Poisson Distributions Cheat Sheet

Edexcel FS1

The Poisson distribution is a discrete probability distribution. We will learn all about when and how we can use it solve problems.

Using the Poisson distribution

If a variable X follows a Poisson distribution, then we write $X \sim Po(\lambda)$. The parameter of the distribution, λ, is generally known as the mean. To calculate probabilities, you can use the formula:

- $P(X=x)=\frac{(e^{-\lambda})(\lambda^x)}{x!}$

Generally, you can also use your calculator to find probabilities but for certain questions where either λ or x is unknown, you may need to make use of the above formula.

- You can use the cumulative Poisson function on your calculator OR the statistical tables to directly find probabilities of the form $P(X \le a)$. This is useful when trying to find more difficult probabilities (e.g. $P(2 \le X \le 4)$ or $P(X > 6)$).

Example 1: The discrete random variable $X \sim Po(3.1)$. Find:
a) $P(X=4)$
b) $P(X \ge 2)$
c) $P(1 \le X \le 4)$

a) Use the formula or your calculator.	$P(X=4)=\frac{(e^{-3.1})(3.1^4)}{4!}=0.173$
b) Rewrite $P(X \ge 2)$ as $1-P(X \le 1)$, then use either the statistical tables or your calculator.	$P(X \ge 2)=1-P(X \le 1)$ $=1-0.185=0.815$
c) Rewrite $P(1 \le X \le 4)$ as $P(X \le 4)-P(X \le 0)$, then use either the statistical tables or your calculator.	$P(1 \le X \le 4)=P(X \le 4)-P(X \le 0)$ $=0.798-0.045=0.753$

Modelling with the Poisson distribution

You need to be able to recognise scenarios where a Poisson distribution would be a suitable model. Most problems will be given in the context of a question.

In order for the Poisson distribution to be a suitable model, the events must occur:

- independently
- singly, in space or time. (two events cannot occur at the same time)
- at a constant average rate (so that the mean number in an interval is proportional to the length of the interval).

Example 2: A call centre agent handles telephone calls at a rate of 18 per hour.

a) Give two reasons to support the use of a Poisson distribution as a suitable model for the number of calls per hour handled by the agent.

a) State two assumptions that are met by the context. Any two from these three would be fine.	- Calls occur singly - Calls occur at a constant rate - Calls occur independently

b) Find the probability that in any randomly selected 15 minute interval the agent handles
(i) exactly 5 calls,
(ii) more than 8 calls.

i) We first need to set up our distribution. In an hour, the average number of calls was 18. So, in 15 minutes, the average number of calls will be $\frac{18}{4}=4.5$.	Let X be the number of calls in a 15 minute interval. Then $X \sim Po(4.5)$
We need to find $P(X=5)$. Either use the formula or your calculator.	$P(X=5)=\frac{(e^{-4.5})(4.5^5)}{5!}=0.171$
ii) We need to find $P(X \ge 9)$. Rewrite $P(X \ge 9)$ as $1-P(X \le 8)$ then use the cumulative function on your calculator or the statistical tables.	$P(X \ge 9)=1-P(X \le 8)=1-0.960$ $=0.040$

Adding Poisson distributions

- If $X \sim Po(\lambda)$ and $Y \sim Po(\mu)$, then $X+Y \sim Po(\lambda+\mu)$, provided X and Y are independent and within the same limitation (e.g. both the values stated are for every 5 minutes).

Example 3: A secretary receives internal calls at a rate of 1 every 5 minutes and external calls at a rate of 2 every 5 minutes. Calculate the probability that the total number of calls is:

a) 3 in a 4-minute period.
b) at least 2 in a 2-minute period.
c) no more than 5 in a 10-minute period.

a) Let X be the number of internal calls every minute and Y be the number of external calls every minute. Then $X \sim Po\left(\frac{1}{5}\right)$ and $Y \sim Po\left(\frac{2}{5}\right)$. Let Z represent the total number of calls every minute. Then $Z \sim Po\left(\frac{3}{5}\right)$. For 4 minutes, $Z_1 \sim Po\left(\frac{12}{5}\right)$.	$P(required)=P(Z_1=3)=\frac{(e^{-2.4})(2.4^3)}{3!}$ $=0.209$
b) For 2 minutes, $Z_2 \sim Po\left(\frac{6}{5}\right)$.	$P(required)=P(Z_2 \ge 2)=1-P(Z_2 \le 1)$ $=1-0.663=0.337$
c) For 10 minutes, $Z_3 \sim Po(6)$.	$P(required)=P(Z_3 \le 5)=0.446$

Mean and variance of a Poisson distribution

- If $X \sim Po(\lambda)$, then mean of X = variance of $X=\lambda$

This is a very important property in determining whether a Poisson distribution is a suitable model for a particular scenario.

Example 4: A student is investigating the numbers of cherries in a fruit scone. A random sample of 100 fruit scones is taken and the results can be summarised as:
$\sum x=143, \sum x^2=347$
a) Calculate the mean and the variance of the data.
b) Explain why the results in part a) suggest that a Poisson distribution may be a suitable model for the number of cherries in a fruit scone.

a) Mean= $\frac{\sum x}{n}$	Mean= $\frac{143}{100}=1.43$
Variance= $\frac{\sum x^2}{n}-\bar{x}^2$	Variance= $\frac{347}{100}-1.43^2=1.43\ (3\ s.f)$
b) State that the mean is approximately equal to the variance.	Mean≈Variance, which is a characteristic of the Poisson distribution.

Mean and variance of a Binomial distribution

- If $X \sim B(n,p)$, then:
  - ⇒ Mean of $X=E(X)=\mu=np$
  - ⇒ Variance of $X=Var(X)=\sigma^2=np(1-p)$

Some questions will require you to use these definitions to find the value of n or p for a particular binomial distribution.

Example 5: A company produces a certain type of delicate component. The probability of any one component being defective is p. The probability of obtaining at least one defective component in a sample of 4 is 0.3439. The company produces 600 components in a day.
Find the mean and variance of the number of defective components produced per day.

Start off by defining the binomial distribution for a sample of 4.	Let X be the number of defective components in a sample of 4. Then $X \sim B(4,p)$
We are told that $P(X \ge 1)=0.3439$.	$P(X \ge 1)=1-P(X=0)=0.3439$ $\therefore P(X=0)=0.6561$
Using the probability mass function for a binomial random variable to find $P(X=0)$ in terms of p.	$P(X=0)=\binom{4}{0}(p)^0(1-p)^4=(1-p)^4$
We already found that $P(X=0)=0.6561$.	$\therefore (1-p)^4=0.6561$
Solve for p.	$1-p=0.6561^{\frac{1}{4}}=0.9$ $\therefore p=0.1$
We want to find the mean/variance for the number of defective components in an entire day, so consider a new binomial distribution where $n=600, p=0.1$.	For a whole day, $X_2 \sim B(600,0.1)$.
For 600 components, $X_2 \sim B(600,0.1)$. So, $n=600, p=0.1$	$mean=np=(4)(0.1)=0.4$
Find the variance.	$variance=np(1-p)=(600)(0.1)(0.9)=54$

Using the Poisson distribution to approximate the binomial distribution

- If $X \sim B(n,p)$ and
  - ⇒ n is large
  - ⇒ p is small

  then $X \approx \sim Po(np)$. We say that X approximately follows a Poisson distribution with mean np.

There is no set range of values that defines what is meant by 'large' or 'small'. Questions will generally tell you when you need to use a Poisson approximation. Remember that the larger n is and the smaller p is, the better the approximation.

Example 6: Each cell of a certain animal contains 11000 genes. It is known that each gene has a probability 0.0005 of being damaged. A cell is chosen at random.
a) Suggest a suitable model for the distribution of the number of damaged genes in the cell.
b) Find the mean and variance of the number of damaged genes in the cell.
c) Using a suitable approximation, find the probability that there are at most 2 damaged genes in the cell.

a) We initially have a binomial distribution:	Let X be the number of damaged genes in a cell. Then $X \sim B(11000,0.005)$.
b) Remembering the mean and variance formulae for a binomial random variable.	$mean=np=11000 \times 0.0005=5.5$ $variance=np(1-p)=5.5(1-0.0005)=5.50\ (3\ s.f.)$
c) Apply the Poisson approximation.	n is large, p is small. $np=11000 \times 0.0005=5.5$ $\therefore X \approx \sim Po(5.5)$
We need to find $P(X \le 2)$	$P(required)=P(X \le 2)=0.088$ (by calculator).

Exam-style questions

We will now go through two more examples of questions which are of a slightly different style than the rest of the examples in this cheat sheet.

Example 7: A machine which manufactures nails is known to produce 2.5% defective nails. The nails are sold in packets of 200.
a) Using a Poisson approximation, calculate the probability that a packet contains more than 6 defective nails
A carpenter buys 6 packets of nails.
b) Estimate the probability that more than half of these packets contain more than 6 defective nails.

a) We initially have a binomial distribution.	Let X be the number of defective nails in a packet of 200. Then $X \sim B(200,0.025)$
Apply the Poisson approximation.	n is large, p is small. $np=200 \times 0.025=5$ $\therefore X \approx \sim Po(5)$
We need to find $P(X \ge 7)$	$P(X \ge 7)=1-P(X \le 6)=1-0.7620=0.238$
b) We now have 6 different packets. The probability that each pack has more than 6 defective nails is constant so we can use a binomial distribution.	Let Y be the number of packets with more than 6 defective nails. Then $Y \sim B(6,0.238)$.
We need to find $P(Y \ge 4)$.	$P(required)=P(Y \ge 4)=1-P(Y \le 3)=1-0.968$ $=0.032$

Example 8: A company receives telephone calls at random at a mean rate of 2.5 per hour.

a) Find the probability that the company receives exactly 3 telephone calls in the next 15 minutes.
b) Find, to the nearest minute, the maximum length of time the telephone can be left unattended so that the probability of missing a telephone call is less than 0.2.

a) Figure out the mean for a 15 minute interval.	For an hour, the mean is 2.5. So, for 15min (quarter of an hour), the mean will be $\frac{2.5}{4}=\frac{5}{8}$.
Set up the correct distribution then use your calculator to find the required probability.	Let X be the number of calls in 15min. Then $X \sim Po\left(\frac{5}{8}\right)$. $P(required)=P(X=3)=0.022$
b) Since we are trying to find the length of time here, we let Y be the number of calls in t hours.	Let X be the number of calls in t hours. Then $Y \sim Po(2.5t)$
The question is really just asking us to find the maximum time for which $P(Y=0)<0.2$	We want $P(Y=0)<0.2$ $P(Y=0)=\frac{(e^{-2.5t})(2.5t^0)}{0!}=e^{-2.5t}$
Set $P(Y=0)<0.2$.	$e^{-2.5t}<0.2$
Now we need to solve for t. Take logs of both sides.	$-2.5t<\ln(0.2)$
Divide by -2.5.	$t>\frac{\ln 0.2}{-2.5}=0.644$
So, the max time is 0.644 **hours**. Convert this into minutes.	0.644 hours = 0.644×60 minutes = 38.6 So 39 *minutes*