

DP IB Maths: AI HL



Your notes

4.7 Binomial Distribution

Contents

- * 4.7.1 The Binomial Distribution
- * 4.7.2 Calculating Binomial Probabilities



Your notes

4.7.1 The Binomial Distribution

Properties of Binomial Distribution

What is a binomial distribution?

- A binomial distribution is a **discrete probability distribution**
- A **discrete random variable** X follows a **binomial distribution** if it **counts the number of successes** when an experiment satisfies the following 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 is constant** (p)
- 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:
 - $P(X = r) = {}^n C_r \times p^r (1 - p)^{n - r}$ for $r = 0, 1, 2, \dots, n$
 - ${}^n C_r = \frac{n!}{r!(n - r)!}$ where $n! = n \times (n - 1) \times (n - 2) \times \dots \times 3 \times 2 \times 1$
 - You will be expected to use the distribution function on your **GDC to calculate probabilities** with the binomial distribution

What are the important properties of a binomial distribution?

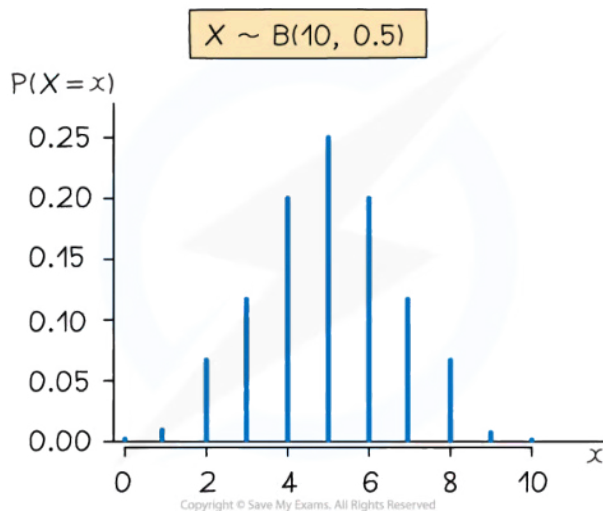
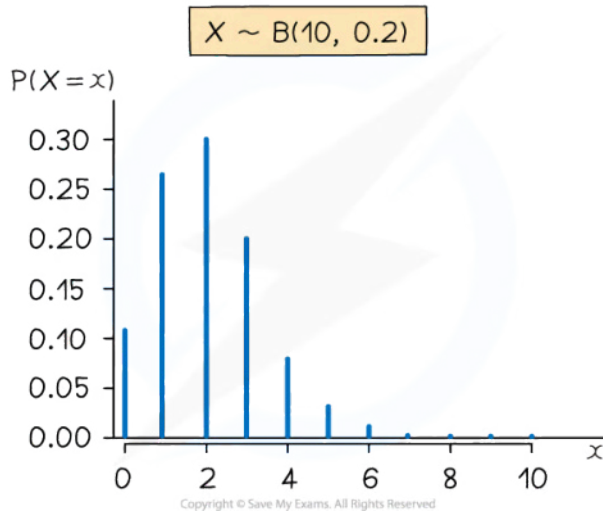
- The **expected number (mean)** of successful trials is

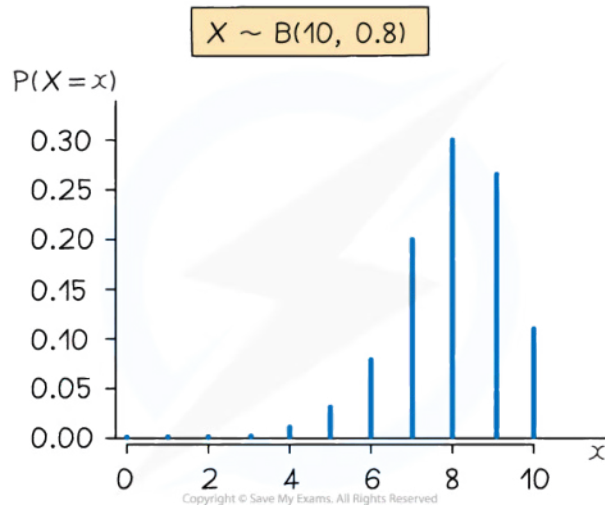
$$E(X) = np$$
 - You are given this in the **formula booklet**
- The **variance** of the number of successful trials is

$$\text{Var}(X) = np(1 - p)$$
 - You are given this in the **formula booklet**
 - 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





Your notes



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
- **Identify** the **parameters**
 - n is the number of trials and p is the probability of success in each trial
- 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

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:
 - 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 $p = 0.5$
- 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
 - The probability is **not constant**

Examiner Tip

- An exam question might involve different types of distributions so make it clear which distribution is being used for each variable



Your notes



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.

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

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

- c) Calculate the expected number of people in the sample that are immune to the virus.

Formula booklet

$$E(X) = 50 \times 0.08$$

4 people

Binomial distribution $X \sim B(n, p)$	
Mean	$E(X) = np$



Your notes



Your notes

4.7.2 Calculating Binomial Probabilities

Calculating Binomial Probabilities

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

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

- You should have a **GDC** that can calculate **binomial probabilities**
- You want to use the "**Binomial Probability Distribution**" function
 - This is sometimes shortened to BPD, Binomial PD or Binomial Pdf
- You will need to enter:
 - The 'x' value - the value of x for which you want to find $P(X = x)$
 - The 'n' value - the **number of trials**
 - The 'p' value - the **probability of success**
- Some calculators will give you the option of **listing the probabilities** for **multiple values of x at once**
- There is a formula that you can use but you are expected to be able to use the distribution function on your GDC

$$P(X = x) = {}^n C_x \times p^x (1 - p)^{n - x}$$

$${}^n C_x = \frac{n!}{x!(n - x)!}$$

How do I calculate $P(a \leq X \leq b)$: the cumulative probabilities for a binomial distribution?

- You should have a **GDC** that can calculate **cumulative binomial probabilities**
 - Most calculators will find $P(a \leq X \leq b)$
 - Some calculators can only find $P(X \leq b)$
 - The identities below will help in this case
- You should use the "**Binomial Cumulative Distribution**" function
 - This is sometimes shortened to BCD, Binomial CD or Binomial Cdf
- You will need to enter:
 - The lower value - this is the **value a**
 - This can be zero in the case $P(X \leq b)$
 - The upper value - this is the **value b**
 - This can be n in the case $P(X \geq a)$
 - The 'n' value - the **number of trials**
 - The 'p' value - the **probability of success**

How do I find probabilities if my GDC only calculates $P(X \leq x)$?



Your notes

- To calculate $P(X \leq x)$ just enter x into the cumulative distribution function
- To calculate $P(X < x)$ use:
 - $P(X < x) = P(X \leq x - 1)$ which works when X is a binomial random variable
 - $P(X < 5) = P(X \leq 4)$
- To calculate $P(X > x)$ use:
 - $P(X > x) = 1 - P(X \leq x)$ which works for any random variable X
 - $P(X > 5) = 1 - P(X \leq 5)$
- To calculate $P(X \geq x)$ use:
 - $P(X \geq x) = 1 - P(X \leq x - 1)$ which works when X is a binomial random variable
 - $P(X \geq 5) = 1 - P(X \leq 4)$
- To calculate $P(a \leq X \leq b)$ use:
 - $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a - 1)$ which works when X is a binomial random variable
 - $P(5 \leq X \leq 9) = P(X \leq 9) - P(X \leq 4)$

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** (\leq and \geq) by using the identities for a binomial distribution
 - $P(X < x) = P(X \leq x - 1)$ and $P(X > x) = P(X \geq x + 1)$
 - For example: $P(X < 5) = P(X \leq 4)$ and $P(X > 5) = P(X \geq 6)$
- It helps to think about the **range of integers** you want
 - Identify the smallest and biggest integers in the range
- If your range has no minimum or maximum then use 0 or n
 - $P(X \leq b) = P(0 \leq X \leq b)$
 - $P(X \geq a) = P(a \leq X \leq n)$
- $P(a < X \leq b) = P(a + 1 \leq X \leq b)$
 - $P(5 < X \leq 9) = P(6 \leq X \leq 9)$
- $P(a \leq X < b) = P(a \leq X \leq b - 1)$
 - $P(5 \leq X < 9) = P(5 \leq X \leq 8)$
- $P(a < X < b) = P(a + 1 \leq X \leq b - 1)$
 - $P(5 < X < 9) = P(6 \leq X \leq 8)$

Examiner Tip

- If the question is in context then write down the inequality as well as the final answer
 - This means you still might gain a mark even if you accidentally type the wrong numbers into your GDC



Your notes



Your notes

Worked example

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

i) $P(X = 10)$.

Identify n and p $n = 40$ $p = 0.35$

Use binomial probability distribution on GDC

$$P(X = 10) = 0.057056\dots$$

$$P(X = 10) = 0.057 \text{ (3sf)}$$

ii) $P(X \leq 10)$.

Identify upper and lower values

$$P(X \leq 10) = P(0 \leq X \leq 10)$$

Use binomial cumulative distribution on GDC

$$P(X \leq 10) = 0.121491\dots$$

$$P(X \leq 10) = 0.121 \text{ (3sf)}$$

iii) $P(8 < X < 15)$.

Identify upper and lower values

$$P(8 < X < 15) = P(9 \leq X \leq 14)$$

Use binomial cumulative distribution on GDC

$$P(9 \leq X \leq 14) = 0.541827\dots$$

$$P(8 < X < 15) = 0.542 \text{ (3sf)}$$