Edexcel International A Level (IAL) Maths: Statistics 2

Your notes

Working with Distributions

Contents

Your notes

Modelling with Distributions

When should I use a binomial distribution?

- A random variable that follows a **binomial distribution** is a **discrete random variable**
- A binomial distribution is used when the random variable **counts something**
  - The number of successful trials
  - The number of members of a sample that satisfy a criterion (satisfying the criteria can be seen as a successful trial)
- There are **four conditions** that X must fulfil to follow a binomial distribution
  - There is a **fixed finite number of trials (*n*)**
  - The trials are **independent**
  - There are **exactly two outcomes** of each trial (**success or failure**)
  - The **probability of success (*p*) is constant**

When should I use a Poisson distribution?

- A random variable that follows a **Poisson distribution** is a **discrete random variable**
- A Poisson distribution is used when the random variable **counts something**
  - The number of occurrences of an event in a given interval of time or space
- There are three conditions that X must fulfil to follow a Poisson distribution
  - The **mean** number of occurrences is **known** and **finite (λ)**
  - The events occur at **random**
  - The events occur **singly** and **independently**

When should I use a normal distribution?

- A random variable that follows a **normal distribution** is a **continuous random variable**
- A normal distribution is used when the random variable **measures something** and the distribution is:
  - **Symmetrical**
  - **Bell-shaped**
- A normal distribution can be used to model real-life data provided the **histogram** for this data is **roughly symmetrical and bell-shaped**
  - If the variable is normally distributed then as more data is collected the outline of the histogram should get smoother and resemble a normal distribution curve

Your notes

Can the binomial or Poisson distribution and the normal distribution be used in the same question?

- Some questions might require you to **first use the normal distribution** to find the **probability of success** and then use a **discrete distribution**
    - Remember a discrete distribution is either a binomial or Poisson distribution
- The key is to make sure you are very clear about what each **parameter/variable represents**

Worked Example

In a population of cows, the masses of the cows can be modelled using a normal distribution with mean 550 kg and standard deviation 80 kg. A farmer classifies cows as *beefy* if they weigh more than 700 kg. The farmer takes a random sample of 10 cows and weighs them.

Find the probability that at most one cow is *beefy*.

Answer:

State the variables and distributions

Let M be the mass of a cow

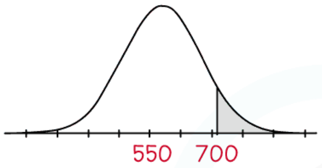
$M \sim N(550, 80^2)$

Let X be the number of beefy cows in the sample

$X \sim B(10, p)$ where p is the probability that a cow is beefy

Calculate the value of p

$p = P(M > 700)$



Standardise: $P(M > 700) = P\left(Z > \dfrac{700 - 550}{80}\right)$

$= P(Z > 1.875)$

$= 1 - \Phi(1.875)$

$= 1 - 0.9697$

Use table of values: $p = 0.0303$

Find the probabilty of at most 1 beefy cow

$X \sim B(10, 0.0303)$

$P(X \leqslant 1) = P_0 + P_1$

$\binom{10}{1} = 10$

$\binom{10}{0} = 1$

$= (0.9697)^{10} + 10(0.0303)(0.9697)^9$

Using formula

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

$= 0.96485...$

0.965 (3sf)

Examiner Tips and Tricks

- Always state what your variables and parameters represent. Make sure you know the conditions for when each distribution is (or is not) a suitable model.

Your notes

Approximating the Poisson Distribution

Your notes

Calculating probabilities using a binomial or Poisson distribution can take a while. Under certain conditions we can use a normal distribution to approximate these probabilities. As we are going from a discrete distribution (binomial or Poisson) to a continuous distribution (normal) we need to apply continuity corrections.

Continuity Corrections

What are continuity corrections?

- The **binomial** and **Poisson distribution are discrete** and the **normal distribution is continuous**
- A **continuity correction** takes this into account when using a normal approximation
- The probability being found will need to be changed from a discrete variable, X to a continuous variable, X_N
  - For example, $X = 4$ for Poisson can be thought of as $3.5 \leq X_N < 4.5$ for normal as every number within this interval rounds to 4
  - Remember that for a normal distribution the probability of a single value is zero so $P(3.5 \leq X_N < 4.5) = P(3.5 < X_N < 4.5)$

How do I apply continuity corrections?

- Think about what is **largest/smallest integer** that can be included in the inequality for the discrete distribution and then find its **upper/lower bound**
- $\mathrm{P}(X = k) \approx \mathrm{P}(k - 0.5 < X_N < k + 0.5)$
- $\mathrm{P}(X \leq k) \approx \mathrm{P}(X_N < k + 0.5)$
  - You add 0.5 as you want to include k in the inequality
- $\mathrm{P}(X < k) \approx \mathrm{P}(X_N < k - 0.5)$
  - You subtract 0.5 as you don't want to include k in the inequality
- $\mathrm{P}(X \geq k) \approx \mathrm{P}(X_N > k - 0.5)$
  - You subtract 0.5 as you want to include k in the inequality
- $\mathrm{P}(X > k) \approx \mathrm{P}(X_N > k + 0.5)$
  - You add 0.5 as you don't want to include k in the inequality
- For a closed inequality such as $\mathrm{P}(a < X \leq b)$
  - Think about each inequality separately and use above

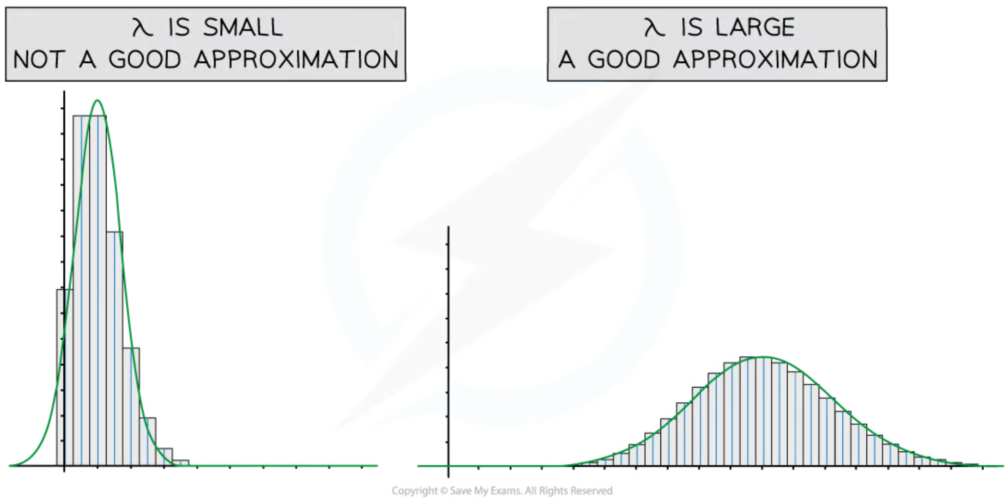
- $\mathrm{P}(X > a) \approx \mathrm{P}(X_N > a + 0.5)$
- $\mathrm{P}(X \le b) \approx \mathrm{P}(X_N > b + 0.5)$
- Combine to give
- $\mathrm{P}(a + 0.5 < X_N < b + 0.5)$

Your notes

Normal Approximation of Poisson

When can I use a normal distribution to approximate a Poisson distribution?

- A Poisson distribution $X \sim Po(\lambda)$ can be **approximated** by a normal distribution $X_N \sim N(\mu, \sigma^2)$ provided
  - **λ is large**
- Remember that the mean and variance of a Poisson distribution are approximately equal, therefore the parameters of the approximating distribution will be:
  - $\mu = \lambda$
  - $\sigma^2 = \lambda$
  - $\sigma = \sqrt{\lambda}$
- The greater the value of λ in a Poisson distribution, the more symmetrical the distribution becomes and the closer it resembles the bell-shaped curve of a normal distribution



Why do we use approximations?

- If there are a large number of values for a Poisson distribution there could be a lot of calculations involved and it is inefficient to work with the Poisson distribution

Your notes

- These days calculators can find Poisson probabilities so approximations are no longer necessary
- However it can still be **easier** to work with a normal distribution
  - You can calculate the probability of a range of values quickly
  - You can use the inverse normal distribution function (most calculators don't have an inverse Poisson distribution function)

How do I approximate a probability?

- **STEP 1**: Find the **mean** and **variance** of the approximating distribution
  - $\mu = \sigma^2 = \lambda$
- **STEP 2**: Apply **continuity corrections** to the inequality
- **STEP 3**: Find the **probability** of the new corrected inequality
  - Find the standard normal probability and use the **table of the normal distribution**
- The probability will not be exact as it is an approximation but provided λ is large enough then it will be a close approximation

Worked Example

The number of hits on a revision web page per hour can be modelled by the Poisson distribution with a mean of 40. Use a normal approximation to find the probability that there are more than 50 hits on the webpage in a given hour.

Answer:

Let X be the number of hits per hour, then $X \sim Po(40)$

$\lambda = 40$ is large so X can be approximated by $X_N \sim N(40, 40)$

Apply continuity corrections to $P(X > 50)$

Strict inequality so use $P(X_N > 50 + 0.5)$

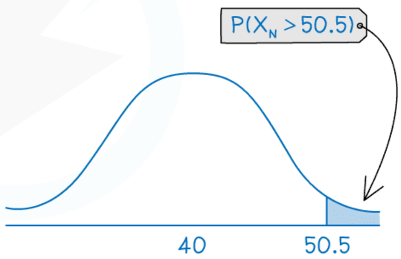
$P(X > 50) \approx P(X_N > 50.5)$

$= P\left(Z > \frac{50.5 - 40}{\sqrt{40}}\right)$

$= 1 - \Phi(1.660)$

$= 1 - 0.9515$

$= 0.0485$



$P(X > 50) \approx 0.0485$

Your notes

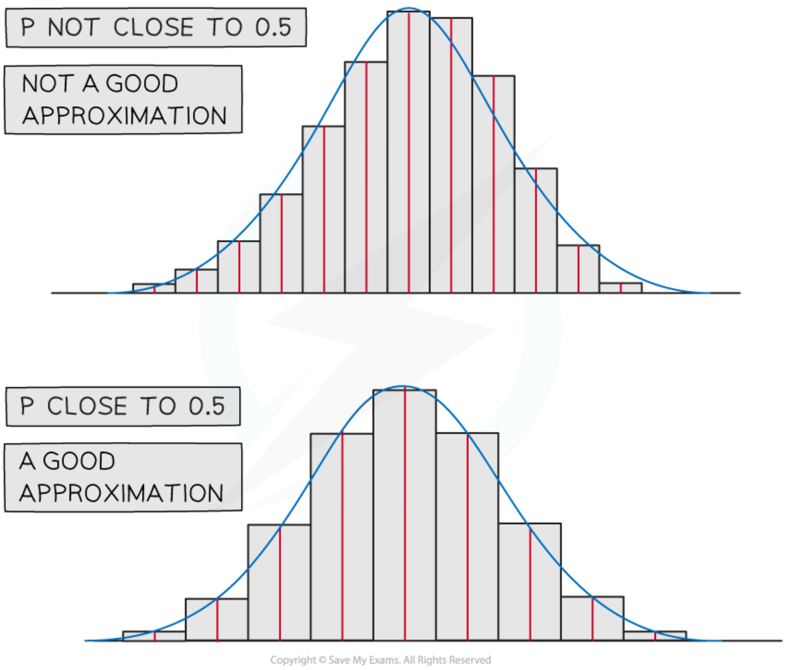
Examiner Tips and Tricks

- The question will make it clear if an approximation is to be used, λ will be bigger than the values in the formula booklet.

Normal Approximation of Binomial

When can I use a normal distribution to approximate a binomial distribution?

- A binomial distribution $X \sim \mathrm{B}(n,p)$ can be **approximated** by a normal distribution $X_N \sim \mathrm{N}(\mu,\sigma^2)$ provided
  - ***n* is large**
  - ***p* is close to 0.5**
- The mean and variance of a binomial distribution can be calculated by:
  - $\mu = np$
  - $\sigma^2 = np(1-p)$



Why do we use approximations?

- If there are a large number of values for a binomial distribution there could be a lot of calculations involved and it is inefficient to work with the binomial distribution
  - These days calculators can calculate binomial probabilities so approximations are no longer necessary
  - However it is **easier** to work with a normal distribution

- You can calculate the probability of a range of values quickly
- You can use the inverse normal distribution function (most calculators don't have an inverse binomial distribution function)

Do I need to use continuity corrections?

- Yes!
- As the binomial distribution is discrete and normal distribution is continuous you will need to use continuity corrections
- $P(X=k) \approx P(k-0.5 < X_N < k+0.5)$
- $P(X \leq k) \approx P(X_N < k+0.5)$
- $P(X < k) \approx P(X_N < k-0.5)$
- $P(X \geq k) \approx P(X_N > k-0.5)$
- $P(X > k) \approx P(X_N > k+0.5)$

How do I approximate a probability?

- **STEP 1**: Find the **mean** and **variance** of the approximating distribution
  - $\mu = np$
  - $\sigma^2 = np(1-p)$
- **STEP 2**: Apply **continuity corrections** to the inequality
- **STEP 3**: Find the **probability** of the new corrected inequality
  - Find the standard normal probability and use the **table of the normal distribution**
    - Find the standard normal probability and use the **table of the normal distribution**
- The probability will not be exact as it is an approximate but provided n is large and p is close to 0.5 then it will be a close approximation

Worked Example

The random variable $X \sim \mathrm{B}(1250, 0.4)$.

Use a suitable approximating distribution to approximate $\mathrm{P}(485 \leq X \leq 530)$.

Answer:

Your notes

Step 1: Find $X_N \sim N(\mu, \sigma^2)$

$\mu = np = (1250)(0.4) = 500$

$\sigma^2 = np(1-p) = (1250)(0.4)(0.6) = 300$

$\sigma = \sqrt{300}$

Step 2: Apply continuity corrections

$485 \leqslant X \leqslant 530$

Include 485 Include 530

$484.5 \leqslant X_N \leqslant 530.5$

Step 3: Calculate the probability

$$P(484.5 \leqslant X_N \leqslant 530.5) = P\left(\frac{484.5 - 500}{\sqrt{300}} \leqslant z \leqslant \frac{530.5 - 500}{\sqrt{300}}\right)$$

$$= P(-0.894... \leqslant z \leqslant 1.760...)$$

$$= \Phi(1.76) - (1 - \Phi(0.89))$$

$$= 0.9608 - (1 - 0.8133)$$

$$= 0.7741$$

0.774 (3.s.f.)

Your notes

Poisson Approximation of Binomial

When can I use a Poisson distribution to approximate a binomial distribution?

- A binomial distribution X~B(*n*, *p*)can be **approximated** by a Poisson distribution $X_p \sim Po(\lambda)$ provided
  - ***n* is large** (typically > 50)
  - ***p* is small**
- The mean of a binomial distribution can be calculated by:
  - $\lambda = np$
- The Poisson distribution is derived from the binomial distribution for conditions where *n* is becoming infinitely large and *p* is becoming infinitely small

Do I need to use continuity corrections?

- No!
- As both the binomial distribution and Poisson distribution are discrete there is no need for continuity corrections

Worked Example

It is known that one person in a thousand who checks a revision website will choose to subscribe. Given that the website received 3000 hits yesterday, use a suitable approximation to find the probability that more than 5 people subscribed.

Answer:

Define the distribution:

Let X be the number of people who subscribe to the website, then

$$X \sim B\left(\underbrace{3000}_{\text{Number of trials}}, \underbrace{\frac{1}{1000}}_{\text{Probability of success}}\right)$$

Check:

$n = 3000$ is large

$p = \frac{1}{1000}$ is small

Define the approximating distribution:

$$X_p \sim P(\underbrace{3}_{np = 3})$$

Find the probability:

$$P(X > 5) \approx P(X_p > 5) = 1 - P(X \leqslant 5)$$

$$= 1 - \underbrace{0.9161}_{\text{Using the tables}}$$

$$= 0.0839$$

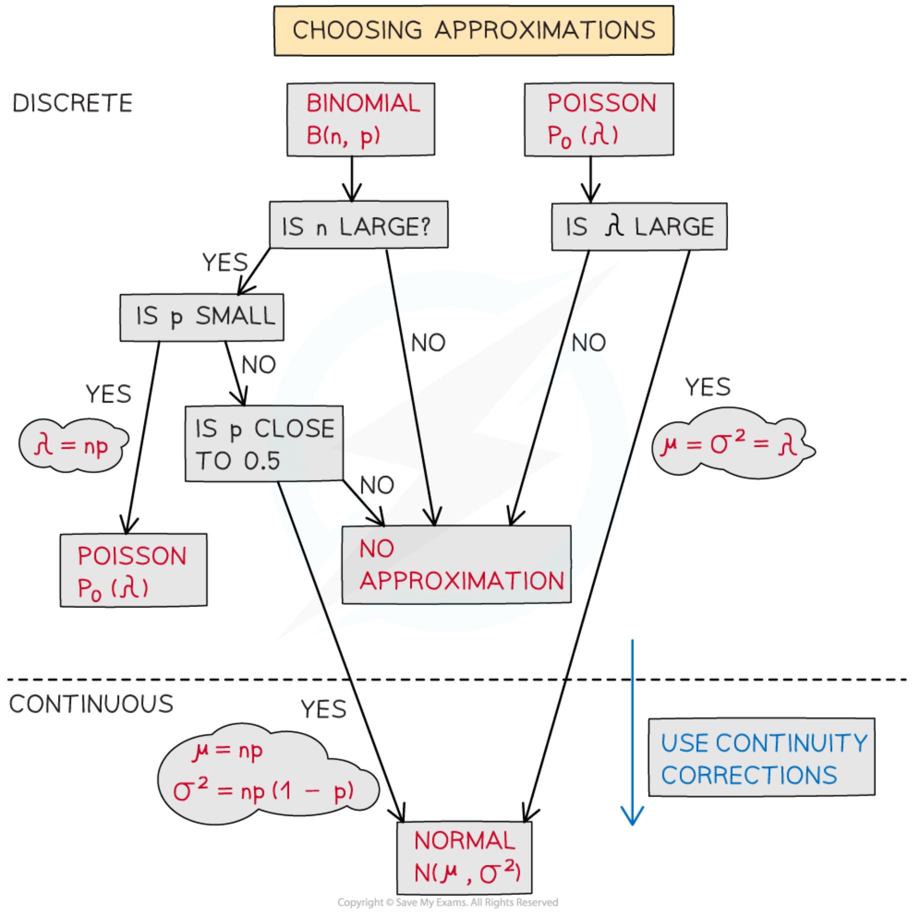
$P(X > 5) \approx 0.0839$ (3.s.f.)

Choosing the Approximation

How will I choose which approximation to use?

- When deciding what approximating distribution to use first make sure you know the reason why you cannot find the probability using the original distribution
  - Is the value of n or λ too large?
  - Will it take too long to carry out the calculations?
- Make sure you know what distribution you are approximating **from**
  - If your distribution is a binomial distribution, you could either use a Poisson (if p is small) or a normal approximation (if p is close to 0.5)
  - If your distribution is a Poisson distribution, you will use a normal approximation
- Use the conditions for approximations to decide which approximation is appropriate
- Calculate the parameters for the approximating distribution

Your notes



Examiner Tips and Tricks

- If you are asked to approximate the binomial distribution but are unsure whether to use Poisson or normal, then calculate the mean and see if it is one of the possible values for λ in the table