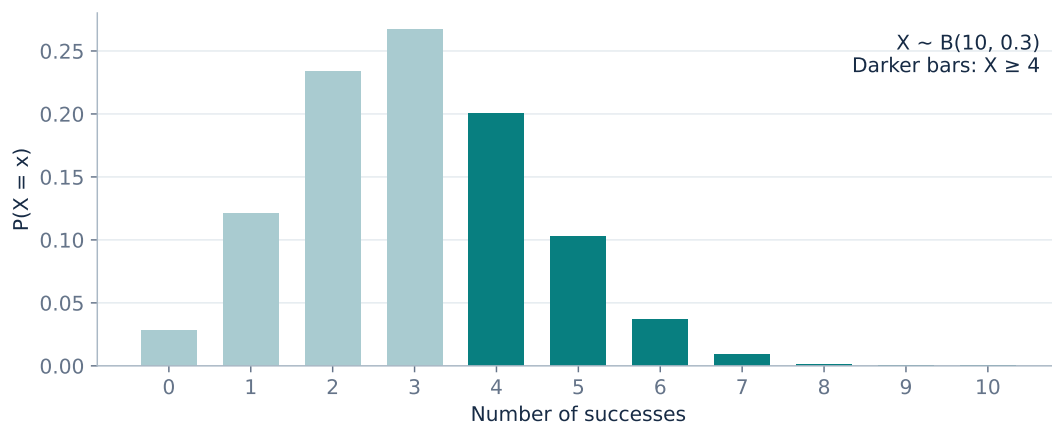
3. Take the square root of the variance.

$$SD(F) = \sqrt{1.8} = 1.3416 \dots$$

Answer. The average is 2 failures per group of 20; the standard deviation is about 1.34 failures. A particular group need not have exactly two failures.

If the mean and variance of a non-degenerate binomial model are known, dividing variance by mean gives $1 - p$. You can then recover p and n . Check that the resulting n is an integer and p lies between 0 and 1.

5. Reading a binomial distribution



The distribution $X \sim B(10, 0.3)$ has support 0 to 10. Highlighted bars represent $P(X \geq 4)$, the sum of the bars at 4 through 10.

Each bar gives the probability of one count, and all bar probabilities sum to 1. The mean is $np = 3$, but it is not a boundary splitting the probability equally. For p below one half the distribution can have a longer right tail. The binomial support always stops at n , regardless of the shape.

6. A reliable solution method

- Define the counted success and write $X \sim B(n, p)$.
- Check that the model assumptions match the context.

- Translate the wording into an integer event.
- Choose the exact formula, a cumulative sum or a complement.
- Keep extra precision during working, then interpret the result.

COMMON MISTAKE Variance is not standard deviation

$np(1-p)$ is the variance. Take its square root when standard deviation is requested. Also remember that a probability lies between 0 and 1; the mean count is not a probability.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

$X \sim B(10, 0.3)$. Find (a) $P(X = 4)$ and (b) $P(X \geq 4)$.

1. For the exact count, use four successes and six failures.

$$P(X = 4) = \binom{10}{4} (0.3)^4 (0.7)^6 = 0.200120949$$

2. For the upper tail, exclude counts zero through three.

$$P(X \geq 4) = 1 - \sum_{r=0}^3 \binom{10}{r} (0.3)^r (0.7)^{10-r} = 0.3503892816$$

Answer. (a) 0.200; (b) 0.350, to three significant figures.

PRACTICE Question 2

Twelve independent trials each have success probability 0.2. Find the probability of at least one success.

1. Let $X \sim B(12, 0.2)$. The complementary event is no successes.
2. Every trial must fail for that event.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0.8)^{12} = 0.9312805233 \dots$$

Answer. The probability is approximately 0.931.

PRACTICE Question 3

Stretch: a binomial variable has mean 6 and variance 4.2. Find n and p .

1. Write the two moment equations.

$$np = 6, \quad np(1 - p) = 4.2$$

2. Divide the variance equation by the mean equation.

$$1 - p = \frac{4.2}{6} = 0.7, \quad p = 0.3$$

3. Substitute into the mean equation.

$$n = \frac{6}{0.3} = 20$$

Answer. $n = 20$ and $p = 0.3$.

PRACTICE Question 4

A bag contains 3 red and 7 blue counters. Four counters are drawn without replacement. Explain why a binomial model for the red count is not exact. If each counter is replaced and the bag mixed after each draw, give the model, mean and variance.

1. Without replacement, the next red probability depends on earlier colours; the trials are dependent.

2. With replacement and mixing, each draw has red probability $3/10$ independently, so $R \sim B(4, 0.3)$.

3. Apply the moments.

$$E(R) = 4(0.3) = 1.2, \quad \text{Var}(R) = 4(0.3)(0.7) = 0.84$$

Answer. The replacement model is $B(4, 0.3)$, with mean 1.2 and variance 0.84.

Ready to move on?

- ☐ I can identify the four binomial conditions.
- ☐ I can explain the combinations factor.
- ☐ I can distinguish “at least”, “more than”, “at most” and “fewer than”.
- ☐ I can find and interpret mean, variance and standard deviation.
- ☐ I can check whether a numerical answer is plausible.

Next: Geometric distribution. Change from a fixed number of trials to waiting until the first success.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 5.4: discrete random variables](#) · [Cambridge 9231 prior-learning guidance](#)

LESSON 05 9709 TO 9231

Geometric distribution

Calculate how long a sequence of independent trials waits for its first success.

By the end of this lesson

- Recognise a geometric experiment and define its trial count.
- Calculate first-success and tail probabilities.
- Use the mean to interpret or determine the success probability.
- Distinguish a waiting-time variable from a binomial success count.

Before you begin: Independent trials and complementary probabilities; Powers and simple probability products; The meaning of an expected value.

1. Count trials up to and including success

A geometric variable counts how many trials are needed to obtain the first success. The stopping point is random: the first trial may succeed, or many failures may occur first. Use the same definition of success on every trial. As with a binomial model, success can mean any outcome being studied, including a fault.

- Trials are independent.
- Every trial has success or failure as its two possible categories.
- The success probability p is constant, with $0 < p \leq 1$.
- Continue until the first success, including that successful trial in the count.

KEY FORMULA Model and support

$$X \sim \text{Geo}(p), \quad X \in \{1, 2, 3, \dots\}$$

Cambridge counts trials including the first success. Write $q = 1 - p$ for the failure probability.

The smallest possible value is 1, because success on the first trial still uses one trial. There is no fixed largest possible value when $0 < p < 1$. Some sources define a geometric variable as the number of failures before success; that variable starts at 0. Always state your definition to avoid shifting every answer by one.

2. First success on a specified trial

The event $X = r$ requires exactly $r - 1$ failures, followed by one success. This ordering is compulsory: an earlier success would stop the experiment. Independence allows the probabilities to be multiplied. No combinations factor is needed because there is only one permitted success-failure pattern.

KEY FORMULA Geometric probability

$$P(X = r) = (1 - p)^{r-1}p = q^{r-1}p$$

For integer $r \geq 1$, with independent trials and constant success probability $0 < p \leq 1$.

WORKED EXAMPLE Worked example 1: first success on trial 4

An independent attempt succeeds with probability 0.3. Attempts continue until success. Find the probability that the first success occurs on attempt 4.

- Let X count attempts including the successful attempt. Then $X \sim \text{Geo}(0.3)$.
- Trial 4 must be preceded by three failures: failure, failure, failure, success.

$$P(X = 4) = (0.7)^3(0.3)$$

- Evaluate the product.

$$P(X = 4) = 0.1029$$

Answer. The probability is 0.1029, approximately 0.103.

COMMON MISTAKE The exponent counts failures

For first success on trial 4, use three failures, not four. The expression $(0.7)^4(0.3)$ would describe first success on trial 5.

3. After, at least and within

Tail probabilities have a particularly simple form. If $X > r$, the first r trials must all fail. There is no need to specify when success happens after them. If $X \geq r$, only the first $r - 1$ trials must fail: success on trial r is allowed.

KEY FORMULA After r trials

$$P(X > r) = q^r, \quad P(X \leq r) = 1 - q^r$$

Use integer $r \geq 0$. "Within the first r trials" means $X \leq r$.

KEY FORMULA At least r trials

$$P(X \geq r) = q^{r-1}$$

Use integer $r \geq 1$. Compare the included starting value carefully.

WORKED EXAMPLE Worked example 2: success within five attempts

Each independent attempt succeeds with probability 0.2. Find the probability of (a) needing more than five attempts and (b) succeeding within five attempts.

1. Let $X \sim \text{Geo}(0.2)$, so the failure probability is $q = 0.8$.

2. More than five attempts requires five initial failures.

$$P(X > 5) = (0.8)^5 = 0.32768$$

3. Within five attempts is the complementary event.

$$P(X \leq 5) = 1 - (0.8)^5 = 0.67232$$

Answer. (a) 0.32768; (b) 0.67232. They add to 1.

For an interval, subtract tails. First success from trial a through trial b , inclusive, means waiting at least a trials but not more than b . Thus $P(a \leq X \leq b) = q^{a-1} - q^b$ for integers $1 \leq a \leq b$. Listing the endpoint trials before calculating helps prevent an off-by-one error.

4. Expected waiting time

The expected value describes the average number of trials over many repetitions of the whole waiting experiment. A smaller success probability gives a longer expected wait. This average is not a deadline, and success does not become more likely merely because several failures have occurred.

KEY FORMULA Mean number of trials

$$E(X) = \frac{1}{p}$$

For $X \sim \text{Geo}(p)$ with $0 < p \leq 1$, counting the successful trial.

WORKED EXAMPLE Worked example 3: recovering the probability

A repeated independent experiment has a geometric waiting time with mean 5 trials. Find the success probability and the probability of success within the first three trials.

1. Use the expected-value formula to recover the probability.

$$\frac{1}{p} = 5 \implies p = 0.2$$

2. The failure probability is 0.8. Within three trials is the complement of three failures.

$$P(X \leq 3) = 1 - (0.8)^3 = 1 - 0.512$$

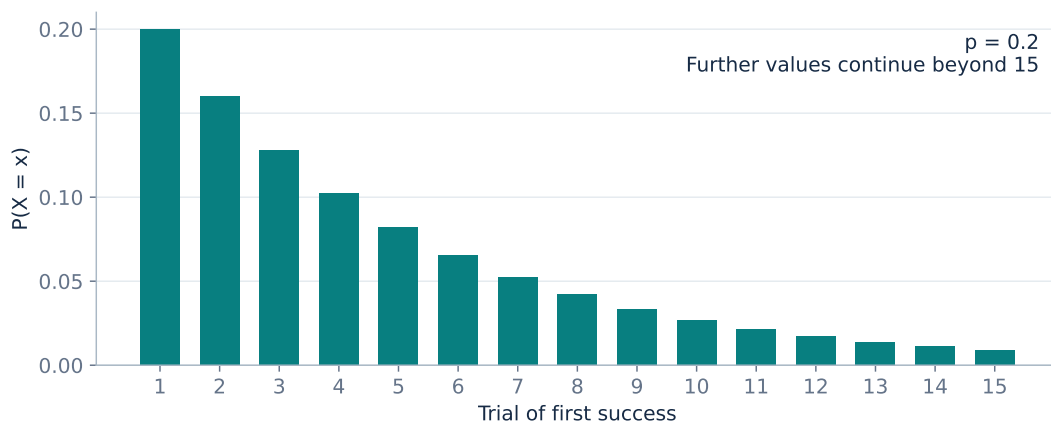
3. Evaluate and interpret the cumulative probability.

$$P(X \leq 3) = 0.488$$

Answer. The success probability per trial is 0.2, and the chance of success within three trials is 0.488.

METHOD TIP Check the scale

An expected waiting time of 5 trials does not mean that trial 5 is the most likely first-success trial. For $0 < p < 1$, trial 1 always has the largest individual first-success probability.

5. Shape and the effect of previous failures

First-success probabilities for $X \sim \text{Geo}(0.2)$, shown for trials 1 to 15. The omitted tail is

$$P(X > 15) = 0.8^{15} \approx 0.0352.$$

For $0 < p < 1$, each bar is q times the preceding bar, so probabilities decrease as the trial number increases. Nevertheless, many later outcomes together can have a substantial total probability. The visible bars need not sum to 1 when the graph stops at a finite trial number.

Independence also explains why past failures do not alter the next attempt: after ten failures, the next trial still succeeds with probability p . More generally, given r failures, the probability of waiting more than

another s trials is q^s . This property is called memorylessness; it follows from the same unchanged trial probabilities.

6. Choosing the right trial model

Feature	Binomial	Geometric
What is counted?	Successes in a fixed number of trials	Trials until first success
What is fixed?	Number of trials n	Success probability p ; stop at success
Possible values	$0, 1, \dots, n$	$1, 2, 3, \dots$
Mean	np	$1/p$

Both models need independent trials with constant success probability. First decide what the random variable counts; similar words in the story can lead to different models. If success becomes easier with practice, or each failure changes the next attempt, the constant-probability assumption needs reconsidering.

COMMON MISTAKE Do not attach an extra success factor to a tail

$P(X > r)$ only specifies initial failures, so it is q^r . Multiplying by p would instead require success immediately after those failures.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

$X \sim \text{Geo}(0.4)$. Find (a) $P(X = 3)$, (b) $P(X > 3)$ and (c) $P(X \geq 3)$.

1. First success on trial 3 needs two failures then success.

$$P(X = 3) = (0.6)^2(0.4) = 0.144$$

2. More than three means three failures; at least three means two failures.

$$P(X > 3) = (0.6)^3 = 0.216, \quad P(X \geq 3) = (0.6)^2 = 0.36$$

Answer. (a) 0.144; (b) 0.216; (c) 0.360.

PRACTICE Question 2

$Y \sim \text{Geo}(0.2)$. Find the probability that the first success is on trial 3, 4, 5 or 6.

1. The required event includes both endpoints: $3 \leq Y \leq 6$. Subtract the tail beyond trial 6 from the tail starting at trial 3.

$$P(3 \leq Y \leq 6) = P(Y \geq 3) - P(Y > 6)$$

2. Use the failure probability 0.8.

$$P(3 \leq Y \leq 6) = (0.8)^2 - (0.8)^6 = 0.377856$$

Answer. The probability is approximately 0.378.

PRACTICE Question 3

A geometric waiting time T has mean 4 trials. Find its success probability, $P(T = 4)$ and $P(T > 4)$.

1. Invert the mean to find the probability.

$$p = \frac{1}{4} = 0.25$$

2. Use three failures then success for the exact count.

$$P(T = 4) = (0.75)^3(0.25) = 0.10546875$$

3. Use four failures for the tail.

$$P(T > 4) = (0.75)^4 = 0.31640625$$

Answer. $p = 0.25$; $P(T = 4) \approx 0.105$; $P(T > 4) \approx 0.316$.

PRACTICE Question 4

Trials are independent with success probability 0.3. In experiment A, six trials are performed and successes counted. In experiment B, trials continue until the first success. State both models and find the chance that experiment B finishes within four trials.

1. Experiment A counts successes in fixed trials, so $A \sim B(6, 0.3)$. Experiment B counts waiting trials, so $B \sim \text{Geo}(0.3)$.

2. Finishing within four trials means at least one success in those trials.

$$P(B \leq 4) = 1 - (0.7)^4 = 0.7599$$

Answer. $A \sim B(6, 0.3)$, $B \sim \text{Geo}(0.3)$; the required probability is 0.7599.

Ready to move on?

- ☐ I count the successful trial and start the support at 1.
- ☐ I can explain why there is no combinations factor.
- ☐ I can distinguish first success on, after, and by a specified trial.
- ☐ I can use the geometric mean and interpret it correctly.
- ☐ I can choose between binomial and geometric variables.

Next: Continuous random variables. Move from probabilities at separate integer values to probabilities represented by areas under a density curve.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 5.4: discrete random variables](#) · [Cambridge 9231 prior-learning guidance](#)

LESSON 06 9709 TO 9231

Continuous random variables

Use area and integration to find probabilities, averages and percentiles for measurements.

By the end of this lesson

- Check and normalise a probability density function.
- Calculate interval probabilities and percentiles.
- Find the mean, variance and standard deviation by integration.
- Handle a simple density with an infinite support.

Before you begin: Integrate powers of x using definite integrals; Rearrange equations and work with exact fractions; Recall expectation and variance for a discrete random variable.

1. Probability is an area

A continuous model describes a measurement that can take any value in an interval, such as time or length. Its probability density function, written $f(x)$, shows how probability is distributed. The support is the set of possible values in the model; the density is zero outside it.

KEY FORMULA Conditions for a valid density

$$f(x) \geq 0, \quad \int_{\text{whole support}} f(x) \, dx = 1$$

Check nonnegativity throughout the support as well as total area. Both conditions are necessary.

A density height is not a probability. For example, a rectangle of height 2 and width 0.5 has area 1, so it can define a valid density. If X is measured in seconds, its density has units of inverse seconds; multiplying height by width produces a dimensionless probability.

KEY FORMULA Probability between two values

$$P(a < X < b) = \int_a^b f(x) \, dx$$

Integrate only where the interval meets the support. For a continuous variable, $P(X = a) = 0$, so including a single endpoint does not change the answer.

COMMON MISTAKE Density is not point probability

Do not write $P(X = a) = f(a)$. A single point has zero width and therefore zero area, even where the density is positive.

2. Find the constant before the probability

If a density contains an unknown constant, use total probability 1 to find it. Then substitute the constant into the density before calculating other areas. Sketching the support first helps you choose sensible integration limits.

WORKED EXAMPLE A rising density

The density is $f(x) = kx$ for $0 < x < 2$ and zero elsewhere. Find k and $P(0.5 < X < 1.5)$.

1. Set the area over the entire support equal to one.

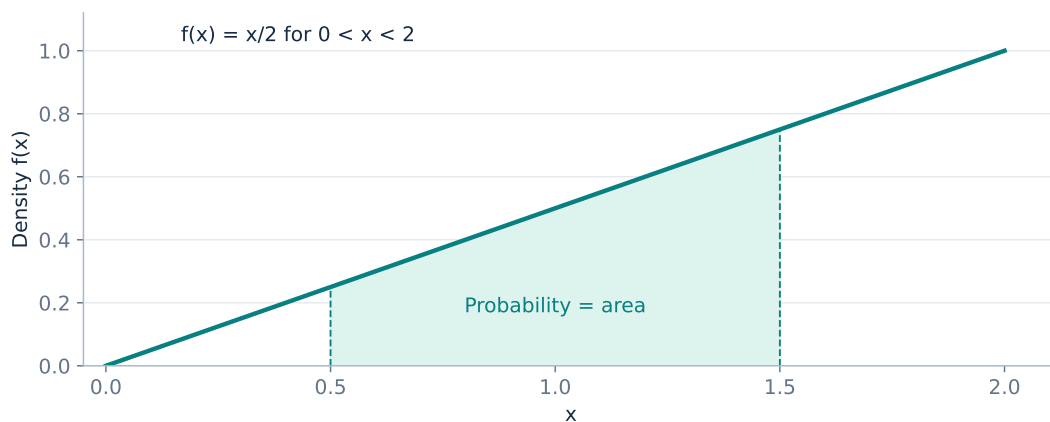
$$\int_0^2 kx \, dx = k \left[\frac{x^2}{2} \right]_0^2 = 2k = 1$$

2. Hence $k = \frac{1}{2}$. The resulting density is nonnegative throughout its support.

3. Integrate this density over the requested interval.

$$P(0.5 < X < 1.5) = \left[\frac{x^2}{4} \right]_{0.5}^{1.5} = \frac{2.25 - 0.25}{4} = \frac{1}{2}$$

Answer. The normalising constant is $k = \frac{1}{2}$ and the probability is 0.5.



For $f(x) = x/2$ on $0 < x < 2$, the shaded area between 0.5 and 1.5 is 0.5. The total triangular area is 1.

3. Mean and variance of a continuous model

Expectation uses the same weighting idea as a discrete distribution, with integration replacing summation. The mean is the long-run average of observations from the model. First calculate the two moments below, and then subtract the square of the mean.

KEY FORMULA Continuous moments

$$E(X) = \int x f(x) dx, \quad E(X^2) = \int x^2 f(x) dx$$

Both integrals are over the whole support, and the moments must exist.

KEY FORMULA Variance and standard deviation

$$\text{Var}(X) = E(X^2) - [E(X)]^2, \quad \text{SD}(X) = \sqrt{\text{Var}(X)}$$

Variance has squared measurement units. Standard deviation has the same units as X .

WORKED EXAMPLE Moments of the rising density

For $f(x) = x/2$ on $0 < x < 2$, find the mean, variance and standard deviation.

1. Multiply the density by x for the first moment.

$$E(X) = \int_0^2 \frac{x^2}{2} dx = \left[\frac{x^3}{6} \right]_0^2 = \frac{4}{3}$$

2. Multiply the density by x squared for the second moment.

$$E(X^2) = \int_0^2 \frac{x^3}{2} dx = \left[\frac{x^4}{8} \right]_0^2 = 2$$

3. Subtract using exact fractions.

$$\text{Var}(X) = 2 - \left(\frac{4}{3} \right)^2 = \frac{2}{9}, \quad \text{SD}(X) = \frac{\sqrt{2}}{3}$$

Answer. Mean = $\frac{4}{3}$, variance = $\frac{2}{9}$ and standard deviation ≈ 0.471 . The mean lies within the support and the variance is positive.

4. Medians and percentiles divide the area

The median m has half the probability to its left. More generally, a percentile is a value below which a specified proportion of the distribution lies. These are statements about area, so the median need not be halfway between the support endpoints.

KEY FORMULA Find a percentile

$$\int_{\text{lower support limit}}^q f(x) dx = p$$

For a median use $p = 0.5$; for the upper quartile use $p = 0.75$. Select a solution within the support.

For $f(x) = x/2$ on $0 < x < 2$, the area to the left of q is $q^2/4$. The median solves $m^2/4 = 1/2$, giving $m = \sqrt{2}$. The upper quartile solves $q^2/4 = 3/4$, giving $q = \sqrt{3}$. Both exceed 1, which fits the greater density towards the right.

5. When the support continues indefinitely

An infinite integration limit means a limit of finite integrals. Replace infinity by a number b , integrate, and then let b increase without bound. Total probability must converge to 1. A valid density does not automatically have a finite mean or variance: those integrals need checking separately.

WORKED EXAMPLE An infinite tail

Let $f(x) = 3/x^4$ for $x \geq 1$ and zero elsewhere. Check the density and find $P(X > 2)$, the mean and the variance.

1. The density is nonnegative for $x \geq 1$. Its total area is

$$\lim_{b \rightarrow \infty} \int_1^b 3x^{-4} dx = \lim_{b \rightarrow \infty} (1 - b^{-3}) = 1$$

2. For the tail probability, change the lower limit to two.

$$P(X > 2) = \lim_{b \rightarrow \infty} \left(\frac{1}{8} - b^{-3} \right) = \frac{1}{8}$$

3. Both moment integrals converge.

$$E(X) = \int_1^{\infty} 3x^{-3} dx = \frac{3}{2}, \quad E(X^2) = \int_1^{\infty} 3x^{-2} dx = 3$$

4. Use the variance identity.

$$\text{Var}(X) = 3 - \left(\frac{3}{2} \right)^2 = \frac{3}{4}$$

Answer. This is a valid density. Its tail probability is $\frac{1}{8}$, mean $\frac{3}{2}$ and variance $\frac{3}{4}$.

6. A reliable calculation routine

- Identify the support and set the density to zero elsewhere.
- Normalise if necessary, then select the integral that matches the question.
- Use exact fractions until the final numerical answer.
- Check probabilities lie between zero and one, percentiles lie in the support, and variances are non-negative.

METHOD TIP Check what you are integrating

For a probability integrate $f(x)$; for a mean integrate $xf(x)$; for a second moment integrate $x^2f(x)$. These are three different calculations.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

Let $f(x) = 2x$ for $0 < x < 1$, zero elsewhere. Find (a) $P(X > 0.5)$ and (b) the median.

1. Integrate from the threshold to the upper support limit.

$$P(X > 0.5) = \int_{0.5}^1 2x \, dx = [x^2]_{0.5}^1 = \frac{3}{4}$$

2. The median has area one half to its left.

$$\int_0^m 2x \, dx = m^2 = \frac{1}{2}$$

Answer. (a) $\frac{3}{4}$. (b) $m = 1/\sqrt{2} \approx 0.707$, taking the positive root within the support.

PRACTICE Question 2

A density is $f(x) = k(1+x)$ on $0 < x < 1$, zero elsewhere. Find k , the mean and the variance.

1. Normalise over the support.

$$k \int_0^1 (1+x) \, dx = \frac{3k}{2} = 1 \quad \Rightarrow \quad k = \frac{2}{3}$$

2. Calculate both moments.

$$E(X) = \frac{2}{3} \left(\frac{1}{2} + \frac{1}{3} \right) = \frac{5}{9}, \quad E(X^2) = \frac{2}{3} \left(\frac{1}{3} + \frac{1}{4} \right) = \frac{7}{18}$$

3. Subtract the squared mean.

$$\text{Var}(X) = \frac{7}{18} - \frac{25}{81} = \frac{13}{162}$$

Answer. $k = \frac{2}{3}$, mean = $\frac{5}{9}$ and variance = $\frac{13}{162}$.

PRACTICE Question 3

A density equals 4 on $1 < x < 1.25$ and zero elsewhere. Explain why it is valid, then find $P(1.1 < X < 1.2)$ and the median.

1. The density is nonnegative and its total area is $4(1.25 - 1) = 1$.

2. The requested probability is the rectangle area $4(1.2 - 1.1) = 0.4$.

3. For the median, $4(m - 1) = 0.5$, so $m = 1.125$.

Answer. The density is valid despite its height exceeding 1. The probability is 0.4 and the median is 1.125.

PRACTICE Question 4

Let $f(x) = 3x^2$ for $0 < x < 1$, zero elsewhere. Find the mean, variance and upper quartile.

1. Calculate the moments over the support.

$$E(X) = \int_0^1 3x^3 \, dx = \frac{3}{4}, \quad E(X^2) = \int_0^1 3x^4 \, dx = \frac{3}{5}$$

2. Find the variance.

$$\text{Var}(X) = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$$

3. Set the area to the left of the upper quartile equal to 0.75.

$$\int_0^q 3x^2 \, dx = q^3 = \frac{3}{4}$$

Answer. Mean = $\frac{3}{4}$, variance = $\frac{3}{80}$ and upper quartile $q = (3/4)^{1/3} \approx 0.909$.

Ready to move on?

- ☐ I can distinguish a density height from a probability.
- ☐ I can use the support to choose correct integration limits.
- ☐ I can calculate moments and subtract the squared mean.
- ☐ I can find a percentile by solving an area equation.
- ☐ I can evaluate a simple improper integral using a limit.

Next: The normal distribution. Apply the same area interpretation to an important continuous model, then use standardisation to calculate its probabilities.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

Cambridge 9709 syllabus, section 6.3: Continuous random variables · Cambridge 9231 prior-learning guidance

LESSON 07 9709 TO 9231

Normal distribution

Use standardisation, areas and percentiles, then approximate suitable binomial probabilities.

By the end of this lesson

- Standardise a normal variable and calculate tail or interval probabilities.
- Find percentiles and recover model parameters from probability information.
- Choose a suitable normal approximation to a binomial distribution.
- Apply the correct continuity correction.

Before you begin: Mean, variance and standard deviation; Continuous probability as area; Binomial mean and variance; Solving simultaneous linear equations.

1. Reading a normal model

A normal distribution is continuous, symmetric and bell-shaped. Its mean μ gives the centre, and its standard deviation σ controls the spread. Total area under its density curve is 1, with half on each side of the mean. A normal model can describe measurements, but the modelling assumption must be justified by the context.

KEY FORMULA Normal notation

$$X \sim N(\mu, \sigma^2), \quad -\infty < X < \infty$$

The second parameter is the variance. Take $\sigma > 0$ and use σ , not σ^2 , when standardising.

For $X \sim N(50, 64)$, the mean is 50 and the standard deviation is $\sqrt{64} = 8$. The mean has the units of X , while variance has squared units. Since the distribution is continuous, $P(X = c) = 0$ for any single value c ; changing an endpoint from $<$ to $\leq$ does not change a normal probability.

2. Convert to a standard normal variable

Standardisation expresses a measurement as a number of standard deviations above or below the mean. Subtracting the mean centres the distribution at zero; dividing by the standard deviation gives variance 1. A negative standardised value lies below the mean.

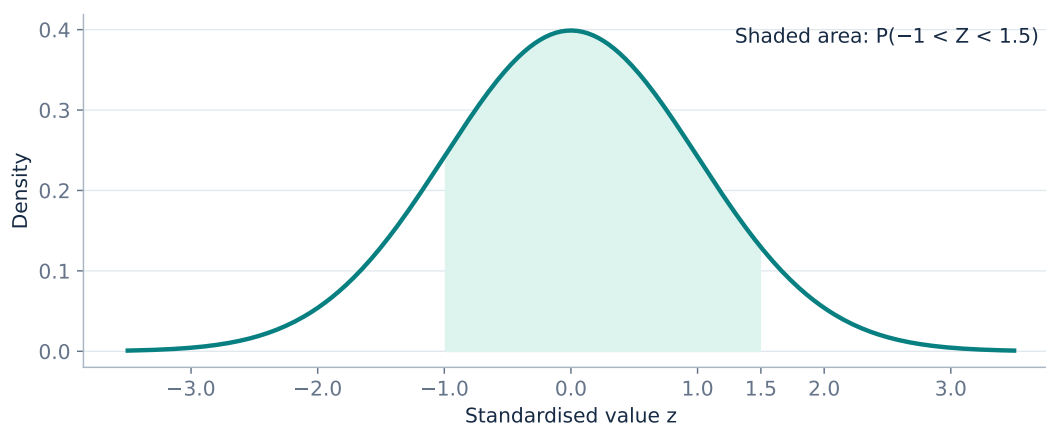
KEY FORMULA Standardisation

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Use when X is normally distributed. Write $\Phi(z) = P(Z \leq z)$ for the standard normal cumulative probability.

Area needed	Expression
Left of z	$\Phi(z)$
Right of z	$1 - \Phi(z)$
Between a and b	$\Phi(b) - \Phi(a)$
Left of $-z$	$\Phi(-z) = 1 - \Phi(z)$

Check what your table provides: some tables report a left cumulative area, while others give the area between zero and a positive value. For a left cumulative table, $\Phi(1.25) \approx 0.8944$. Sketch and shade the required region before reading a value or using a calculator.

3. Tails and intervals

The shaded standard normal area is $P(-1 < Z < 1.5) = \Phi(1.5) - \Phi(-1) \approx 0.7745$.

WORKED EXAMPLE Worked example 1: two areas under one model

$X \sim N(50, 64)$. Find (a) $P(X > 60)$ and (b) $P(42 < X < 62)$.

- The standard deviation is 8. For part (a), standardise the boundary and use the upper tail.

$$P(X > 60) = P\left(Z > \frac{60 - 50}{8}\right) = 1 - \Phi(1.25)$$

- Evaluate the right-hand area.

$$P(X > 60) = 0.1056498 \dots$$

3. For part (b), standardise both endpoints.

$$\frac{42 - 50}{8} = -1, \quad \frac{62 - 50}{8} = 1.5$$

4. Subtract cumulative areas in the correct order.

$$P(42 < X < 62) = \Phi(1.5) - \Phi(-1) = 0.7745375 \dots$$

Answer. (a) 0.106; (b) 0.775, to three significant figures.

METHOD TIP Use the sketch as a check

The boundary 60 is above the mean 50, so the area to its right must be below 0.5. A result such as 0.8944 would identify the opposite tail.

4. Percentiles and unknown parameters

A percentile reverses the probability calculation: the area is known and the boundary is unknown. If $P(X \leq c) = a$, first find the standard normal value z_a with $\Phi(z_a) = a$. Then undo standardisation. For an upper-tail probability b , use left area $a = 1 - b$.

KEY FORMULA Inverse normal calculation

$$c = \mu + z_a \sigma, \quad \Phi(z_a) = a$$

For a normal variable and a cumulative probability $0 < a < 1$.

For example, the 90th percentile uses $z_{0.90} \approx 1.2816$, while an upper 5% cutoff uses $z_{0.95} \approx 1.645$. If both μ and σ are unknown, two distinct percentile statements produce two simultaneous equations.

WORKED EXAMPLE Worked example 2: recovering centre and spread

A normal variable has 10th percentile 40 and 90th percentile 60. Find its mean and standard deviation.

1. By symmetry, $z_{0.10} = -1.2815516 \dots$ and $z_{0.90} = 1.2815516 \dots$

$$40 = \mu - 1.2815516\sigma, \quad 60 = \mu + 1.2815516\sigma$$

2. Add the equations to eliminate the standard deviation.

$$100 = 2\mu \implies \mu = 50$$

3. Subtract the equations, using unrounded values in the calculation.

$$20 = 2(1.2815516)\sigma \implies \sigma \approx 7.80304$$

Answer. The mean is 50 and the standard deviation is approximately 7.80.

COMMON MISTAKE Use the requested side of the cutoff

If 10% of observations lie above a cutoff, its cumulative probability is 0.90, not 0.10. The cutoff should be above the mean.

5. When a binomial normal approximation is useful

A binomial distribution can be approximated by a normal distribution when both expected successes and expected failures are sufficiently large. Common guidance is to check $np > 5$ and $n(1 - p) > 5$. These conditions guide an approximation; they do not make the two distributions identical. A large n alone is not enough if p is extremely close to 0 or 1.

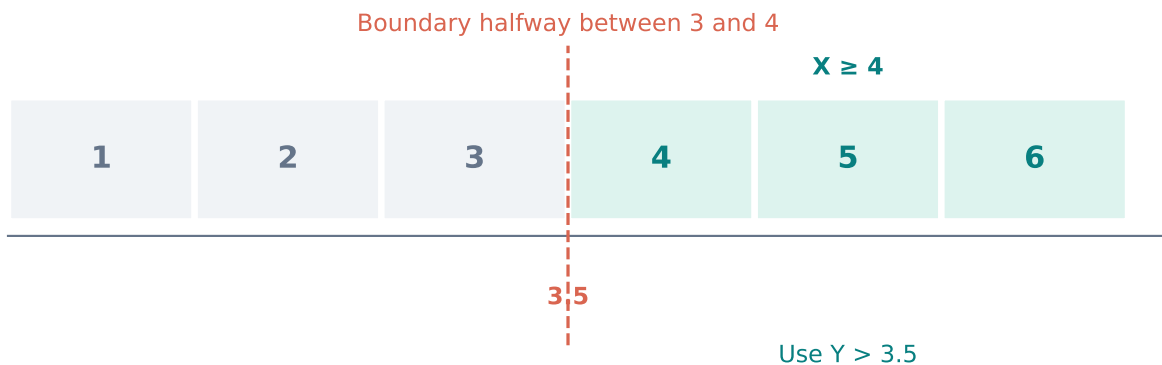
KEY FORMULA Match the binomial moments

$$X \sim B(n, p) \implies Y \sim N(np, np(1 - p))$$

Use Y as an approximation when both np and $n(1 - p)$ are sufficiently large. Apply continuity correction before standardising.

The binomial variable is integer-valued, whereas the normal variable is continuous. Represent each integer count by a unit-width interval centred on it. A continuity correction moves the probability boundary to the appropriate edge of this interval.

6. Continuity correction



To include integer counts $X \geq 4$, start at the left edge of the box centred at 4: use the continuous region $Y > 3.5$.

Binomial event	Normal approximation region
$X \leq k$	$Y < k + 0.5$
$X < k$	$Y < k - 0.5$
$X \geq k$	$Y > k - 0.5$
$X > k$	$Y > k + 0.5$
$a \leq X \leq b$	$a - 0.5 < Y < b + 0.5$

WORKED EXAMPLE Worked example 3: a corrected upper tail

$X \sim \text{B}(100, 0.4)$. Use a normal approximation to estimate $P(X \geq 48)$.

1. Check the conditions and match the mean and variance.

$$np = 40, \quad n(1 - p) = 60, \quad Y \sim \text{N}(40, 24)$$

2. Count 48 is included, so use the boundary halfway between 47 and 48.

$$P(X \geq 48) \approx P(Y > 47.5)$$

3. Standardise using the square root of the variance.

$$P(Y > 47.5) = 1 - \Phi\left(\frac{47.5 - 40}{\sqrt{24}}\right) \approx 0.0629$$

Answer. The approximate probability is 0.0629.

For strict inequalities, identify the last or first included integer before correcting. For example, $X < 8$ includes at most 7, so its upper boundary is 7.5. Correcting directly to 8.5 would include an unwanted integer.

7. Final checks before accepting an answer

- Have you identified σ from the variance?
- Does your shaded region match the required tail or interval?
- For inverse calculations, does the percentile lie on the correct side of the mean?
- For a binomial approximation, have you checked both conditions and corrected before standardising?

Normal areas will reappear in sampling distributions, confidence intervals and hypothesis tests. Keeping the model, standardised boundary and area statement visible now makes those later methods much easier to follow.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

$Y \sim N(60, 100)$. Find (a) $P(Y > 75)$ and (b) $P(45 < Y < 70)$.

1. The standard deviation is 10. Standardise each boundary.

$$P(Y > 75) = 1 - \Phi(1.5) = 0.0668072 \dots$$

2. For the interval, subtract the two cumulative probabilities.

$$P(45 < Y < 70) = \Phi(1) - \Phi(-1.5) = 0.7745375 \dots$$

Answer. (a) 0.0668; (b) 0.775.

PRACTICE Question 2

For $Y \sim N(60, 100)$, find (a) the 90th percentile and (b) the cutoff with 5% of values above it.

1. Use the inverse formula with standard deviation 10.

$$c_{0.90} = 60 + 1.2815516(10) = 72.815516 \dots$$

2. An upper 5% tail has left cumulative area 0.95.

$$c_{0.95} = 60 + 1.6448536(10) = 76.448536 \dots$$

Answer. (a) Approximately 72.8; (b) approximately 76.4.

PRACTICE Question 3

$X \sim B(80, 0.25)$. Justify a normal approximation and estimate (a) $P(X \leq 25)$ and (b) $P(16 \leq X \leq 24)$.

1. Both expected counts exceed 5.

$$np = 20, \quad n(1 - p) = 60, \quad Y \sim N(20, 15)$$

2. Include count 25 using the boundary 25.5.

$$P(X \leq 25) \approx \Phi\left(\frac{5.5}{\sqrt{15}}\right) \approx 0.922$$

3. Include both integer endpoints with outward half-unit corrections.

$$P(16 \leq X \leq 24) \approx \Phi\left(\frac{4.5}{\sqrt{15}}\right) - \Phi\left(\frac{-4.5}{\sqrt{15}}\right) \approx 0.755$$

Answer. The approximating model is $N(20, 15)$; the estimates are (a) 0.922 and (b) 0.755.

PRACTICE Question 4

$T \sim N(100, 16)$. A student standardises 108 as $(108 - 100)/16 = 0.5$. Correct the calculation and find $P(T > 108)$. Is continuity correction needed?

1. The second parameter is variance, so divide by its square root.

$$z = \frac{108 - 100}{\sqrt{16}} = 2$$

2. Use the upper-tail probability.

$$P(T > 108) = 1 - \Phi(2) = 0.0227501 \dots$$

3. The original variable is already continuous and normal. No continuity correction is needed.

Answer. $P(T > 108) \approx 0.0228$, with no continuity correction.

Ready to move on?

- ☐ I use standard deviation when standardising.
- ☐ I can calculate left-tail, right-tail and interval areas.
- ☐ I can find an inverse percentile and solve for unknown parameters.
- ☐ I can check when a binomial approximation is suitable.
- ☐ I can apply continuity correction to strict and inclusive inequalities.

Next: Poisson distribution. Build another essential discrete model, this time for events counted in a fixed interval.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 5.5: normal distribution](#) · [Cambridge 9231 prior-learning guidance](#)

LESSON 08 9709 TO 9231

Poisson distribution

Model event counts in a fixed interval and turn model probabilities into expected frequencies.

By the end of this lesson

- Recognise the assumptions of a Poisson model.
- Match the parameter to the interval and calculate probabilities.
- Use Poisson moments and recover a parameter from a zero-count probability.
- Calculate expected frequencies for later goodness-of-fit work.

Before you begin: Discrete random variables and probability sums; Complementary probabilities; Factorials, exponentials and logarithms; Mean, variance and standard deviation.

1. What a Poisson variable counts

A Poisson variable counts events in a fixed interval of time, length, area or volume. Examples might include calls arriving in ten minutes or flaws along one metre of material. The interval is fixed; the event count is random. Unlike the binomial distribution, the model has no fixed maximum count.

KEY FORMULA Model and support

$$X \sim \text{Po}(\lambda), \quad X \in \{0, 1, 2, \dots\}$$

Take $\lambda > 0$. The parameter λ is the expected event count in the particular interval being modelled.

- The average event rate is constant over the interval.
- Counts in disjoint intervals are independent.
- Events do not systematically arrive in clusters or batches.
- The chance of two or more events in a sufficiently short interval is negligible compared with the chance of one.

Use the context to assess these assumptions. A rush-hour surge may contradict a constant rate; linked equipment failures may contradict independence. Exactly scheduled arrivals every ten minutes have a predictable pattern and do not fit the random-arrival model. A stated average rate alone is not enough to establish a Poisson distribution.

2. Probabilities of individual counts

KEY FORMULA Poisson probability

$$P(X = r) = \frac{e^{-\lambda} \lambda^r}{r!}$$

For integer $r \geq 0$ when $X \sim \text{Po}(\lambda)$. The notation $e^{-\lambda}$ is also written $\exp(-\lambda)$.

The factorial applies to the observed count, not to the parameter. At zero, use $0! = 1$ and $\lambda^0 = 1$, giving $P(X = 0) = e^{-\lambda}$. Keep the full exponential value during calculation, especially if the probability will later be multiplied by a sample size.

WORKED EXAMPLE Worked example 1: exactly three faults

An illustrative model gives a mean of 2.4 faults per kilometre of cable. Assume faults follow a Poisson model. Find the probability of exactly three faults in one kilometre.

- Let X count faults in one kilometre, so $X \sim \text{Po}(2.4)$.
- Substitute the expected count and the required integer count.

$$P(X = 3) = \frac{e^{-2.4}(2.4)^3}{3!}$$

- Evaluate the power and factorial.

$$P(X = 3) = e^{-2.4} \frac{13.824}{6} = 0.2090141 \dots$$

Answer. The probability is approximately 0.209.

COMMON MISTAKE A mean can be non-integer

$\lambda = 2.4$ is an expected count, so it need not be an integer. The actual variable X still takes integer values only; $X = 2.4$ is not a possible count.

3. Match the parameter to the interval

Convert the rate before calculating any probability. If a process has constant mean rate a events per unit and the requested interval has length t in the same units, its expected count is at . A shorter interval gives a smaller Poisson parameter. Scale the expected count, not the finished probability.

KEY FORMULA Interval scaling

$$\lambda = \text{mean rate} \times \text{interval length}$$

The rate and length must use matching units, and the Poisson assumptions must remain reasonable throughout the interval.

WORKED EXAMPLE Worked example 2: at least one call

Calls follow a Poisson model with rate 6 per hour. Find the probability of at least one call in a 15-minute interval.

1. Convert 15 minutes to a fraction of an hour.

$$\lambda = 6 \left(\frac{15}{60} \right) = 1.5$$

2. Let $Y \sim \text{Po}(1.5)$. “At least one” is the complement of zero.

3. Use the zero-count probability.

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-1.5} = 0.7768698 \dots$$

Answer. The probability is approximately 0.777.

METHOD TIP Write the interval beside the parameter

Writing “ $\lambda = 1.5$ per 15-minute interval” makes the conversion visible. Using $\lambda = 6$ here would calculate an hourly probability instead.

4. Mean, variance and an unknown parameter**KEY FORMULA** Poisson moments

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda, \quad \text{SD}(X) = \sqrt{\lambda}$$

For $X \sim \text{Po}(\lambda)$. The numerical values of the mean and variance are equal; the standard deviation is their square root.

For $X \sim \text{Po}(4)$, the mean is 4 events and the standard deviation is 2 events. Equality of mean and variance is a feature of the model, not a rule for every observed dataset. Sample summaries fluctuate; compare the model with the process assumptions as well as the data.

A zero-count probability can determine the parameter. If $P(X = 0) = a$, where $0 < a < 1$, then $e^{-\lambda} = a$. Taking natural logarithms gives $-\lambda = \ln a$, so $\lambda = -\ln a$. The negative sign is necessary because $\ln a$ is negative for a probability between 0 and 1.

KEY FORMULA Recovering the parameter from zero counts

$$\lambda = -\ln(P(X = 0))$$

Use when the zero-count probability is known under a Poisson model and lies strictly between 0 and 1.

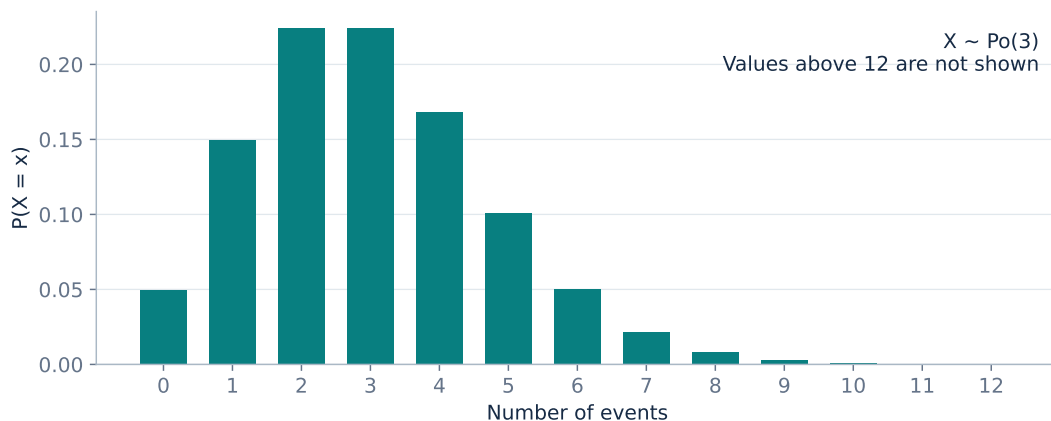
5. Several counts and complementary tails

To find a cumulative probability, add the individual probabilities for the allowed counts. For example, $P(X \leq 2)$ includes 0, 1 and 2. An upper tail is often shorter to calculate using a complement: $P(X \geq 3) = 1 - P(X \leq 2)$. Because the counts are integers, “more than 3” instead means $1 - P(X \leq 3)$.

KEY FORMULA A useful small cumulative sum

$$P(X \leq 2) = e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} \right)$$

For $X \sim \text{Po}(\lambda)$; the three terms correspond to counts 0, 1 and 2.



Probabilities for $X \sim \text{Po}(3)$, displayed for counts 0 to 12. Larger counts remain possible; the omitted probability above 12 is about 0.0000161.

The full collection of probabilities sums to 1. A truncated chart may omit a small positive tail. The largest bar does not have to identify a unique most likely value: when $\lambda = 3$, counts 2 and 3 have equal probabilities.

6. From model probabilities to expected frequencies

An expected frequency is the number predicted on average in a category across repeated observations. It differs from an observed frequency, which is an actual integer count. Later, a goodness-of-fit test compares observed frequencies with the frequencies predicted by a model.

KEY FORMULA Expected frequency

$$E_i = Np_i$$

N is the number of observations or intervals, and p_i is the model probability of category i . Categories should be disjoint and exhaustive.

WORKED EXAMPLE Worked example 3: predicted counts across intervals

The count in each of 120 independent intervals follows $\text{Po}(2)$. Find expected frequencies for counts 0, 1, 2 and 3 or more.

1. Calculate the four category probabilities.

$$p_0 = e^{-2}, \quad p_1 = 2e^{-2}, \quad p_2 = 2e^{-2}, \quad p_{3+} = 1 - 5e^{-2}$$

2. Multiply each probability by 120, retaining full precision.

$$E_0 = 16.2402\dots, \quad E_1 = 32.4805\dots, \quad E_2 = 32.4805\dots$$

3. Find the final category and check the total.

$$E_{3+} = 38.7988\dots, \quad \sum E_i = 120$$

Answer. The expected frequencies are approximately 16.240, 32.480, 32.480 and 38.799, respectively.

COMMON MISTAKE Do not round expected frequencies to whole counts

The values describe model averages and can be fractional. Keep their decimal precision for subsequent calculations; rounding each one to an integer changes the model comparison.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

$X \sim \text{Po}(4)$. Find (a) $P(X \leq 2)$ and (b) $P(X \geq 3)$.

1. Sum the probabilities for zero, one and two.

$$P(X \leq 2) = e^{-4} \left(1 + 4 + \frac{4^2}{2!} \right) = 13e^{-4} = 0.2381033\dots$$

2. The second event is the complement.

$$P(X \geq 3) = 1 - 13e^{-4} = 0.7618967\dots$$

Answer. (a) 0.238; (b) 0.762.

PRACTICE Question 2

A Poisson variable has $P(X = 0) = 0.2$. Find its parameter, mean and variance.

1. Equate the zero-count formula to the given probability.

$$e^{-\lambda} = 0.2$$

2. Take natural logarithms and change the sign.

$$\lambda = -\ln(0.2) = 1.6094379 \dots$$

3. Under a Poisson model, both $E(X)$ and $\text{Var}(X)$ equal λ .

Answer. The parameter, mean and variance are all approximately 1.61.

PRACTICE Question 3

Events follow a Poisson model at rate 12 per hour. For a 10-minute interval, find the model, the mean and standard deviation, and the probability of exactly one event.

1. Convert the interval to hours.

$$\lambda = 12 \left(\frac{10}{60} \right) = 2, \quad Y \sim \text{Po}(2)$$

2. Use the moment formulas.

$$E(Y) = 2, \quad \text{SD}(Y) = \sqrt{2} \approx 1.41$$

3. Apply the individual-count formula.

$$P(Y = 1) = 2e^{-2} = 0.2706706 \dots$$

Answer. $Y \sim \text{Po}(2)$; mean 2 events, standard deviation about 1.41 events, and $P(Y = 1) \approx 0.271$.

PRACTICE Question 4

Two hundred independent intervals each have a count $X \sim \text{Po}(3)$. Find (a) the expected number of zero-count intervals and (b) the probability of at least four events in one interval.

1. Multiply the zero-count probability by the number of intervals.

$$E_0 = 200e^{-3} = 9.9574137 \dots$$

2. Exclude counts zero through three for the upper tail.

$$P(X \geq 4) = 1 - e^{-3} \left(1 + 3 + \frac{9}{2} + \frac{27}{6} \right) = 0.3527681 \dots$$

Answer. (a) About 9.96 zero-count intervals on average; (b) approximately 0.353.

Ready to move on?

- ☐ I can state when a Poisson model is reasonable.
- ☐ I match the parameter to the requested interval before calculating.
- ☐ I can find individual, cumulative and complementary probabilities.
- ☐ I know the Poisson mean, variance and standard deviation.
- ☐ I can calculate expected frequencies and retain their precision.

Next: Linear combinations of random variables. Use distribution means and variances to study sums, differences and scaled measurements.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 6.1: Poisson distribution](#) · [Cambridge 9231 prior-learning guidance](#)

LESSON 09 9709 TO 9231

Linear combinations of random variables

Find the mean, spread and distribution of totals, differences and scaled random quantities.

By the end of this lesson

- Find the expectation and variance of a linear transformation.
- Combine independent random variables and distinguish repeated copies from a repeated variable.
- Calculate probabilities for sums and differences of independent normal variables.

Before you begin: Expectation, variance and standard deviation; Independence; Normal distributions and standardisation.

1. What is a linear combination?

A linear combination multiplies random variables by fixed constants and then adds the results, possibly with a fixed shift. Expressions such as $2X + 1$, $X - Y$ and $3X - 2Y + 5$ are linear combinations. Squaring a variable, as in X^2 , is a nonlinear transformation and requires different reasoning.

Expectation describes location and variance describes spread. Before using a rule, identify the random quantities and the fixed constants. A delivery time may vary from day to day, while a fixed five-minute preparation allowance does not introduce extra randomness.

KEY FORMULA Linearity of expectation

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

The constants a , b and c are fixed and the expectations exist. Independence is not needed.

For example, if two journey stages have expected times of 12 and 8 minutes, their combined expected time is 20 minutes. This remains true even when traffic conditions make their times dependent. Dependence matters when calculating the spread of the total.

2. Shifting and scaling one variable

Adding a constant shifts every possible value by the same amount. The mean shifts equally, so each distance from the mean remains unchanged. Multiplying by a constant scales those distances. Variance uses squared distances, so its multiplier is the square of the scale factor.

KEY FORMULA One-variable transformation rules

$$E(aX + b) = aE(X) + b, \quad \text{Var}(aX + b) = a^2 \text{Var}(X)$$

The constants are fixed and X has finite variance. A fixed shift contributes no variance.

KEY FORMULA Standard deviation after scaling

$$\text{SD}(aX + b) = |a| \text{SD}(X)$$

The absolute value keeps the standard deviation nonnegative, including when the scale factor is negative.

WORKED EXAMPLE A negative scale factor

$E(X) = 10$ and $\text{Var}(X) = 9$. Define $Y = 4 - 2X$. Find the expectation, variance and standard deviation of Y .

1. Apply the shift and scale to the expectation.

$$E(Y) = 4 - 2(10) = -16$$

2. Square the coefficient of X when transforming variance.

$$\text{Var}(Y) = (-2)^2(9) = 36$$

3. Take the positive square root.

$$\text{SD}(Y) = \sqrt{36} = 6$$

Answer. The expectation is -16 , the variance is 36 and the standard deviation is 6 . A negative mean is possible; a negative variance is not.

COMMON MISTAKE The shift has no spread

For $Y = 4 - 2X$, do not add 4 to the variance. Also, $\text{Var}(-2X)$ is $4 \text{Var}(X)$, not $-2 \text{Var}(X)$.

3. Variance of independent combinations

For independent variables, each source of variation contributes its own variance to a sum. The coefficients must be squared before the contributions are added. Subtracting an independent quantity still introduces uncertainty, so its variance contribution is positive.

KEY FORMULA Variance rule for independent variables

$$\text{Var}(aX + bY + c) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$$

Use this rule when X and Y are independent and have finite variances. Independence is a sufficient condition. Without a suitable justification, do not assume that variances add.

KEY FORMULA An independent difference

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$$

Here the coefficients are 1 and -1 , whose squares are both 1.

WORKED EXAMPLE Two independent measurements

X and Y are independent, with $E(X) = 12$, $\text{Var}(X) = 4$, $E(Y) = 8$ and $\text{Var}(Y) = 9$. Find the expectation and variance of $W = 3X - 2Y + 5$.

1. Combine the expectations with their original coefficients.

$$E(W) = 3(12) - 2(8) + 5 = 25$$

2. Use independence and squared coefficients for the variance.

$$\text{Var}(W) = 3^2(4) + (-2)^2(9) = 36 + 36 = 72$$

Answer. $E(W) = 25$ and $\text{Var}(W) = 72$. If needed, $\text{SD}(W) = \sqrt{72} \approx 8.49$.

METHOD TIP Combine variances before square roots

Standard deviations do not generally add. Calculate the total variance first, then take one square root to find the standard deviation.

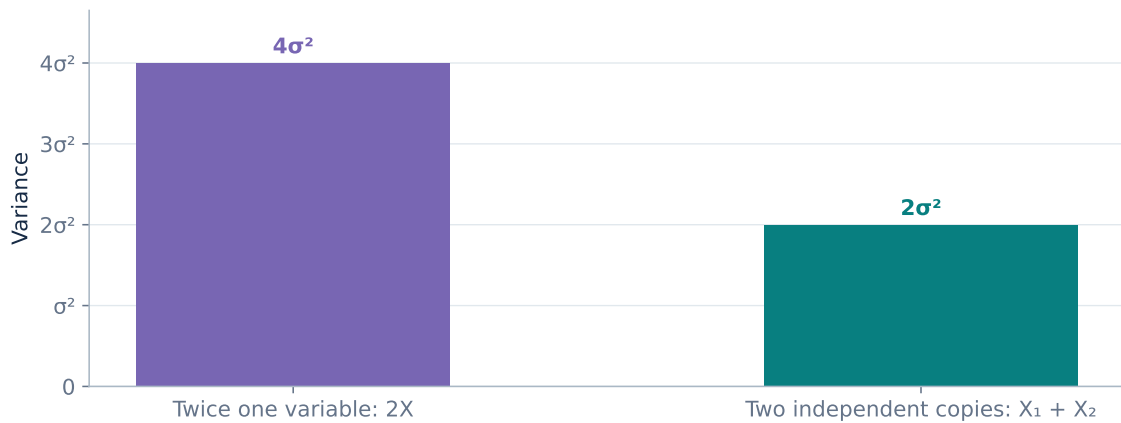
4. Doubling one value and adding two copies

The expression $2X$ means that one realised value of X is doubled. The expression $X_1 + X_2$ can represent two separate, independent observations from the same distribution. Both have the same expectation, but they generally have different variances.

KEY FORMULA Same mean, different variance

$$\text{Var}(2X) = 4\sigma^2, \quad \text{Var}(X_1 + X_2) = 2\sigma^2$$

$\text{Var}(X) = \sigma^2$ and X_1, X_2 are independent copies of X . Both expectations equal $2E(X)$.



Doubling one observation gives variance $4 \text{Var}(X)$. Adding two independent copies gives variance $2 \text{Var}(X)$.

Suppose a package mass has mean 100 g and variance 4 g^2 . Doubling one package mass gives mean 200 g and variance 16 g^2 . Adding the masses of two independent packages gives the same mean but variance 8 g^2 . The first operation doubles every deviation together; the second combines separate deviations.

COMMON MISTAKE The same variable is not an independent copy

Writing $X + X$ does not make its two terms independent. It is exactly $2X$. Use distinct subscripts and state independence when the model involves separate observations.

5. Combining independent normal variables

A linear combination of independent normal variables is normal. Determine its mean and variance using the same rules before calculating a probability. In the notation $N(\mu, \sigma^2)$, the second parameter is the variance, not the standard deviation.

KEY FORMULA Normal combination

$$aX + bY + c \sim N(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

$X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent. For probability standardisation, the resulting variance must be positive.

WORKED EXAMPLE Which measurement is larger?

Independent measurements satisfy $A \sim N(50, 9)$ and $B \sim N(44, 16)$. Find $P(A > B)$.

1. Turn the comparison into a single random variable: let $D = A - B$.

$$E(D) = 50 - 44 = 6, \quad \text{Var}(D) = 9 + 16 = 25$$

2. The difference is normal with standard deviation 5.

$$D \sim N(6, 25), \quad P(A > B) = P(D > 0)$$

3. Standardise and use normal symmetry. Here Φ is the standard normal cumulative distribution function.

$$P(D > 0) = P\left(Z > \frac{0 - 6}{5}\right) = P(Z > -1.2) = \Phi(1.2)$$

Answer. $P(A > B) \approx 0.8849$. It exceeds 0.5, as expected because A has the higher mean.

The expectation and variance rules do not require normality. Normality is the additional assumption that identifies the whole distribution and allows this exact normal probability calculation.

6. Totals, averages and a reliable method

If $X_1, \dots, X_n$ are independent observations from the same population, each with mean μ and variance σ^2 , their total has expectation $n\mu$ and variance $n\sigma^2$. Dividing the total by n creates the random sample mean $\bar{X}$.

KEY FORMULA The random sample mean

$$E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

For n independent observations with the same finite mean and variance. Normality is not needed for these two moment formulas.

For a probability question, first define the total or difference, then find its mean and variance, then identify its distribution. Finally, standardise using its standard deviation. Keeping these stages separate prevents using an original component's spread for the combined quantity.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

X and Y are independent, with $E(X) = 5$, $\text{Var}(X) = 2$, $E(Y) = 7$ and $\text{Var}(Y) = 4$. Find the expectation and variance of $S = 2X + Y - 3$.

1. Use linearity for the expectation.

$$E(S) = 2(5) + 7 - 3 = 14$$

2. Square the coefficients and add the independent variance contributions.

$$\text{Var}(S) = 2^2(2) + 1^2(4) = 12$$

Answer. The expectation is 14 and the variance is 12.

PRACTICE Question 2

$U \sim N(20, 16)$ and $V \sim N(17, 9)$ are independent. Find $P(U < V)$.

1. Define $D=U-V$ and find its normal distribution.

$$D \sim N(20 - 17, 16 + 9) = N(3, 25)$$

2. Translate the comparison and standardise using standard deviation 5.

$$P(U < V) = P(D < 0) = \Phi\left(\frac{0 - 3}{5}\right) = \Phi(-0.6)$$

Answer. $P(U < V) \approx 0.2743$.

PRACTICE Question 3

Three independent task times each have distribution $N(20, 4)$ minutes. Find the probability that their total exceeds 65 minutes.

1. The total is normal because its three components are independent and normal.

$$T \sim N(3 \times 20, 3 \times 4) = N(60, 12)$$

2. Use the total standard deviation, $\sqrt{12}$.

$$P(T > 65) = 1 - \Phi\left(\frac{65 - 60}{\sqrt{12}}\right) = 1 - \Phi(1.443 \dots)$$

Answer. The probability is approximately 0.0745.

PRACTICE Question 4

$E(X) = 3$ and $\text{Var}(X) = 5$. Let $A = 2X + 1$ and $B = X_1 + X_2 + 1$, where X_1, X_2 are independent copies of X . Compare their expectations and variances.

1. Both constructions add two copies of the mean and a fixed shift.

$$E(A) = E(B) = 2(3) + 1 = 7$$

2. A doubles the same variable, so its variance scales by four.

$$\text{Var}(A) = 2^2(5) = 20$$

3. B adds two independent variables, so their variances add.

$$\text{Var}(B) = 5 + 5 = 10$$

Answer. Both have expectation 7. Their variances are 20 for A and 10 for B, so A has the larger standard deviation.

Ready to move on?

- ☐ I can separate fixed constants from random quantities.
- ☐ I can explain why a shift leaves variance unchanged.
- ☐ I square coefficients when combining independent variances.
- ☐ I can distinguish doubling one observation from adding independent observations.
- ☐ I can form and standardise a normal total or difference.

Next: Sampling and estimation. The sample mean is a linear combination whose variance explains how increasing sample size improves precision.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 6.2](#) · [Cambridge 9231 prior-learning guidance](#)

LESSON 10 9709 TO 9231

Sampling and estimation

Connect individual measurements, random sample means and estimates of unknown population values.

By the end of this lesson

- Distinguish a population parameter from a sample statistic.
- Find the mean, variance and standard error of a sample mean.
- Choose an exact normal model or a justified CLT approximation.
- Estimate a population mean and variance from raw or summarised data.

Before you begin: Calculate means and descriptive variances; Use expectation and variance for independent sums; Standardise a normal random variable and find normal probabilities.

1. What is random, and what is fixed?

A population is the complete group or probability model we want to describe. A sample contains the observations available to us. A parameter describes the population: its mean μ and variance σ^2 are fixed quantities, even when we do not know their values.

Notation	Meaning
X or X_i	One random observation before its value is known.
$\bar{X}$	The random mean of a sample of size n .
x_i and $\bar{x}$	The observed values and their calculated mean.
μ, σ^2	Population mean and population variance.
s^2	An estimate of population variance from the observed sample.

A statistic is calculated from a sample. Before sampling it is random because another sample could give a different result. After observing the data, $\bar{x}$ is a particular number; it is used to estimate μ . Keep this distinction when writing probabilities: $P(\bar{X} > 54)$ describes possible future sample means.

2. Sampling assumptions matter

The formulas here assume independent observations from the same population, with finite mean and variance. Random selection supports a defensible link between the sample and population. A conve-

nience sample can systematically miss part of a population; taking more observations does not automatically remove that bias.

Independence means that knowing one observation does not change the distribution of another. Sampling with replacement from a finite population provides independent draws. Sampling without replacement is not exactly independent; the simple variance formula below then needs justification or an appropriate correction. In these notes, independence is given or assumed explicitly.

METHOD TIP A random sample can still look unusual

Random sampling does not make every sample perfectly representative. Sampling variability is precisely why we study the distribution of $\bar{X}$.

3. The distribution of sample means

KEY FORMULA Sample mean and its moments

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}, \quad E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

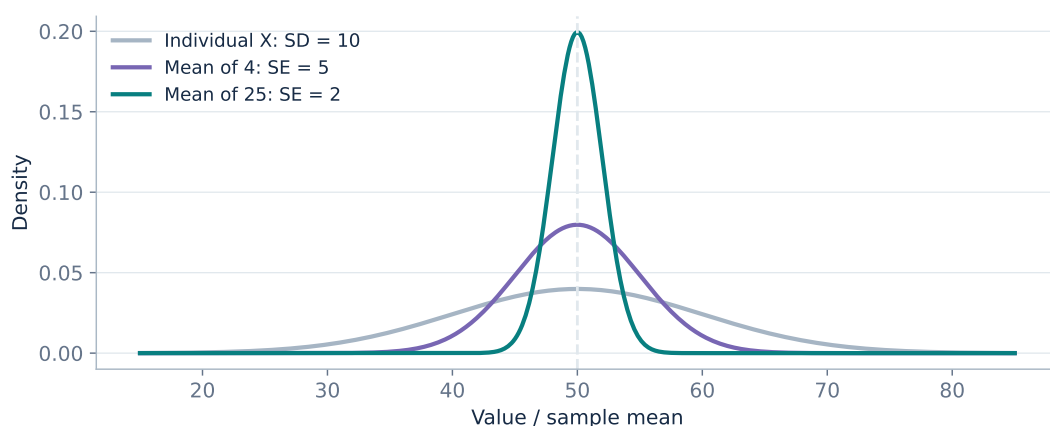
These moment formulas hold for independent observations from the same population with finite variance; normality is not required.

The expectation follows by adding n means and dividing by n . For variance, independence allows us to add the n variances, while dividing the sum by n divides its variance by n^2 . Thus $n\sigma^2/n^2 = \sigma^2/n$. Sample means have the same centre as individual observations but less variation.

KEY FORMULA Standard error of a mean

$$\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Standard error means the standard deviation of a statistic. It measures the spread of sample means, rather than the spread of individual observations.



For a normal population with mean 50 and standard deviation 10, sample means have standard deviation 5 when $n = 4$, and 2 when $n = 25$. All three curves are centred at 50.

4. Exact normal sampling distributions

When the population is normal, any independent sample mean is exactly normal. There is no requirement that the sample size exceed a particular threshold. In the notation $N(\mu, \sigma^2)$, the second parameter is the variance, not the standard deviation.

KEY FORMULA Normal population

$$X_i \sim N(\mu, \sigma^2) \implies \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

The observations must be independent and have this same normal distribution.

WORKED EXAMPLE A sample mean above a threshold

Individual measurements follow $N(50, 100)$. For a random independent sample of 25, find the distribution of $\bar{X}$ and $P(\bar{X} > 54)$.

1. Divide the population variance by the sample size.

$$\bar{X} \sim N(50, 100/25) = N(50, 4), \quad \text{SE} = 2$$

2. Standardise using the standard error.

$$P(\bar{X} > 54) = P\left(Z > \frac{54 - 50}{2}\right) = P(Z > 2) \approx 0.0228$$

Answer. $\bar{X} \sim N(50, 4)$ and the probability is approximately 0.0228. Using the individual standard deviation 10 would answer a different question.

5. A normal approximation for other populations

The central limit theorem explains why a sample mean can be approximately normal even when individual observations are not. For independent observations from the same population with finite variance, the approximation improves as the sample size increases. Strong skewness or heavy tails may require much larger samples than a nearly symmetric population.

KEY FORMULA Suitable large-sample approximation

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Use this only when a normal approximation is suitable. There is no universal rule that $n = 30$ guarantees accuracy for every population.

WORKED EXAMPLE A skewed waiting-time population

Waiting times have mean 24 minutes and standard deviation 10 minutes. A normal approximation is suitable for the mean of 100 independent random observations. Estimate $P(\bar{X} < 22)$.

1. The variance formula still applies although the population is skewed.

$$\text{Var}(\bar{X}) = \frac{100}{100} = 1, \quad \text{SE} = 1$$

2. Use the stated normal approximation.

$$P(\bar{X} < 22) \approx P\left(Z < \frac{22 - 24}{1}\right) = P(Z < -2) \approx 0.0228$$

Answer. The estimated probability is 0.0228. The sampling model is approximate, so retain an approximation sign.

6. Estimating unknown population values

An estimator is unbiased when its expected value equals the parameter it estimates. Since $E(\bar{X}) = \mu$, the sample mean is an unbiased estimator of the population mean. This is a statement about repeated sampling; an individual estimate may still be too high or too low.

KEY FORMULA Unbiased estimate of population variance

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{\sum x_i^2 - (\sum x_i)^2/n}{n - 1}$$

Use $n > 1$ and an independent sample from a population with finite variance. The variance describing the listed data themselves instead uses denominator n .

The deviations from the sample mean sum to zero. Once $n - 1$ deviations are known, the last is determined. The denominator $n - 1$ also corrects the downward bias caused by measuring deviations from the estimated mean. The square root s estimates the population standard deviation; unbiasedness of s^2 does not mean s is itself unbiased.

WORKED EXAMPLE Estimating from raw data

An independent random sample gives 4, 6, 8, 10. Estimate the population mean and variance, then calculate the estimated standard error of the mean.

1. Find the observed mean and squared deviations.

$$\bar{x} = 7, \quad \sum (x_i - \bar{x})^2 = 9 + 1 + 1 + 9 = 20$$

2. Use the sample-variance denominator.

$$s^2 = \frac{20}{3}, \quad s = \sqrt{\frac{20}{3}}$$

3. Divide the estimated standard deviation by the square root of n .

$$\widehat{SE} = \frac{s}{\sqrt{4}} = \sqrt{\frac{5}{3}} \approx 1.291$$

Answer. The mean estimate is 7, the variance estimate is $20/3$, and the estimated standard error is approximately 1.291.

7. Choose the appropriate standard error

If the population standard deviation is known, use $\sigma/\sqrt{n}$. If it is unknown, calculate s from the sample and use the estimated standard error $s/\sqrt{n}$. Estimating the standard error does not, by itself, justify an exact normal confidence interval or test.

COMMON MISTAKE An estimate does not become a known parameter

Replacing σ by s introduces additional uncertainty. A suitable large sample can support approximate normal inference. For a small sample from a normal population with unknown variance, the t methods in Further Statistics are needed.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

A normal population has mean 30 and variance 16. For 16 independent random observations, find the distribution of $\bar{X}$ and $P(29.5 < \bar{X} < 31)$.

1. Divide the variance by n and take its square root.

$$\bar{X} \sim N(30, 16/16) = N(30, 1), \quad SE = 1$$

2. Standardise both endpoints.

$$P(-0.5 < Z < 1) = \Phi(1) - \Phi(-0.5) \approx 0.8413 - 0.3085$$

Answer. $\bar{X} \sim N(30, 1)$ and the probability is approximately 0.5328.

PRACTICE Question 2

For the independent random sample 3, 5, 5, 7, estimate the population mean, variance and standard error of the sample mean.

1. Calculate the mean and sum of squared deviations.

$$\bar{x} = 5, \quad \sum (x_i - \bar{x})^2 = 4 + 0 + 0 + 4 = 8$$

2. Use n minus one for variance and the square root of n for standard error.

$$s^2 = \frac{8}{3}, \quad \widehat{SE} = \frac{\sqrt{8/3}}{2} = \sqrt{\frac{2}{3}}$$

Answer. Mean estimate = 5, variance estimate = 8/3, estimated standard error ≈ 0.8165 .

PRACTICE Question 3

An independent random sample has $n = 10$, $\sum x_i = 52$ and $\sum x_i^2 = 300$. Find $\bar{x}$, s^2 and the estimated standard error.

1. Use the summarised-data formula.

$$\bar{x} = \frac{52}{10} = 5.2, \quad s^2 = \frac{300 - 52^2/10}{9} = \frac{148}{45}$$

2. Take the square root after dividing the estimated variance by n.

$$\widehat{SE} = \sqrt{\frac{148/45}{10}} \approx 0.5735$$

Answer. $\bar{x} = 5.2$, $s^2 = 148/45 \approx 3.289$, and estimated standard error ≈ 0.5735 .

PRACTICE Question 4

(a) How must sample size change to halve the standard error when σ stays fixed? (b) Does this change guarantee that a sample mean from every population is normally distributed?

1. Set the new standard error equal to half the old one.

$$\frac{\sigma}{\sqrt{n_{\text{new}}}} = \frac{1}{2} \frac{\sigma}{\sqrt{n_{\text{old}}}} \Rightarrow n_{\text{new}} = 4n_{\text{old}}$$

2. Normality is exact for independent samples from a normal population. For another population, a suitable CLT approximation needs justification.

Answer. (a) Multiply sample size by 4. (b) No: a larger sample does not provide a universal normality guarantee.

Ready to move on?

- ☐ I can tell an individual observation from a sample mean.
- ☐ I can distinguish the random mean from its observed value.
- ☐ I can calculate variance and standard error on the correct scale.
- ☐ I can state when a normal sampling model is exact or approximate.
- ☐ I can calculate unbiased variance estimates from raw and summarised data.

Next: Hypothesis testing foundations. Use a sampling distribution to assess evidence about a population mean.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 6.4: Sampling and estimation](#) · [Cambridge 9231 prior-learning guidance](#)

LESSON 11 9709 TO 9231

Hypothesis testing foundations

Use a sample mean to assess evidence about a population claim and state a careful conclusion.

By the end of this lesson

- Write null and alternative hypotheses about a population mean.
- Choose one or two tails from the research question.
- Carry out exact known-variance and suitable approximate normal mean tests.
- Interpret significance, critical regions and conclusions correctly.

Before you begin: Calculate the standard error of a sample mean; Use normal probabilities and standard normal critical values; Distinguish population parameters from sample estimates.

1. Compare the observation with a population claim

A hypothesis test starts with a claim about a population parameter. We temporarily assume that claim is true, then ask whether the observed sample result would be unusually extreme under that assumption. An extreme result gives evidence against the claim; an ordinary result provides insufficient evidence to reject it.

The null hypothesis, H_0 , supplies the reference value. For a population mean we write $H_0 : \mu = \mu_0$, where μ_0 is the claimed value. The alternative hypothesis, H_1 , describes the departure being investigated. Hypotheses concern μ , not the observed sample mean $\bar{x}$: the sample mean is evidence about the population.

METHOD TIP Choose the direction first

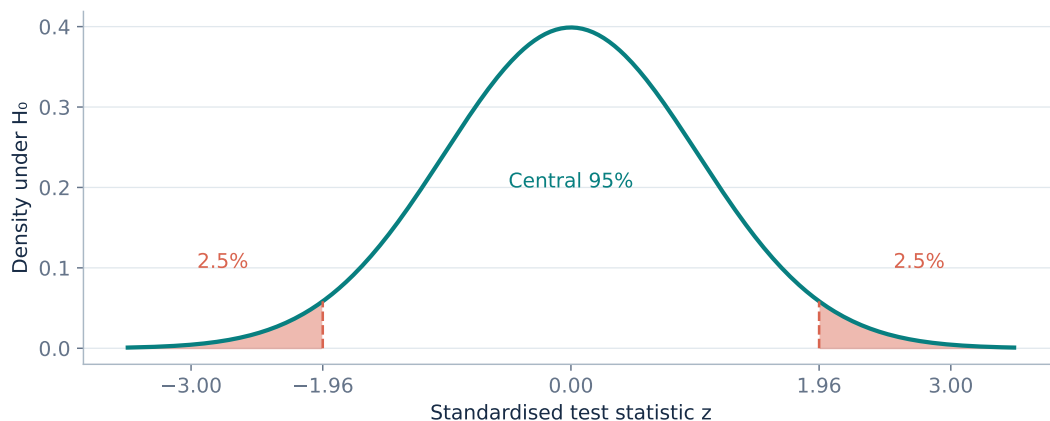
Select H_1 from the investigation before examining the sample result. Choosing the favourable tail after seeing the data changes the intended testing procedure.

2. Significance levels and critical regions

The critical region is the set of sufficiently extreme statistic values for which we reject H_0 . A significance level of 5% assigns total probability 0.05 to this region under the null model. For an approximate test, this probability is approximately 0.05. It is not the probability that H_0 is true.

Alternative	Tail or tails	5% normal critical region
$H_1 : \mu > \mu_0$	Upper	$z > 1.645$
$H_1 : \mu < \mu_0$	Lower	$z < -1.645$
$H_1 : \mu \neq \mu_0$	Two	$z < -1.96$ or $z > 1.96$

A two-tailed test looks for a departure in either direction. Split the significance level equally: at 5%, each tail contains 2.5%. The displayed cutoffs are rounded standard normal quantiles; an equality convention does not affect probabilities in a continuous model.



For a two-tailed normal test at 5%, the rejection regions lie beyond approximately -1.96 and 1.96 , with probability 0.025 in each tail under H_0 .

3. Standardise the sample mean under the null

KEY FORMULA Known population standard deviation

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}, \quad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Z is the random statistic before sampling; z is its observed value. Under H_0 , $Z \sim N(0, 1)$ for an independent random sample from a normal population with known σ .

The denominator is the standard error of the sample mean, not the standard deviation of individual observations. A suitable large sample from a non-normal population with finite variance can justify an approximate normal test through the central limit theorem. There is no universal sample-size threshold that makes this approximation reliable.

- Check random sampling and independent observations.
- Identify whether the population standard deviation is known or estimated.
- State whether normality of the test statistic is exact or an approximation.

4. Worked tests with known variance

WORKED EXAMPLE Is a filling machine underfilling?

Bag masses are normal with known standard deviation 8 g. An independent random sample of 25 bags has mean 496.8 g. Test at 5% whether the population mean is less than 500 g.

1. Define μ as the population mean bag mass and choose a lower-tailed alternative.

$$H_0 : \mu = 500, \quad H_1 : \mu < 500$$

2. Calculate the standard error and observed statistic.

$$SE = \frac{8}{\sqrt{25}} = 1.6, \quad z = \frac{496.8 - 500}{1.6} = -2$$

3. The critical region is $z < -1.645$. Since $-2 < -1.645$, reject H_0 .

Answer. There is sufficient evidence at the 5% significance level that the population mean bag mass is below 500 g.

WORKED EXAMPLE Has a mean fill volume changed?

Bottle volumes are normal with known standard deviation 12 ml. An independent random sample of 36 bottles has mean 254 ml. Test at 5% whether the population mean differs from 250 ml.

1. A change in either direction requires two tails.

$$H_0 : \mu = 250, \quad H_1 : \mu \neq 250$$

2. Calculate the statistic using the standard error.

$$SE = \frac{12}{\sqrt{36}} = 2, \quad z = \frac{254 - 250}{2} = 2$$

3. At 5%, reject for $|z| > 1.96$. Here $2 > 1.96$.

Answer. Reject H_0 . There is sufficient evidence at 5% that the population mean volume differs from 250 ml.

5. A suitable large sample with unknown variance

When the population variance is unknown, estimate it using s^2 with denominator $n - 1$. For a suitable large sample, replacing σ by s gives an approximate normal test. This is not an exact known-variance test. Small normal samples with unknown variance require the t methods taught in Further Statistics.

KEY FORMULA Approximate statistic using an estimated standard error

$$z_{\text{approx}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Use an independent random sample and a justified large-sample normal approximation. The sample standard deviation s estimates σ .

WORKED EXAMPLE A large sample of item masses

An independent random sample of 100 items gives $\bar{x} = 51.8$ g and $s = 10$ g. A large-sample normal approximation is suitable. Test at 5% whether the population mean exceeds 50 g.

1. State an upper-tailed alternative.

$$H_0 : \mu = 50, \quad H_1 : \mu > 50$$

2. Estimate the standard error.

$$\widehat{SE} = \frac{10}{\sqrt{100}} = 1, \quad z_{\text{approx}} = \frac{51.8 - 50}{1} = 1.8$$

3. Since $1.8 > 1.645$, the result falls in the approximate upper-tailed 5% critical region.

Answer. Reject H_0 . The approximate test gives sufficient evidence at 5% that the population mean mass exceeds 50 g.

6. What the conclusion does and does not say

If the statistic is in the critical region, reject H_0 and state the evidence for H_1 in context. Otherwise, do not reject H_0 : there is insufficient evidence for the proposed departure at that significance level. A non-significant result does not prove that the null value is correct.

In the bottle example, an observed mean of 253 ml would give $z = (253 - 250)/2 = 1.5$. This lies outside the two-tailed critical region. The conclusion would be insufficient evidence that the population mean differs from 250 ml, even though the sample mean itself differs from 250.

COMMON MISTAKE Statistical evidence is not practical size

A decision describes evidence under a model. It does not, by itself, tell you whether the departure is large enough to matter in practice.

7. Optional: expressing the evidence with a p-value

A p-value is the probability, assuming H_0 and the model assumptions, of a statistic at least as extreme as the observed value in the direction or directions specified by H_1 . Reject when the p-value is at most the significance level.

KEY FORMULA Normal p-values

$$p_{\text{upper}} = 1 - \Phi(z), \quad p_{\text{lower}} = \Phi(z), \quad p_{\text{two}} = 2[1 - \Phi(|z|)]$$

Here $\Phi(a) = P(Z \leq a)$ for $Z \sim N(0, 1)$. These formulas use standard normal symmetry.

For the bottle test, $p = 2P(Z > 2) \approx 0.0455 < 0.05$, agreeing with the critical-region decision. The p-value is not the probability that the null hypothesis is true.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

Measurements are normal with known $\sigma = 15$. An independent random sample of 36 gives $\bar{x} = 125$. Test at 1% whether the population mean exceeds 120. Use the upper-tail critical value 2.326.

1. State the hypotheses and calculate the standard error.

$$H_0 : \mu = 120, \quad H_1 : \mu > 120, \quad SE = 15/\sqrt{36} = 2.5$$

2. Calculate and compare the statistic.

$$z = \frac{125 - 120}{2.5} = 2 < 2.326$$

Answer. Do not reject H_0 . There is insufficient evidence at 1% that the population mean exceeds 120.

PRACTICE Question 2

Measurements are normal with known $\sigma = 6$. An independent random sample of 16 gives $\bar{x} = 36.7$. Test at 5% whether the population mean differs from 40.

1. A difference requires two tails.

$$H_0 : \mu = 40, \quad H_1 : \mu \neq 40$$

2. Calculate the statistic.

$$SE = 6/\sqrt{16} = 1.5, \quad z = \frac{36.7 - 40}{1.5} = -2.2$$

3. Since $-2.2 < -1.96$, the result is in the critical region.

Answer. Reject H_0 . There is sufficient evidence at 5% that the population mean differs from 40.

PRACTICE Question 3

For a normal population with known $\sigma = 10$, take an independent random sample of 25. Express the 5% critical region for testing $H_0 : \mu = 100$ against $H_1 : \mu > 100$ in terms of the observed mean $\bar{x}$.

1. The standard error is two, and the test is upper-tailed.

$$\frac{\bar{x} - 100}{2} > 1.645$$

2. Rearrange the inequality.

$$\bar{x} > 100 + 2(1.645) = 103.29$$

Answer. Reject when $\bar{x} > 103.29$, using the rounded critical value.

PRACTICE Question 4

An independent random sample of 64 gives $\bar{x} = 19.4$ and $s = 4$. A large-sample normal approximation is suitable. Test at 5% whether the population mean is below 20.

1. State the hypotheses and estimate the standard error.

$$H_0 : \mu = 20, \quad H_1 : \mu < 20, \quad \widehat{SE} = 4/\sqrt{64} = 0.5$$

2. Compare the approximate statistic with the lower-tail threshold.

$$z_{\text{approx}} = \frac{19.4 - 20}{0.5} = -1.2 > -1.645$$

Answer. Do not reject H_0 . This approximate test provides insufficient evidence at 5% that the population mean is below 20.

Ready to move on?

- ☐ I can write hypotheses using the population mean.
- ☐ I can choose the correct tail or tails before calculating.
- ☐ I can state the conditions for the statistic I use.
- ☐ I can distinguish a known standard error from an estimated one.
- ☐ I can give a contextual conclusion without claiming to prove the null hypothesis.

Next: Confidence intervals foundations. Estimate a population mean with an interval that expresses sampling uncertainty.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

Cambridge 9709 syllabus, section 6.5: Hypothesis tests · Cambridge 9231 prior-learning guidance

LESSON 12 9709 TO 9231

Confidence intervals foundations

Estimate a population mean while showing how much uncertainty remains after sampling.

By the end of this lesson

- Calculate and interpret a normal confidence interval for a population mean.
- Distinguish exact known-variance intervals from approximate intervals.
- Explain how confidence, variability and sample size affect width.
- Find a minimum sample size for a target margin of error.

Before you begin: Distinguish a population mean from an observed sample mean; Calculate known and estimated standard errors; Use central areas of the standard normal distribution.

1. A sample estimate comes with uncertainty

The observed sample mean $\bar{x}$ is a point estimate of the population mean μ . Another independent sample will usually give a different mean. A confidence interval combines the point estimate with a margin of error to express this sampling uncertainty.

KEY FORMULA Structure of an interval

$$\text{estimate} \pm \text{margin of error}$$

The two endpoints are the estimate minus the margin and the estimate plus the margin. The full width is twice the margin.

The population mean is fixed but unknown. Before sampling, $\bar{X}$ and the resulting interval are random because they depend on the sample. Once the data are observed, $\bar{x}$ and the interval endpoints are fixed numbers. This distinction explains the repeated-sampling meaning of confidence.

2. A normal interval with known population variance

KEY FORMULA Confidence interval for a population mean

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Use a random independent sample and a known population standard deviation σ . The population must be normal, or a suitable large-sample normal approximation must apply.

Here n is the sample size, $\sigma/\sqrt{n}$ is the standard error of the sample mean, and z^* is a standard normal critical value chosen for the confidence level. The margin of error is z^* times the standard error.

For a normal population with known variance, the normal method is exact for any sample size, apart from rounding critical values. For a non-normal population with finite variance, a suitable central limit theorem approximation gives approximate coverage. A sample size of 30 is not a universal guarantee.

KEY FORMULA Why the 95% method works

$$P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

Under the normal sampling model, this follows by rearranging $-1.96 \leq (\bar{X} - \mu)/(\sigma/\sqrt{n}) \leq 1.96$. The approximation here reflects the rounded critical value.

3. Choose the central area

A two-sided confidence interval leaves equal probabilities in the two tails. For 95% confidence, the central area is 0.95 and each tail has area 0.025. A larger confidence level requires a larger critical value.

Confidence level	Probability in each tail	z^*
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.576

METHOD TIP Match the interval to the confidence level

Do not use 1.645 for a two-sided 95% interval: it leaves 5% in each tail and therefore gives 90% central coverage.

4. Calculate and interpret an interval

WORKED EXAMPLE The population mean fill volume

Fill volumes are normally distributed with known standard deviation 12 ml. An independent random sample of 64 containers has mean 502 ml. Calculate a 95% confidence interval for the population mean.

1. Calculate the standard error of the sample mean.

$$SE = \frac{12}{\sqrt{64}} = 1.5 \text{ ml}$$

2. Multiply by the 95% critical value.

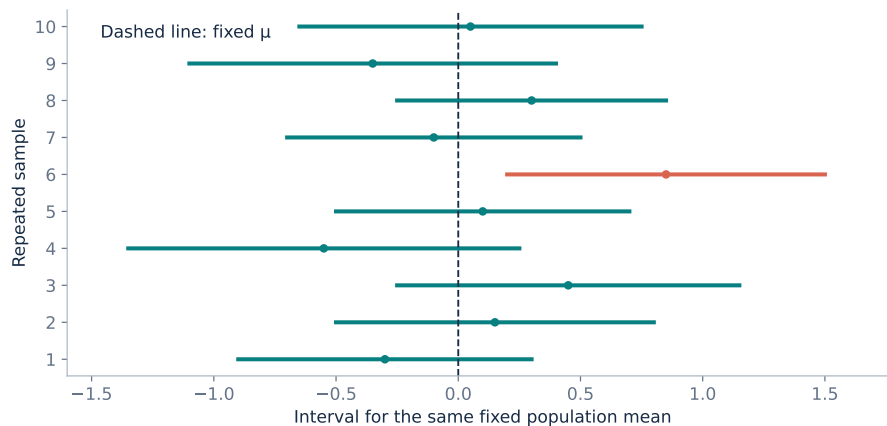
$$\text{margin} = 1.96(1.5) = 2.94 \text{ ml}$$

3. Subtract and add the margin to the observed mean.

$$502 \pm 2.94 = (499.06, 504.94) \text{ ml}$$

Answer. We are 95% confident that the population mean fill volume lies between 499.06 ml and 504.94 ml.

The confidence statement refers to the procedure: if we repeatedly took samples in the same way and constructed an interval each time, about 95% of those intervals would contain the fixed population mean. Some intervals miss it. We cannot tell from one sample alone whether our particular interval is one of them.



Illustrative intervals from repeated samples surround different sample means. The population mean is fixed. The diagram explains coverage; its small collection of intervals is not intended to contain exactly 95% successful intervals.

5. Interval width and sample size

KEY FORMULA Width and margin

$$\text{width} = \frac{2z^*\sigma}{\sqrt{n}}, \quad \text{margin} = \frac{z^*\sigma}{\sqrt{n}}$$

These expressions use a known population standard deviation and the normal interval above.

- Higher confidence increases the critical value, producing a wider interval.
- Greater population variability increases the width.
- A larger sample reduces the width through the square root of n .
- Multiplying sample size by four halves the margin when confidence and population variability stay fixed.

To plan a sample, rearrange the margin formula and round the required sample size upwards. Rounding to the nearest integer can fail to meet the target. If the question specifies the full width, halve it first to obtain the margin.

WORKED EXAMPLE Planning a sufficiently precise estimate

A population has known standard deviation 10. Find the smallest sample size giving a 95% normal confidence interval with margin at most 2. Assume independent random sampling and an appropriate normal sampling model.

1. Write the required bound on the margin.

$$\frac{1.96(10)}{\sqrt{n}} \leq 2$$

2. Rearrange, using positive quantities.

$$n \geq \left(\frac{1.96(10)}{2} \right)^2 = 96.04$$

3. The sample size must be a whole number at least 96.04, so round upwards.

Answer. The smallest suitable sample size is $n = 97$.

6. An approximate interval using an estimated variance

If the population standard deviation is unknown, estimate it using the sample standard deviation s , where s^2 uses denominator $n - 1$. For a suitable large sample, the estimated standard error $s/\sqrt{n}$ can be used in an approximate normal interval. The additional estimation means the normal coverage is no longer exact.

KEY FORMULA Large-sample normal interval

$$\bar{x} \pm z^* \frac{s}{\sqrt{n}}$$

Use independent random observations and a justified large-sample normal approximation. A small normal sample with unknown variance requires a t interval, studied in Further Statistics.

WORKED EXAMPLE A mean with unknown population variance

An independent random sample of 100 observations gives $\bar{x} = 80$ and $s = 15$. A large-sample normal approximation is suitable. Find an approximate 95% confidence interval for μ .

1. Estimate the standard error.

$$\widehat{SE} = \frac{15}{\sqrt{100}} = 1.5$$

2. Calculate the approximate margin and interval.

$$80 \pm 1.96(1.5) = 80 \pm 2.94 = (77.06, 82.94)$$

Answer. An approximate 95% confidence interval for the population mean is (77.06, 82.94).

7. Avoid three interpretation mistakes

COMMON MISTAKE Confusing a mean with individual observations

A 95% confidence interval for μ does not contain 95% of individual observations. Its purpose is to estimate the population mean; an interval for individual values addresses a different question.

COMMON MISTAKE Treating the fixed mean as random

In this repeated-sampling framework, do not describe a calculated interval as assigning probability 0.95 to the fixed parameter. Explain that the interval method captures the true mean in about 95% of repeated samples.

COMMON MISTAKE Assuming a narrow interval removes every problem

The margin reflects sampling variability under the model. It does not automatically account for biased sampling, measurement problems or failures of independence.

Check your understanding

Try each question before reading its worked solution. Keep unrounded values during calculations.

PRACTICE Question 1

A normal population has known standard deviation 9. An independent random sample of 36 has mean 42. Find a 90% confidence interval for the population mean.

1. Use the 90% critical value and calculate the standard error.

$$SE = 9/\sqrt{36} = 1.5, \quad z^* = 1.645$$

2. Calculate the margin and endpoints.

$$\text{margin} = 1.645(1.5) = 2.4675, \quad 42 \pm 2.4675 = (39.5325, 44.4675)$$

Answer. The 90% confidence interval is approximately (39.53, 44.47).

PRACTICE Question 2

A population has known standard deviation 6. Find the smallest sample size for a 99% normal confidence interval with margin at most 1. Assume independent random sampling and a suitable normal sampling model.

1. Write the margin inequality.

$$\frac{2.576(6)}{\sqrt{n}} \leq 1$$

2. Square and round upwards.

$$n \geq (2.576 \times 6)^2 = 238.887936$$

Answer. The smallest sample size is 239.

PRACTICE Question 3

A normal population has known standard deviation 8. An independent random sample of 64 has mean 25. Find a 95% confidence interval and the sample size needed to halve its margin at the same confidence level.

1. The standard error is one.

$$SE = 8/\sqrt{64} = 1, \quad 25 \pm 1.96 = (23.04, 26.96)$$

2. Halving the margin requires four times as many observations.

$$n_{\text{new}} = 4(64) = 256$$

Answer. The interval is (23.04, 26.96). A sample size of 256 halves the margin.

PRACTICE Question 4

An independent random sample has $n = 64$, $\sum x_i = 1920$ and $\sum x_i^2 = 61632$. Population variance is unknown, and a large-sample normal approximation is suitable. Find an approximate 95% confidence interval for the population mean.

1. Estimate the mean and population variance from the summaries.

$$\bar{x} = 1920/64 = 30, \quad s^2 = \frac{61632 - 1920^2/64}{63} = \frac{4032}{63} = 64$$

2. Use s , rather than s squared, in the estimated standard error.

$$s = 8, \quad \widehat{SE} = 8/\sqrt{64} = 1$$

3. Form the approximate interval.

$$30 \pm 1.96 = (28.04, 31.96)$$

Answer. An approximate 95% confidence interval for μ is (28.04, 31.96).

Ready to move on?

- ☐ I can choose the critical value for the stated confidence level.
- ☐ I can use the correct known or estimated standard error.
- ☐ I can interpret confidence through repeated sampling.
- ☐ I can distinguish margin of error from full interval width.
- ☐ I can rearrange a margin requirement and round the sample size upwards.

Next: Further Statistics 9231. Continue to t intervals and intervals for differences between population means, building on the same estimate-and-standard-error structure.

Course alignment: selected 9709 Statistics 1 and 2 foundations for Further Mathematics 9231. Numerical examples are original teaching examples.

[Cambridge 9709 syllabus, section 6.4: Sampling and estimation](#) · [Cambridge 9231 prior-learning guidance](#) · [Cambridge 9231 syllabus, section 4: Further Probability and Statistics](#)