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

Your notes

Poisson Distribution

Contents

Your notes

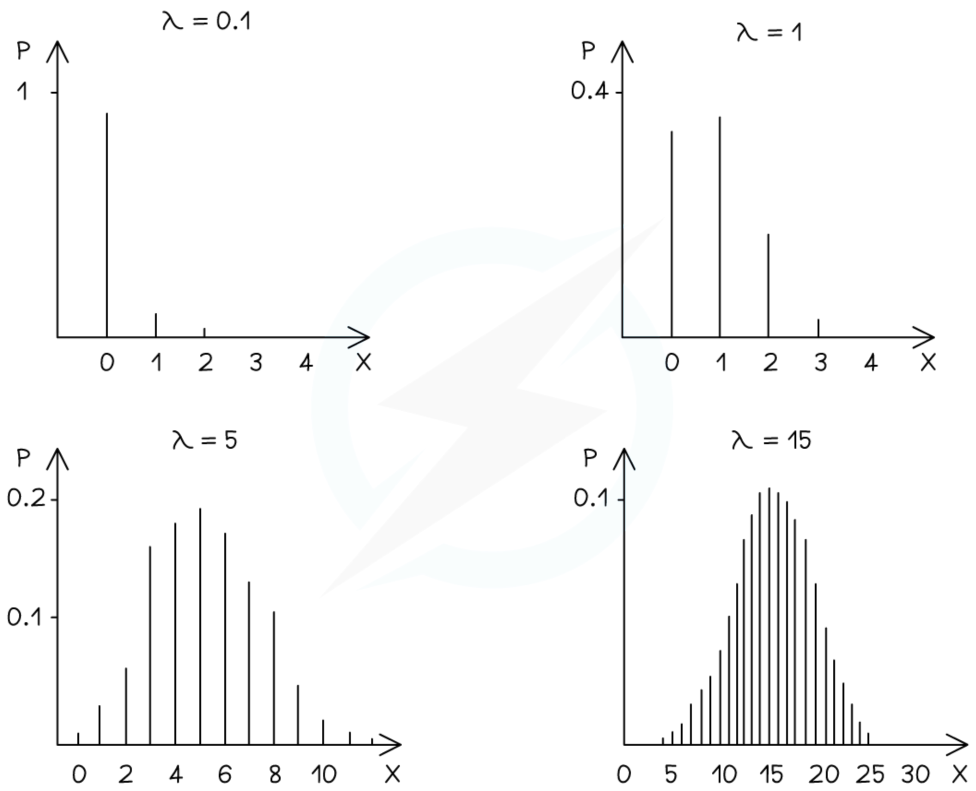
Properties of Poisson Distribution

What is a Poisson distribution?

- A Poisson distribution is a **discrete probability distribution**
- The **discrete random variable** X follows a **Poisson distribution** if it **counts the number of events that occur at random in a given time or space**
- For a Poisson distribution to be valid it **must** satisfy the following properties:
  - Events occur **singly** and at **random** in a **given interval** of time or space
  - The **mean number of occurrences in the given interval(λ)** is known and finite
    - λ has to be positive but does not have to be an integer
  - Each occurrence is **independent** of the other occurrences
- If X follows a Poisson distribution then it is denoted $X \sim \mathrm{Po}(\lambda)$
  - λ is the mean number of trials
- The formula for the probability of r occurrences in a given interval is:
  - $P(X=r)=e^{-\lambda}\times\frac{\lambda^r}{r!}$ for r=0, 1, 2, ...,n
  - e is the constant 2.718...
  - $r!=r(r-1)(r-2)....2\times 1$

What are the important properties of a Poisson distribution?

- The **mean** and **variance** of a Poisson distribution are roughly **equal**
- The distribution can be represented visually using a vertical line graph
  - If λ is **close to 0** then the graph has a **tail to the right** (positive skew)
  - If λ is **at least 5** then the graph is **roughly symmetrical**
- The Poisson distribution becomes more symmetrical as the value of the mean (λ) increases



Worked Example

X is the random variable 'The number of cars that pass a traffic camera per day'.

State the conditions that would need to be met for X to follow a Poisson distribution.

Cars must pass the traffic camera singly and independently

This is likely, cars normally move separately, unless one is being towed

Cars must pass the traffic camera at random

This is less likely as some cars may pass the camera as part of a daily routine

The mean number of cars passing the traffic camera must be known and finite

This could be found by looking at data from the traffic camera

Modelling with Poisson Distribution

How do I set up a Poisson model?

Your notes

Your notes

- Find the **mean** and **variance** and check that they are **roughly equal**
  - You may have to **change the mean** depending on the given time/space interval
- Make sure you **clearly state** what your **random variable** is
  - For example, let *X* be the number of typing errors per page in an academic article
- **Identify** what **probability** you are looking for

What can be modelled using a Poisson distribution?

- Anything that occurs **singly** and **randomly** in a given interval of time or space and satisfies the conditions
- For example, let X be the random variable 'the number of emails that arrive into your inbox per day'
  - There is a given interval of a day, this is an example of an interval of time
  - We can assume the emails arrive into your inbox at random
  - We can assume each email is independent of the other emails
    - This is something that you would have to consider before using the Poisson distribution as a model
  - If you know the mean number of emails per day a Poisson distribution can be used
- Sometimes the given interval will be for space
  - For example, the number of daisies that exist on a square metre of grass
  - look carefully at the units given as you may have to change them when calculating probabilities

Worked Example

State, with reasons, whether the following can be modelled using a Poisson distribution and if so write the distribution.

(i) Faults occur in a length of cloth at a mean rate of 2 per metre.

(ii) On average 4% of a certain population has green eyes.

(iii) An emergency service company receives, on average, 15 calls per hour.

(i) The mean is known and finite
We can assume faults occur singly and randomly
We can assume each fault occurs independently of others

Yes $X \sim Po(2)$

(ii) There is no given interval of time or space

No

(iii) The mean is known and finite
We can assume calls occur singly and randomly
We can assume each call occurs independently of others

Yes $X \sim Po(15)$

Your notes

Examiner Tips and Tricks

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

Calculating Poisson Probabilities

Throughout this section we will use the random variable $X \sim \mathrm{Po}(\lambda)$. For a Poisson distribution, the probability of a X taking a non-integer or negative value is always zero. Therefore any values mentioned in this section (other than λ) will be assumed to be non-negative integers.

Where does the formula for a Poisson distribution come from?

- The formula for calculating an individual **Poisson probability** is
  - $\mathrm{P}(X=r) = p_r = e^{-\lambda} \times \frac{\lambda^r}{r!}$ for r = 0, 1, 2 ,....
- The derivation of the formula comes from the binomial distribution, however it is outside the scope of this syllabus and will not be proved here
  - Whilst the binomial distribution relies on knowing a fixed number of trials, the Poisson can allow for any number of trials within a time period
  - Only the mean number of occurrences of an event within the given period needs to be known

How do I calculate the cumulative probabilities for a Poisson distribution?

- To find an individual probability use the formula $p_r = e^{-\lambda} \times \frac{\lambda^r}{r!}$

- Most of the time you will be required to calculate **cumulative Poisson probabilities** rather than individual ones
  - Use the formula to find the individual probabilities and then add them up
  - Make sure you are confident working with inequalities for **discrete values**
  - Only integer values will be included so it is easiest to look at which integer values you should include within your calculation
  - Sometimes it is quicker to find the probabilities that are **not being asked for** and subtract from one
- $\mathrm{P}(X \le x)$ is asking you to find the probabilities of **all values up to and including *x***
  - This means all values that are **at most *x***
  - Don't forget to include P(*X* = 0)
- $P(X < x)$ is asking you to find the probabilities of **all values up to but not including *x***
  - This means all values that are **less than x**
  - Stop at *x*-1
- $\mathrm{P}(X \ge x)$ is asking you to find the probabilities of **all values greater than and including x**
  - This means all values that are **at least x**
  - As there is **not a fixed number of trials** this includes an **infinite number of possibilities**
  - To calculate this, use the identity $P(X \ge x) = 1 - P(X < x)$
- $P(X > x)$is asking you to find the probabilities of **all values greater than but not including x**
  - This means all values that are **more than x**
  - Rewrite $P(X > x) = 1 - P(X \le x)$ as to calculate this
- If calculating $P(a \le X \le b)$pay attention to whether the probability of a and b should be included in the calculation or not
  - For example, $P(4 < X \le 10)$:
    - You want the integers 5 to 10

How can calculating probabilities for a Poisson distribution be made easier?

- Having to type a lot of calculations into your calculator can be time consuming and cause errors
- Consider the calculation $P(X \le 3) = e^{-\lambda} \times \frac{\lambda^0}{0!} + e^{-\lambda} \times \frac{\lambda^1}{1!} + e^{-\lambda} \times \frac{\lambda^2}{2!} + e^{-\lambda} \times \frac{\lambda^3}{3!}$
  - Note that $e^{-\lambda}$exists in every term and can be factorised out

Your notes

Your notes

- Recall that $\lambda^0 = 1$ and 0! = 1
- Recall also that $\lambda^1 = \lambda$ and 1! = 1

- This calculation could be factorised and simplified to
  - $P(X \le 3) = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\right)$
  - This is much simpler and easier to type into your calculator in exam conditions

How do I change the mean for a Poisson distribution?

- Sometimes the mean may be given for a different interval of time or space than that which you need to calculate the probability for
- A given value of λ can be adjusted to fit the necessary time period
  - For example if a football team score a mean of 2 goals an hour and we want to find the probability of them scoring a certain number of goals in a 90 minute game, then we would use the distribution $X \sim \text{Po}(3)$
    - 90 = 1.5 (60) so use 1.5λ
- A very useful property of the Poisson distribution is that if X and Y are **two independent Poisson** distributions then their sums, ***X* + *Y*** is also a **Poisson** distribution
  - If $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\mu)$ then $X + Y \sim \text{Po}(\lambda + \mu)$
  - Note that, for an integer value of a and b greater than 1, $aX + bY$ no longer follows a Poisson distribution

Worked Example

Xiao makes silly mistakes in his maths homework at a mean rate of 2 per page.

(a) Define a suitable distribution to model the number of silly mistakes Xiao would make in a piece of homework that is five pages long. State any assumptions you have made.

a) Let X be the number of silly mistakes Xiao makes per page, then $X \sim \text{Po}(2)$

2 silly mistakes per page = 10 silly mistakes for 5 pages

Then use the distribution Po (10)

Po (10)

(b) Find the probability that in any random page of Xiao's homework book there are

(i) exactly three silly mistakes

(ii) at most three silly mistakes

(iii) more than three silly mistakes.

b) For $X \sim Po(2)$

Note that this question part refers to one page

$\lambda = 2, r = 3$

(i) $P(X = 3) = e^{-2} \times \frac{2^3}{3!} = 0.18044...$

$P(X = 3) = 0.180 \quad (3sf)$

At most

(ii) $P(X \leq 3) = e^{-2} \times \frac{2^0}{0!} + e^{-2} \times \frac{2^1}{1!} + e^{-2} \times \frac{2^2}{2!} + e^{-2} \times \frac{2^3}{3!}$

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

If you can, go straight to this line

$= 0.8571...$

$P(X \leq 3) = 0.857 \quad (3sf)$

More than

(iii) $P(X > 3) = 1 - P(X \leq 3)$

There are an infinite number of values greater than 3

$= 1 - 0.8571$

$= 0.1428...$

$P(X > 3) = 0.143 \quad (3sf)$

Your notes

Examiner Tips and Tricks

- Look carefully at the given time or space interval to check if you need to change the mean before carrying out calculations. Be prepared for this to change between question parts!