Cambridge (CIE) A Level Maths: Probability & Statistics 2

Your notes

Working with Distributions

Contents

Choosing Distributions

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 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

Will I still be expected to use the binomial and geometric distribution

- Knowledge of using the binomial and geometric distribution is expected for Statistics 2

- Remember the **three conditions** for both distributions
  - The trials are **independent**
  - There are **exactly two outcomes** of each trial (**success or failure**)
  - The **probability of success(p) is constant**
- You will be expected to recognise when a random variable follows a **binomial** or **geometric** distribution and use their properties
  - A **binomial distribution** will have a **fixed finite number of trials(n)**
  - A **geometric distribution** will continue the trials until the **first success**

Examiner Tips and Tricks

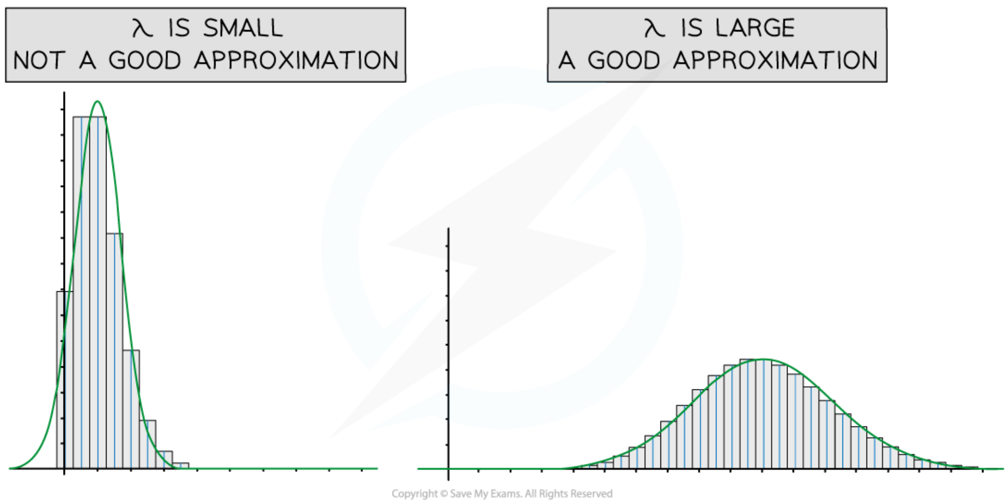
- Always state what your variables and parameters represent. Make sure you know the conditions for when each distribution is (or is not) a suitable model.

Your notes

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 sufficiently large** (> 15)
- 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



What are continuity corrections?

- The **Poisson distribution is 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 $3.5 \le X_N < 4.5$ 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 $P(3.5 \le X_N < 4.5) = P(3.5 < X_N < 4.5)$

Your notes

Do I need to use continuity corrections?

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

How do I approximate a probability?

- **STEP 1**: Find the **mean** and **variance** of the approximating distribution
  - $\mu = \sigma^2 = \lambda$
- **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 (approximately > 15) 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. Find the probability that there are more than 50 hits on the webpage in a given hour.

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 $P(X_N > 50 + 0.5)$

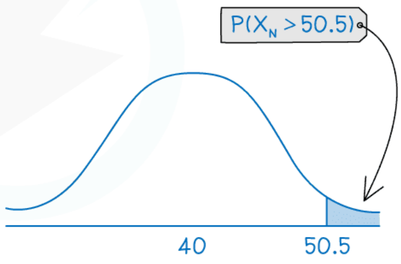
$P(X > 50) \approx P(X_N > 50.5)$

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

$$= 1 - \Phi(1.660)$$

$$= 1 - 0.9515$$

$$= 0.0485$$



$P(X > 50) \approx 0.0485$

Poisson Approximation of Binomial

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

- A binomial distribution B(n,p)can be **approximated** by a Poisson distribution provided
  - ***n* is large** (typically > 50)
  - ***p* is small**
  - ***np* is approximately less than 5**
- 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, find the probability that more than 5 people subscribed.

Your notes

Define the distribution:

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

$$X \sim B\left(3000, \frac{1}{1000}\right)$$

Number of trials

Probability of success

Check:

$n = 3000$ is large

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

$np = 3000 \times \frac{1}{1000} = 3 < 5$ so sufficiently small

Define the approximating distribution:

$$X_P \sim P(3)$$

$np = 3$

Find the probability:

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

$$= 1 - e^{-3}\left(1 + 3 + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!}\right)$$

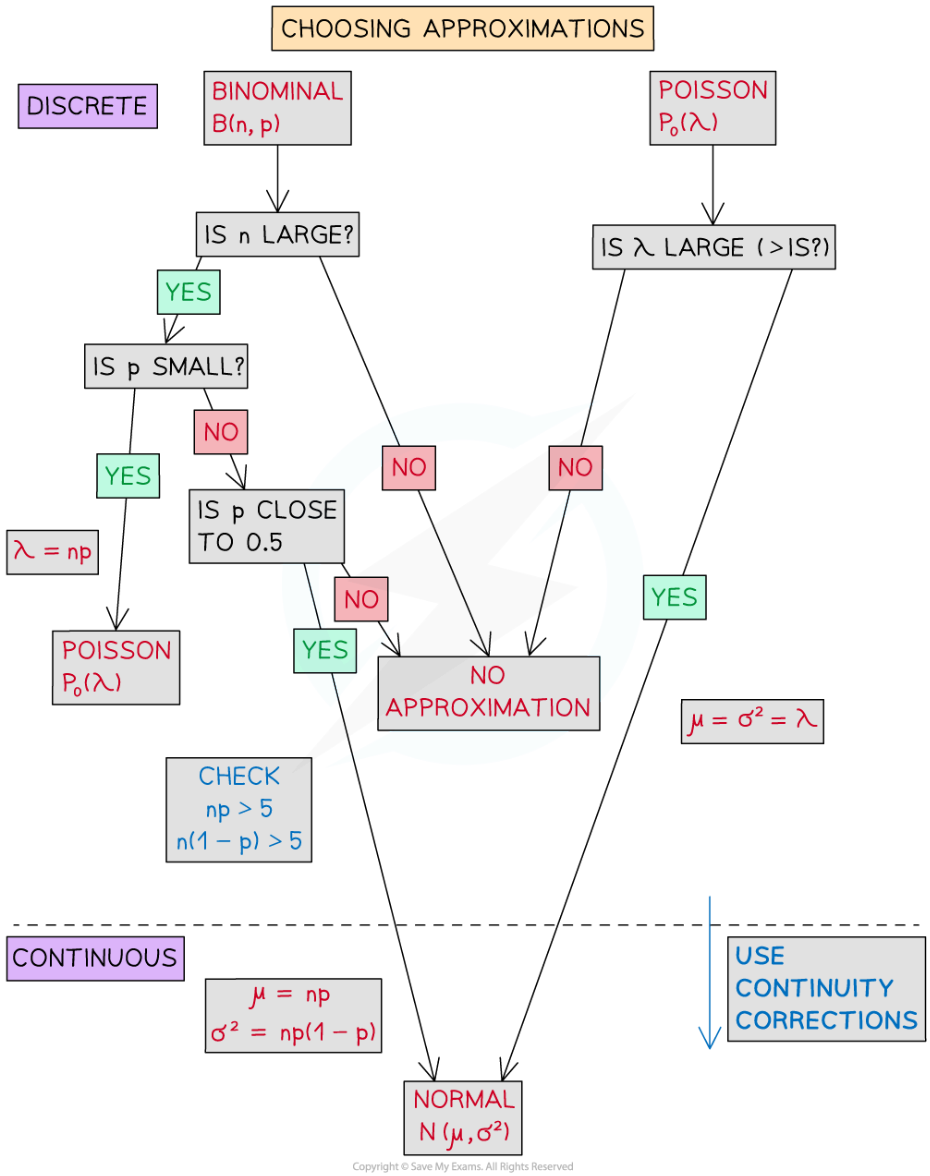
$$= 0.083917\ldots$$

$P(X > 5) \approx 0.0839$ (3sf)

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 or a normal approximation
  - 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 unsure, use the normal distribution to approximate the Poisson and the Poisson distribution to approximate the binomial as these are the two you will be examined on in Statistics 2. However, you should make sure you know the conditions for the normal approximation to the binomial as this may be tested.