Edexcel International A Level (IAL) Maths: Statistics 2

Your notes

Binomial Distribution

Contents

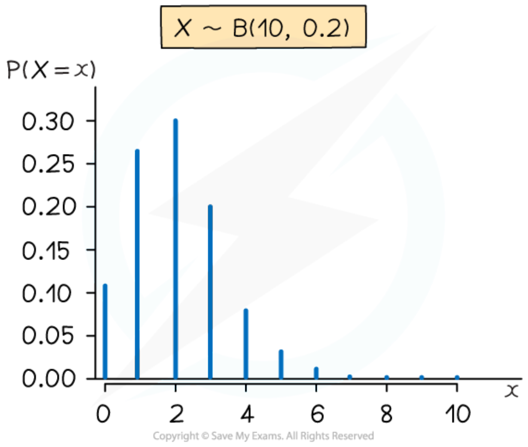
Properties of Binomial Distribution

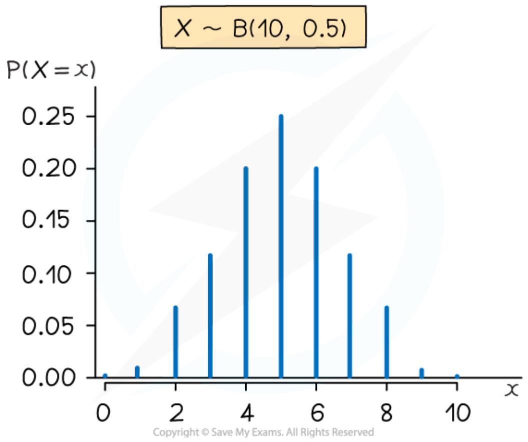
What is a binomial distribution?

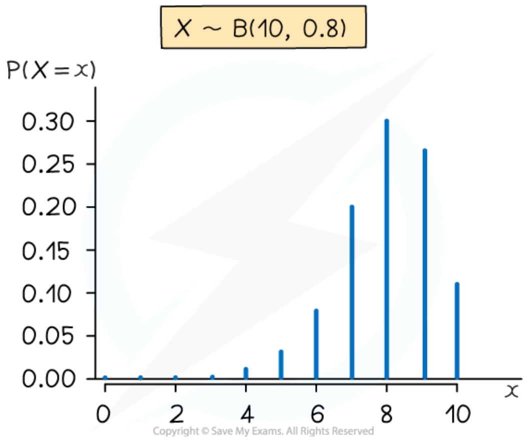
- A binomial distribution is a **discrete probability distribution**
- The **discrete random variable *X*** follows a **binomial distribution** if it **counts the number of successes** when an experiment satisfies the conditions:
  - There are a **fixed finite number of trials** (n)
  - The outcome of each trial is **independent** of the outcomes of the other trials
  - There are **exactly two outcomes** of each trial (**success or failure**)
  - The **probability of success (*p*) is constant**
- If *X* follows a binomial distribution then it is denoted $X \sim B(n,p)$
  - n is the number of trials
  - p is the probability of success
- The **probability of failure is 1–*p*** which is sometimes denoted as ***q***
- The formula for the probability of *r* successful trials is given by:
  - $P = (X = r) = \binom{n}{r} p^r (1-p)^{n-r}$ for $r = 0, 1, 2, \ldots, n$
  - This is equal to the term which includes p^r in the expansion of $(p+q)^n$ where $q = 1 - p$ (this shows the link with the Binomial Expansion)

What are the important properties of a binomial distribution?

- The **expected number (mean)** of successful trials is $\boldsymbol{np}$
- The **variance** of the number of successful trials is $\boldsymbol{np(1-p)}$
  - Square root to get the standard deviation
- The distribution can be represented visually using a vertical line graph
  - If *p* is **close to 0** then the graph has a **tail to the right**
  - If *p* is **close to 1** then the graph has a **tail to the left**
  - If *p* is **close to 0.5** then the graph is **roughly symmetrical**
  - **If *p* = 0.5** then the graph is **symmetrical**







Your notes

Modelling with Binomial Distribution

How do I set up a binomial model?

- **Identify** what a **trial** is in the scenario

- For example: rolling a dice, flipping a coin, checking hair colour
- **Identify** what the **successful outcome** is in the scenario
  - For example: rolling a 6, landing on tails, having black hair
- Make sure you **clearly state** what your **random variable** is
  - For example, let X be the number of students in a class of 30 with black hair

Your notes

What can be modelled using a binomial distribution?

- Anything that satisfies the **four conditions**
- For example, let T be the number of times a fair coin lands on tails when flipped 20 times: $T \sim \mathrm{B}\left(20, \frac{1}{2}\right)$
  - A trial is flipping a coin: There are 20 trials so $n = 20$
  - We can assume each coin flip does not affect subsequent coin flips: They are independent
  - A success is when the coin lands on tails: Two outcomes - tails or not tails (heads)
  - The coin is fair: The probability of tails is constant with $p = \frac{1}{2}$
- Sometimes it might seem like there are more than two outcomes
  - For example, let Y be the number of yellow cars that are in a car park full of 100 cars
  - Although there are more than two possible colours of cars, here the trial is whether a car is yellow so there are two outcomes (yellow or not yellow)
  - Y would still need to fulfil the other conditions in order to follow a binomial distribution
- Sometimes a sample may be taken from a population
  - For example, 30% of people in a city have blue eyes, a sample of 30 people from the city is taken and X is the number of them with blue eyes
  - As long as the population is large and the sample is random then it can be assumed that each person has a 30% chance of having blue eyes

What can not be modelled using a binomial distribution?

- Anything where the number of trials is **not fixed** or is **infinite**
  - The number of emails received in an hour
  - The number of times a coin is flipped until it lands on heads
- Anything where the outcome of **one trial affects** the outcome of the **other trials**
  - The number of caramels that a person eats when they eat 5 sweets from a bag containing 6 caramels and 4 marshmallows

- If you eat a caramel for your first sweet then there are less caramels left in the bag when you choose your second sweet

- Anything where there are **more than two outcomes** of a trial
  - A person's shoe size
  - The number a dice lands on when rolled
- Anything where the **probability of success changes**
  - The number of times that a person can swim a length of a swimming pool in under a minute when swimming 50 lengths
    - The probability of swimming a lap in under a minute will decrease as the person gets tired

Your notes

Worked Example

It is known that 8% of a large population are immune to a particular virus. Mark takes a sample of 50 people from this population. Mark uses a binomial model for the number of people in his sample that are immune to the virus

(a) State the distribution that Mark uses.

(b) State the two assumptions that Mark must make in order to use a binomial model.

Answer:

a) A trial is checking if a person is immune to the virus
A success is if the person is immune.

Let X be the number of people in the sample immune to the virus

$X \sim B(50, 0.08)$

Number of people in sample

Probability of being immune to the virus

b) Mark needs to assume that:
- each person in the population has an 8% chance of being immune
- the sample is random and the people are independent a person being immune does not affect the immunity of others

For example:
if all 50 came from the same family then they would not be independent

Your notes

Examiner Tips and Tricks

- If you are asked to criticise a binomial model always consider whether the trials are independent, this is usually the one that stops a variable from following a binomial distribution!

Your notes

Calculating Binomial Probabilities

Throughout this section we will use the random variable $X \sim \mathrm{B}(n,p)$. For binomial, the probability of a X taking a non-integer or negative value is always zero. Therefore any values mentioned in this section will be assumed to be non-negative integers.

What are the tables for the binomial cumulative distribution function?

- In your **formulae booklet** you get **tables** which list the values of for different values of *x*, *p* and *n*
  - n can be **5, 6, 7, 8, 9 10, 12, 15, 20, 25, 30, 40, 50**
  - p can be **05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5**
  - x can be different values depending on *n*
- The probabilities are rounded to **4 decimal places**
- The values of p **only go up to 0.5**
  - You can instead count the number of failures $Y \sim B(n, 1-p)$ if the probability of success is bigger than 0.5
  - Remember $X + Y = n$, which leads to identities:
    - $\mathrm{P}(X = k) = \mathrm{P}(Y = n-k)$
    - $\mathrm{P}(X \leq k) = \mathrm{P}(Y \geq n-k)$
    - $\mathrm{P}(X \geq k) = \mathrm{P}(Y \leq n-k)$

How do I calculate, P(X = x) the probability of a single value for a binomial distribution?

- You can use the formula given in the formulae booklet
  - $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
    - The number of times this can happen is calculated by the **binomial coefficient** $\binom{n}{x} = {}^nC_x = \frac{n!}{x!(n-x)!}$
- You can also use the tables for the ***Binomial Cumulative Distribution Function*** in the formulae booklet
  - $\mathrm{P}(X = k) = \mathrm{P}(X \leq k) - \mathrm{P}(X \leq k-1)$

How do I calculate, P(X ≤ x), the cumulative probabilities for a binomial distribution?

Your notes

- If x is small, you could find the probability of each possible value of x and then add them together
- Otherwise, you will have to use the tables for the ***Binomial Cumulative Distribution Function*** in the formulae booklet
- If p is bigger than 0.5 then you will have to use the number of failures $Y \sim B(n, 1-p)$
  - $P(X \le x) = P(Y \ge n-x)$

How do I find P(X ≥ x)?

- $X \ge x$: This means all values of *X* which are **at least *x***
  - These are **all** values of ***X*** **except** the ones that are **less than *x***
- $P(X \ge x) = 1 - P(X < x)$
- As *x* is an integer then $P(X < x) = P(X \le x-1)$ as the probability of *X* is zero for non-integer values for a binomial distribution
- Therefore, to calculate $P(X \ge x)$:
  - $\mathbf{P(X \ge x) = 1 - P(X \le x-1)}$
  - For example: $P(X \ge 10) = 1 - P(X \le 9)$

How do I find P(a ≤ X ≤ b)?

- $a \le X \le b$: This means all values of *X* which are **at least *a* and at most *b***
  - This is **all** the values of *X* which are **no greater than *b* except** the ones which are **less than *a***
- $P(a \le X \le b) = P(X \le b) - P(X < a)$
- As *X* is an integer then $P(X < a) = P(X \le a-1)$ as the $P(X = x) = 0$ for non-integer value of x for a binomial distribution
- Therefore to calculate $P(a \le X \le b)$:
  - $\boldsymbol{P(a \le X \le b) = P(X \le b) - P(X \le a-1)}$
  - For example: $P(4 \le X \le 9) = P(X \le 9) - P(X \le 3)$

What if an inequality does not have the equals sign (strict inequality)?

- For a binomial distribution (as it is discrete) you could rewrite all strict inequalities (< and >) as weak inequalities (≤ and ≥) by using the identities for a binomial distribution
  - $\mathbf{P(X < x) = P(X \le x-1)}$ and $\mathbf{P(X > x) = P(X \ge x+1)}$

- For example: $\mathrm{P}(X < 5) = \mathrm{P}(X \leq 4)$ and $P(X > 5) = P(X \geq 6)$
  - Though it helps to understand how they work
- It helps to think about the **range of integers** you want
- Always find the **biggest integer** that you want to **include** and the **biggest integer** that you then want to **exclude**
- For example, : $\mathrm{P}(4 < X \leq 10)$
  - You want the integers 5 to 10
  - You want the integers up to 10 excluding the integers up to 4
  - $P(X \leq 10) - P(X \leq 4)$
- For example, P(X > 6) :
  - You want the all the integers from 7 onwards
  - You want to include all integers excluding the integers up to 6
  - 1 – P(X ≤ 6)
- For example, P(X < 8) :
  - You want the integers 0 to 7
  - P(X ≤ 7)

Your notes

Worked Example

The random variable $X \sim \mathrm{B}(40, 0.35)$. Find:

(a) $\mathrm{P}(X = 10)$

(b) $\mathrm{P}(X \leq 10)$

(c) $\mathrm{P}(X \geq 10)$

(d) $\mathrm{P}(8 < X < 10)$

Answer:

Identify n and p: $n = 40 \quad p = 0.35$

a) Use the formula: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \qquad \binom{n}{x} = {}^nC_r$

$P(X = 10) = {}^{40}C_{10} \times 0.35^{10} \times 0.65^{30}$

0.0571 (3.s.f.)

Your notes

b) Use the tables with $n = 40$, $p = 0.35$ and $x = 10$

$P(X \leqslant 10) = 0.1215$

0.122 (3.s.f.)

c) $X \geqslant 10$: 10, ..., 40

Want all values except the ones up to 9

$P(X \geqslant 10) = 1 - P(X \leqslant 9)$

$= 1 - 0.0644$ Use the table

$= 0.9356$

0.936 (3.s.f.)

d) $8 < X < 15$: 9, ... , 14

Want all values up to 14 except values up to 8

$P(8 < X < 15) = P(X \leqslant 14) - P(X \leqslant 8)$

$= 0.5721 - 0.0303$

$= 0.5418$

0.542 (3.s.f.)

Examiner Tips and Tricks

- Some calculators will calculate probabilities for binomial distributions
- These are great for checking your answers once you have answered your question showing the appropriate method