

YEAR 1 STATISTICS · CHAPTER 6

Statistical Distributions

Specification: Edexcel Paper 3 §4.1 — Discrete distributions · AQA section N

KEY FACTS — THIS CHAPTER

- A **discrete random variable** X takes separate values, each with a probability. Its **probability distribution** lists them, and the probabilities always sum to 1: $\sum P(X = x) = 1$.
- In a **discrete uniform** distribution every value is equally likely: $P(X = x) = \frac{1}{n}$.
- $X \sim B(n, p)$ means a **binomial** distribution: n fixed trials, two outcomes, constant probability p of success, and independent trials.
- $P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}$, using the $\binom{n}{r}$ from Chapter 7 of Pure.
- **Cumulative** probabilities: $P(X \leq r)$ comes straight from a calculator. Then $P(X > r) = 1 - P(X \leq r)$ and $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a - 1)$.
- Check the four conditions before modelling anything as binomial.

A **random variable** is a quantity whose value depends on chance: the score on a die, the number of faulty items in a box. This chapter gives the language for describing all their probabilities at once, then studies the one distribution the course uses most.

1. Discrete random variables

A **discrete** random variable takes distinct, separate values. Capital letters name the variable and small letters its values, so $P(X = 3)$ means “the probability that X takes the value 3”.

KEY RESULT

The **probability distribution** of X gives $P(X = x)$ for every value x . Since one of them must happen,

$$\sum_x P(X = x) = 1$$

WORKED EXAMPLE 6.1 • FINDING AN UNKNOWN PROBABILITY

The discrete random variable X has the distribution below. Find k , and then $P(X \geq 3)$.

x	1	2	3	4
$P(X = x)$	k	$2k$	$3k$	$4k$

The probabilities sum to 1:

$$k + 2k + 3k + 4k = 10k = 1 \quad \Rightarrow \quad k = 0.1$$

Then $P(X \geq 3) = P(X = 3) + P(X = 4) = 0.3 + 0.4 = 0.7$.

A **discrete uniform** distribution is the simplest case: every value equally likely. A fair six-sided die gives $P(X = x) = \frac{1}{6}$ for $x = 1, 2, \dots, 6$.

2. The binomial distribution

Many situations count **successes** in a fixed number of attempts: heads in 20 tosses, faulty bulbs in a box of 50. All of them share the same four features.

THE BINOMIAL CONDITIONS

$X \sim B(n, p)$ needs:

1. a **fixed** number of trials, n ;
2. **two** outcomes per trial — success or failure;
3. a **constant** probability of success, p ;
4. **independent** trials.

Then X counts the successes, and takes values $0, 1, 2, \dots, n$.

COMMON MISTAKE

Sampling **without replacement** breaks conditions 3 and 4: taking one red ball out changes the chance for the next. A binomial model is only reasonable there if the population is very large compared with the sample, so the probability barely moves.

THE BINOMIAL FORMULA

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

for $r = 0, 1, \dots, n$.

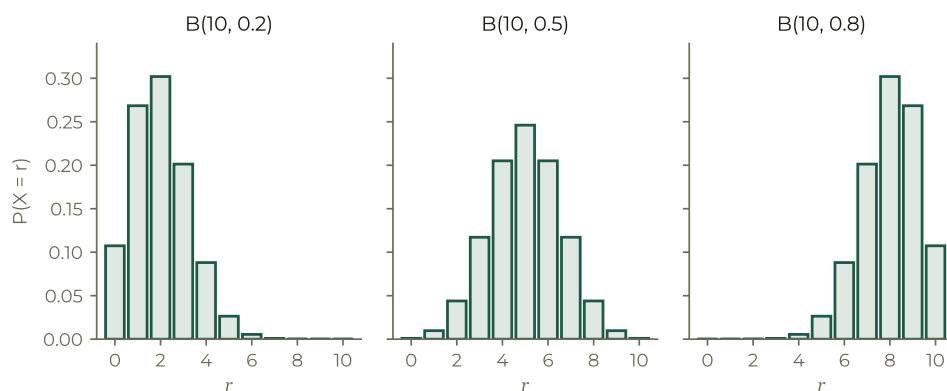
The $\binom{n}{r}$ counts the orders in which r successes can happen — exactly the coefficient from Pascal's triangle in Chapter 7 of Pure. The p^r and $(1 - p)^{n-r}$ give the probability of any one such order.

WORKED EXAMPLE 6.2 · USING THE FORMULA

A biased coin lands heads with probability 0.3. It is tossed 10 times. Find the probability of exactly 4 heads.

Let X be the number of heads, so $X \sim B(10, 0.3)$:

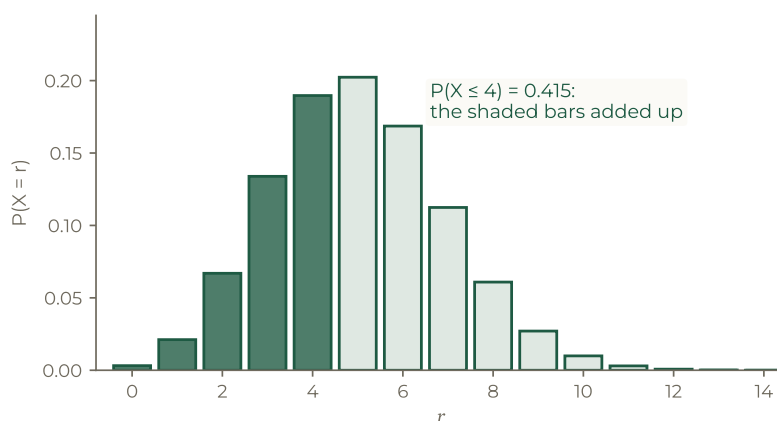
$$\begin{aligned} P(X = 4) &= \binom{10}{4} (0.3)^4 (0.7)^6 \\ &= 210 \times 0.0081 \times 0.117649 = 0.200 \text{ (3 s.f.)} \end{aligned}$$



The same n , three different p . The peak shifts right as p grows, and the shape is symmetric only when $p = 0.5$.

3. Cumulative probabilities

Adding many separate terms is slow, so questions use **cumulative** probabilities $P(X \leq r)$ — the total of all the bars up to r .



$P(X \leq 4)$ is the total height of the shaded bars: 0.415 for $B(20, 0.25)$.

KEY RESULT

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

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a - 1)$$

The -1 matters: X only takes whole-number values, so “at least 7” is “more than 6”.

ON YOUR CALCULATOR

Use the **binomial probability distribution** function for $P(X = r)$ and the **binomial cumulative distribution** function for $P(X \leq r)$. Both need n , p and r . Write the probability statement down before you reach for the calculator: most lost marks here are translation errors, not arithmetic.

WORKED EXAMPLE 6.3 · CUMULATIVE PROBABILITIES

$X \sim B(20, 0.25)$. Find (a) $P(X \leq 4)$, (b) $P(X \geq 7)$, (c) $P(4 \leq X \leq 8)$.

(a) Straight from the cumulative function: $P(X \leq 4) = 0.4148 \approx 0.415$.

(b) “At least 7” is “more than 6”:

$$P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.7858 = 0.214 \text{ (3 s.f.)}$$

(c) Subtract everything below 4, which is everything up to 3:

$$\begin{aligned} P(4 \leq X \leq 8) &= P(X \leq 8) - P(X \leq 3) \\ &= 0.9591 - 0.2252 = 0.734 \text{ (3 s.f.)} \end{aligned}$$

COMMON MISTAKE

$P(X \geq 7) \neq 1 - P(X \leq 7)$. That subtracts the $r = 7$ bar as well. Sketch the bars if you're unsure which end points belong in the range.

4. Choosing the model

Exam questions often ask you to justify a binomial model, or to criticise one.

WORKED EXAMPLE 6.4 · IS A BINOMIAL MODEL REASONABLE?

A teacher says the number of students, out of 30 in a class, who are late on a given morning can be modelled as $B(30, 0.1)$. Comment on this model.

There are a fixed number of students (30), and each is either late or not, so conditions 1 and 2 hold.

The model assumes every student has the same probability of being late (0.1), and that they are independent. Neither is likely: some students travel further and are late more often, and students who share a bus will be late **together**, so one being late makes another more likely.

IN CONTEXT

The binomial model is a reasonable first approximation, but it will understate the chance of several students being late at once, because lateness tends to come in clusters.

Practice

Twelve questions in three tiers. Try each one properly before opening the solution. Give probabilities to 3 significant figures unless they are exact, and always define your variable — “let X be the number of...” — before using a binomial model.

Warm-up**Question 1**

[3 marks]

The discrete random variable X has the distribution shown.

x	1	2	3
$P(X = x)$	k	$2k$	$3k$

- (a) Find k .
- (b) Find $P(X \leq 2)$.

SOLUTION

(a) $k + 2k + 3k = 6k = 1$, so $k = \frac{1}{6}$.

(b) $P(X \leq 2) = \frac{1}{6} + \frac{2}{6} = \frac{1}{2}$.

Question 2

[4 marks]

$X \sim B(8, 0.2)$. Find

- (a) $P(X = 2)$;
- (b) $P(X = 0)$.

SOLUTION

(a) $P(X = 2) = \binom{8}{2} (0.2)^2 (0.8)^6 = 28 \times 0.04 \times 0.262144 = 0.294$ (3 s.f.).

(b) $P(X = 0) = (0.8)^8 = 0.168$ (3 s.f.).

Question 3

[3 marks]

A fair coin is tossed 12 times. Let X be the number of heads.

- (a) State the distribution of X .
- (b) Find $P(X = 6)$ and $P(X \leq 3)$.

SOLUTION

(a) $X \sim B(12, 0.5)$.

(b) $P(X = 6) = 0.226$ and $P(X \leq 3) = 0.0730$ (both 3 s.f.).

Question 4

[4 marks]

State the four conditions needed for a binomial model, and say which one fails when three cards are dealt from a pack and X is the number of aces.

SOLUTION

A fixed number of trials; two outcomes per trial; a constant probability of success; independent trials.

Dealing without replacement breaks the constant probability (and the independence): once an ace has gone, the chance that the next card is an ace falls from $\frac{4}{52}$ to $\frac{3}{51}$.

Standard**Question 5**

[6 marks]

A factory's components are faulty with probability 0.07, independently of each other. A sample of 50 is tested. Let X be the number of faulty components.

- (a) State the distribution of X .
- (b) Find $P(X = 3)$.
- (c) Find $P(X \leq 2)$.
- (d) Find the probability that at least 5 are faulty.

SOLUTION

(a) $X \sim B(50, 0.07)$.

(b) $P(X = 3) = 0.222$ (3 s.f.).

(c) $P(X \leq 2) = 0.311$ (3 s.f.).

(d) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.729 = 0.271$ (3 s.f.).

Question 6

[6 marks]

 $X \sim B(15, 0.3)$. Find

- (a) $P(X = 5)$;
- (b) $P(X \leq 5)$;
- (c) $P(X > 5)$;
- (d) $P(3 \leq X \leq 7)$.

SOLUTION

- (a) 0.206 (3 s.f.).
- (b) 0.722 (3 s.f.).
- (c) $1 - P(X \leq 5) = 1 - 0.7216 = 0.278$ (3 s.f.).
- (d) $P(X \leq 7) - P(X \leq 2) = 0.9500 - 0.1268 = 0.823$ (3 s.f.).

Question 7

[5 marks]

A fair six-sided die is rolled 10 times. Let X be the number of sixes.

- (a) State the distribution of X and justify the binomial conditions.
- (b) Find the probability of exactly two sixes.
- (c) Find the probability of at least one six.

SOLUTION

- (a) $X \sim B(10, \frac{1}{6})$: 10 fixed rolls, each a six or not, the probability stays $\frac{1}{6}$ every time, and the rolls are independent.
- (b) $P(X = 2) = \binom{10}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^8 = 0.291$ (3 s.f.).
- (c) $P(X \geq 1) = 1 - P(X = 0) = 1 - \left(\frac{5}{6}\right)^{10} = 0.838$ (3 s.f.).

Question 8

[4 marks]

For each situation, state whether a binomial model is reasonable, and why.

- (a) The number of heads when a fair coin is tossed 50 times.
- (b) The number of rainy days in a week in one town.

- (c) The number of red beads when 5 beads are taken, without replacement, from a bag of 8 red and 7 blue.

SOLUTION

- (a) Reasonable: 50 fixed trials, two outcomes, constant $p = 0.5$, independent tosses.
- (b) Not reasonable: the weather on consecutive days is **not independent** — wet days cluster — and the probability of rain is not the same every day of the year.
- (c) Not reasonable: without replacement from a small bag, the probability changes sharply with each bead taken, so p is not constant.

Exam style**Question 9**

[8 marks]

A company knows that 40% of visitors to its website buy something. On a particular day, 25 visitors are chosen at random. Let X be the number who buy something.

- (a) State the distribution of X , and one assumption needed.
- (b) Find $P(X = 10)$.
- (c) Find $P(X \leq 8)$.
- (d) Find the probability that at least 12 of them buy something.
- (e) Find $P(8 \leq X \leq 12)$.

SOLUTION

- (a) $X \sim B(25, 0.4)$, assuming the visitors buy independently of each other and each has the same probability 0.4 of buying.
- (b) $P(X = 10) = 0.161$ (3 s.f.).
- (c) $P(X \leq 8) = 0.274$ (3 s.f.).
- (d) $P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.7323 = 0.268$ (3 s.f.).
- (e) $P(X \leq 12) - P(X \leq 7) = 0.8462 - 0.1536 = 0.693$ (3 s.f.).

Question 10

[6 marks]

A seed supplier claims that 90% of its seeds germinate. A gardener plants 20 seeds.

- (a) Assuming the claim is correct, find the probability that all 20 germinate.
- (b) Find the probability that at least 18 germinate.
- (c) Only 14 of the gardener's seeds germinate. Without doing a formal test, comment on the supplier's claim.

SOLUTION

Let X be the number that germinate, so $X \sim B(20, 0.9)$.

(a) $P(X = 20) = 0.9^{20} = 0.122$ (3 s.f.).

(b) $P(X \geq 18) = 1 - P(X \leq 17) = 1 - 0.3231 = 0.677$ (3 s.f.).

(c) If the claim were true, you would expect about 18 seeds to germinate, and getting as few as 14 is unlikely. This is evidence against the claim, though one gardener's results could be affected by soil or watering rather than the seeds. (Chapter 7 makes this judgement precise.)

Question 11

[6 marks]

$X \sim B(4, 0.25)$.

- (a) Write out the full probability distribution of X , giving each probability to 4 decimal places.
- (b) Verify that the probabilities sum to 1.
- (c) State the most likely value of X .

SOLUTION

(a)

r	0	1	2	3	4
$P(X = r)$	0.3164	0.4219	0.2109	0.0469	0.0039

(b) $0.3164 + 0.4219 + 0.2109 + 0.0469 + 0.0039 = 1.0000$.

(c) $X = 1$ is the most likely value, with probability 0.4219.

Question 12

[7 marks]

In a large population, 5% of people carry a particular gene. A researcher tests 30 people chosen at random. Let X be the number of carriers.

- (a) Explain why a binomial model is reasonable here, even though the sampling is without replacement.
- (b) Find the probability that none of the 30 carries the gene.
- (c) Find the probability that at least 3 carry it.
- (d) The researcher finds 6 carriers. Suggest what this might indicate.

SOLUTION

(a) The population is very large, so removing 30 people barely changes the proportion of carriers. The probability stays effectively 0.05 for each person, and the results are effectively independent.

(b) $X \sim B(30, 0.05)$, so $P(X = 0) = 0.95^{30} = 0.215$ (3 s.f.).

(c) $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.8122 = 0.188$ (3 s.f.).

(d) With $p = 0.05$ you would expect about 1 or 2 carriers in 30, so 6 is far more than expected. It suggests the true proportion in this population may be higher than 5%, or that the sample was not random — perhaps the people tested were related, or selected because of a family history.