Chapter 8 The Poisson distribution

The Poisson distribution

THE AVONFORD STAR

Full moon madness hits Avonford bypass.

Since opening two years ago, Avonford bypass has seen more than its fair share of accidents but last night was way beyond that ever experienced before. There were no less than 4 separate accidents during the hours of darkness. And it was full moon!!
Was it, we wonder, full moon madness? Or was it just one of those statistical quirks that happen from time to time?
Our Astrology expert, Jessie Manning told us that this was only to be expected when the moon dominates Saturn.
However, the local vicar, the Rev Paul Cheney took a different view. “We must be careful of jumping to the wrong conclusions” he said when we telephoned him this morning. “This is a load of dangerous rubbish that will lead more vulnerable people to believe dangerous things. I am not a Statistician so I cannot tell you what the chances are of there being 4 accidents in one evening, but I reckon that it is a statistical possibility”.

How would you decide whether four accidents in a night are reasonably likely?
The first thing is to look at past data, and so learn about the distribution of accidents.
Since the bypass was opened nearly two years ago, the figures (not including the evening described in the article) are as follows:
(A day is taken to run from one midday to the next.)

Number of accidents per day, x	0	1	2	3	> 3
Frequency, f	395	235	73	17	0

These figures look as though the data could be drawn from a Poisson distribution. This distribution gives the probability of the different possible number of occurrences of an event in a given time interval under certain conditions. If you are thinking of using a Poisson distribution, here is a check list to see if it is suitable.

- The events occur independently
- The events occur at random
- The probability of an event occurring in a given time interval does not vary with time

In this case, the given time interval is one day, or 24 hours. An event is an accident.

The total number of accidents has been $0 \times 395 + 1 \times 235 + 2 \times 73 + 3 \times 17 = 432$
The number of days has been $395 + 235 + 73 + 17 = 720$
So the mean number of accidents per day has been $\frac{432}{720} = 0.6$

The Poisson distribution is an example of a probability model. It is usually defined by the mean number of occurrences in a time interval and this is denoted by λ.

The probability that there are r occurrences in a given interval is given by $\frac{\lambda^r}{r!}e^{-\lambda}$.
The value of e is 2.71281 828 459…… There is a button for it on your calculator.

So, the probability of

0 occurrences is $e^{-\lambda}$
1 occurrence is $\lambda e^{-\lambda}$
2 occurrences is $\frac{\lambda^2}{2!}e^{-\lambda}$
3 occurrences is $\frac{\lambda^3}{3!}e^{-\lambda}$
and so on.

In this example, $\lambda = 0.6$ and so the probabilities and expected frequencies in 720 days are as follows

Number of accidents per day	0	1	2	3	4	5	> 5
Probability (4 d.p.)	0.5488	0.3293	0.0988	0.0197	0.0030	0.0004	0
Expected frequency (1 d.p.)	395.1	237.1	71.1	14.2	2.1	0.3	0

? Explain where the various figures in this table have come from.

? Compare the expected frequencies with those observed. Is the Poisson distribution a good model?

The table shows that with this model you would expect 2.4 days in 720 (i.e. just over 1 a year) where there would be 4 or more accidents. It would seem as though The Rev. Paul Cheney was right; the seemingly high number of accidents last night could be just what the statistical model would lead you to expect. There is no need to jump to the conclusion that there was another factor, such as full moon, that influenced the data.

ACTIVITY 8.1
Use your calculator to find the probability of 0, 1, 2, 3 occurrences of an event which has a Poisson distribution with mean $\lambda = 2.5$.

Use of tables

Another way to find probabilities in a Poisson distribution is to use tables of *Cumulative Poisson probabilities,* like those given in the MEI Students' Handbook.
In these tables you are not given $P(X = r)$ but $P(X \leq r)$. This means that it gives the sum of all probabilities from 0 up to r.

In the example of the accidents on Avonford Bypass the mean, $\bar{x}$, was 0.6 and probabilities of $X = 0$ and 1 were calculated to be 0.5488 and 0.3293.

To find these values in the tables, look at the column for $\lambda = 0.6$. The first entry in this column is 0.5488, representing the probability that there are no accidents.

$x \backslash \lambda$	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865
4	.....	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977
5	.....	.....	.....	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997
6	.....	.....	.....	.....	.....		1.0000	1.0000	1.0000
7	.....	.....	.....	.....	.....	.....	.....	.....	.....

Fig 1.1

The second entry is 0.8781. This is the probability that there will be 0 or 1 accidents.
To find the probability that there is one accident, subtract these two values giving 0.8781 – 0.5488 = 0.3293.
In the same way, the probability that there are 2 accidents is found by taking the second entry from the third.
Continuing the process gives the following.

Number of accidents	Probability	Number of accidents	Probability
0	0.5488	0	0.5488
0 or 1	0.8781	1	0.8781 – 0.5488 = 0.3293
0, 1 or 2	0.9769	2	0.9769 – 0.8781 = 0.0988
0, 1, 2 or 3	0.9966	3	0.9966 – 0.9769 = 0.0197
0, 1, 2, 3 or 4	0.9996	4	0.9996 – 0.9966 = 0.0030

? How can you use the tables to find the probability of exactly 5 accidents in any night?
Check that you get the same answer entering $\frac{(0.6)^5}{5!}e^{-0.6}$ into your calculator.

! You can see that the probability of having 4 accidents in one night is 0.0030. The probability of having 4 *or more* accidents in one night is 1 – probability of having 3 or fewer accidents, which is 0.9966. So the probability of having 4 or more accidents is 1 – 0.9966 = 0.0034.
In other words 34 in every 10 000 days or roughly 2.5 days in 720. This confirms that it is not necessary to look for other explanations for the 4 accidents in the same night.

! You will see that the tables in the Students' Handbook cover values of λ from 0.01 to 8.90. You will clearly have a problem if you are trying to calculate probabilities with a value of λ that is not given in the tables. In such cases you will need to use the formula.

ACTIVITY 8.2

You were asked to find the probabilities of 0, 1, 2, 3 occurrences λ =2.5 in the previous activity. Now use the cumulative tables to find these probabilities.

Example 8.1

The mean number of typing errors in a document is 1.5 per page.
Find the probability that on a page chosen at random there are

(i) no mistakes,

(ii) more than 2 mistakes.

SOLUTION

If you assume that spelling mistakes occur independently and at random then the Poisson distribution is a reasonable model to use.

(i) For λ = 1.5 the tables give P(0 mistakes) = 0.2231.

x \ λ	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90
0	0.3679	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496
1	0.7358	0.6990	0.6626	0.6268	0.5918	0.5578	0.5249	0.4932	0.4628	0.4337
2	0.9197	0.9004	0.8795	0.8571	0.8335	0.8088	0.7834	0.7572	0.7306	0.7037
3	0.9810	0.9743	0.9662	0.9569	0.9463	0.9344	0.9212	0.9068	0.8913	0.8747
4	0.9963	0.9946	0.9923	0.9893	0.9857	0.9814	0.9763	0.9704	0.9636	0.9559
5	0.9994	0.9990	0.9985	0.9978	0.9968	0.9955	0.9940	0.9920	0.9896	0.9868
6	0.9999	0.9999	0.9997	0.9996	0.9994	0.9991	0.9987	0.9981	0.9974	0.9966
7	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9996	0.9994	0.9992
8	.....	.....	.....	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998
9	.....	.....	.....	.....	.....	.....	.....	1.0000	1.0000	1.0000

Fig 1.2

(ii) P(more than 2 mistakes) = 1 − P(up to 2 mistakes)
= 1 − 0.8088
= 0.1912.

? How would you answer this question using the Poisson Formula? Check that you get the same answers.

Historical note

Simeon Poisson was born in France in 1781. He worked as a mathematician in Paris for most of his life after giving up the study of medicine. His contribution to mathematics embraced electricity, magnetism and planetary orbits and ideas in integration as well as in statistics. He wrote over 300 papers and articles.
The modelling distribution that takes his name was originally derived as an approximation to the binomial distribution.

Exercise 8A

1 The number of cars passing a point on a country lane has a mean 1.8 per minute. Using the Poisson distribution, find the probability that in any one minute there are

(i) no cars, **(ii)** 1 car, **(iii)** 2 cars, **(iv)** 3 cars, **(v)** more than 3 cars.

2 A fire station experiences an average call-out rate of 2.2 every period of three hours. Using the Poisson distribution, find the probability that in any period of 3 hours there will be

(i) no callouts, **(ii)** 1 callout, **(iii)** 2 callouts, **(iv)** 3 callouts,
(v) 4 callouts, **(v)** more than 4 callouts.

3 The number of radioactive particles emitted in a minute from a meteorite is recorded on a Geiger counter. The mean number is found to be 3.5 per minute. Using the Poisson distribution, find the probability that in any one minute there are

(i) no particles,

(ii) 2 particles,

(iii) at least 5 particles.

4 Bacteria are distributed independently of each other in a solution and it is known that the number of bacteria per millilitre follows a Poisson distribution with mean 2.9.
Find the probability that a sample of 1 ml of solution contains

(i) 0, **(ii)** 1, **(iii)** 2, **(iv)** 3, **(v)** more than 3 bacteria.

5 The demand for cars from a car hire firm may be modelled by a Poisson distribution with mean 4 per day.

(i) Find the probability that in a randomly chosen day the demand is for

(***A***) 0, (***B***) 1, (***C***) 2, (***D***) 3 cars.

(ii) The firm has 5 cars available for hire. Find the probability that demand exceeds the number of cars available.

6 A book of 500 pages has 500 misprints. Using the Poisson distribution, estimate to three decimal places the probabilities that a given page contains

(i) exactly 3 misprints,

(ii) more than 3 misprints.

7 190 raisins are put into a mixture which is well stirred and made into 100 small buns. Which is the most likely number of raisins found in a bun?

8 Small hard particles are found in the molten glass from which glass bottles are made. On average 20 particles are found in 100 kg of molten glass. If a bottle made of this glass contains one or more such particles it has to be discarded.
Bottles of mass 1 kg are made using this glass.

(i) Criticise the following argument:
Since the material for 100 bottles contains 20 particles, approximately 20% will have to be discarded.

(ii) Making suitable assumptions, which should be stated, develop a correct argument using a Poisson model and find the percentage of faulty 1 kg bottles to 3 significant figures.

9 A hire company has two lawnmowers which it hires out by the day. The number of demands per day may be modelled by a Poisson distribution with mean 1.5. In a period of 100 working days, how many times do you expect

(i) neither lawnmower to be used,

(ii) some requests for a lawnmower to have to be refused?

(MEI)

Conditions for modelling data with a Poisson distribution

You met the idea of a probability model in Z1. The binomial distribution is one example. The Poisson distribution is another model. A model in this context means a theoretical distribution that fits your data reasonably well.

You have already seen that the Poisson distribution provides a good model for the data for the Avonford Star article on accidents on the bypass.

Here are the data again.

Number of accidents per day, x	0	1	2	3	> 3
Frequency, f	395	235	73	17	0

For these data, $n = 720$, $\sum xf = 432$ and $\sum x^2 f = 680$

So $\bar{x} = \dfrac{432}{720} = 0.6$

$S_{xx} = \sum x^2 f - n\bar{x}^2 = 680 - 720 \times 0.6^2 = 420.8$

So the variance $= \dfrac{S_{xx}}{n-1} = 0.585$

You will notice that the mean, 0.6, and the variance, 0.585, are very close in value. This is a characteristic of the Poisson distribution and provides a check on whether it is likely to provide a good model for a particular data set.

In the theoretical Poisson distribution, the mean and the variance are equal. However, it is usual to call λ the **parameter** of a Poisson distribution, rather than either the mean or the variance. The common notation for describing a Poisson distribution is Poisson(λ); so Poisson(2.4) means the Poisson distribution with parameter 2.4.

You should check that the conditions on page 1 apply - that the events occur at random, independently and with fixed probability.

Example 8.2

A mail order company receives a steady supply of orders by telephone. The manager wants to investigate the pattern of calls received so he records the number of calls received per day over a period of 40 days as follows.

Number of calls per day	0	1	2	3	4	5	> 5
Frequency of calls	8	13	10	6	2	1	0

(i) Calculate the mean and variance of the data. Comment on your answers.

(ii) State whether the conditions for using the Poisson distribution as a model apply.

(iii) Use the Poisson distribution to predict the frequencies of 0, 1, 2, 3… calls per hour.

(iv) Comment on the fit.

SOLUTION

(i) Summary statistics for these data are:

$n = 40,\ \sum xf = 64,\ \sum x^2 f = 164$

So mean, $\bar{x} = \dfrac{\sum xf}{n} = \dfrac{64}{40} = 1.6$

$S_{xx} = 164 - 40 \times 1.6^2 = 61.6$

So variance, $s^2 = \dfrac{61.6}{39} = 1.5795$

The mean is close to the variance, so it may well be appropriate to use the Poisson distribution as a model.

(ii) It is reasonable to assume that

- the calls occur independently
- the calls occur at random
- the probability of a call being made on any day of the week does not vary with time, given that there is a steady supply of orders.
-

(iii) From the cumulative tables with $\lambda = 1.6$ gives the following.

Calls	Probability	Calls	Probability	Expected frequency (probability × 40)
0	0.2019	0	0.2019	8.1
0 or 1	0.5249	1	0.5249 – 0.2019 = 0.3230	12.9
0, 1 or 2	0.7834	2	0.7834 – 0.5249 = 0.2585	10.3
0, 1, 2 or 3	0.9212	3	0.9212 – 0.7834 = 0.1378	5.5
0, 1, 2, 3 or 4	0.9763	4	0.9763 – 0.9212 = 0.0551	2.2
0, 1, 2, 3, 4 or 5	0.9940	5	0.9940 – 0.9763 = 0.0177	0.7
0, 1, 2, 3, 4, 5 or 6	0.9987	6	0.9987 – 0.9940 = 0.0047	0.2

(iv) This is the table showing the comparisons.

.

Number of calls per day	0	1	2	3	4	5	> 5
Actual frequency of calls	8	13	10	6	2	1	0
Theoretical frequency of calls (1 d.p.)	8.1	12.9	10.3	5.5	2.2	0.7	0.2

The fit is very good, as might be expected with the mean and variance so close together.

Example 8.3

Avonford Town Football Club recorded the number of goals scored in each one of their 30 matches in one season as follows.

Goals, x	0	1	2	3	4	> 4
Frequency, f	12	12	4	1	1	0

(i) Calculate the mean and variance for this set of data.

(ii) State whether the conditions for using the Poisson distribution apply.

(iii) Calculate the expected frequencies for a Poisson distribution having the same mean number of goals per match.

(iv) Comment on the fit.

SOLUTION

(i) For this set of data

$n = 30, \quad \sum xf = 27, \quad \sum x^2 f = 53$

So mean, $\bar{x} = \dfrac{\sum xf}{n} = \dfrac{27}{30} = 0.9,$

$S_{xx} = \sum x^2 f - n\bar{x}^2 = 53 - 30 \times 0.9^2 = 28.7$

So variance, $s^2 = \dfrac{28.7}{29} = 0.9897$

(ii) It is reasonable to assume that

- The goals are scored independently
- The goals are scored at random.
- the probability of scoring a goal is constant from one match to the next.

In addition, the value of the mean is close to the value of the variance. Hence, the Poisson distribution can be expected to provide a reasonably good model.

(iii) From the cumulative probability tables for $\lambda = 0.9$.

Goals	Probability	Goals	Probability	Expected frequency (probability × 30)
0	0.4066	0	0.4066	12.2
0 or 1	0.7725	1	0.7725 – 0.4066 = 0.3659	11.0
0, 1 or 2	0.9371	2	0.9371 – 0.7725 = 0.1646	4.9
0, 1, 2 or 3	0.9865	3	0.9865 – 0.9371 = 0.0494	1.5
0, 1, 2, 3 or 4	0.9977	4	0.9977 – 0.9865 = 0.0112	0.3

(iv) As expected from the closeness of the mean and the variance values, the fit is very good.

This is the table showing the comparisons.

Goals, x	0	1	2	3	4	> 4
Actual frequency, f	12	12	4	1	1	0
Theoretical frequency	12.3	11.0	4.9	1.5	0.3	0

? In example 8.2 above, it was claimed that the goals scored in a match were independent of each other. To what extent do you think this is true?

Exercise 8B

1 The number of bacteria in 50 100cc samples of water are given in the following table.

Number of bacteria per sample	0	1	2	3	4 or more
Number of samples	23	16	9	2	0

(i) Find the mean number and the variance of bacteria in a 100cc sample.

(ii) State whether the conditions for using the Poisson distribution as a model apply.

(iii) Using the Poisson distribution with the mean found in part (i), estimate the probability that another 100cc sample will contain

(*A*) no bacteria,

(*B*) more than 4 bacteria.

2 Avonford Town Council agree to install a pedestrian crossing near to the library on Prince Street if it can be shown that the probability that there are more than 4 accidents per month exceeds 0.1.
The accidents recorded in the last 10 months are as follows:

3 2 2 1 0 2 5 4 3 1

(i) Calculate the mean and variance for this set of data.

(ii) Is the Poisson distribution a reasonable model in this case?

(iii) Using the Poisson distribution with the mean found in part (i), find the probability that, in any month taken at random, there are more than 4 accidents.
Hence say whether Avonford Town Council should install the pedestrian crossing.

3 The numbers of customers entering a shop in forty consecutive periods of one minute are given below

3	0	0	1	0	2	1	0	1	1
0	3	4	1	2	0	2	0	3	1
1	0	1	2	0	2	1	0	1	2
3	1	0	0	2	1	0	3	1	2

(i) Draw up a frequency table and illustrate it by means of a vertical line graph.

(ii) Calculate values of the mean and variance of the number of customers entering the shop in a one minute period.

(iii) Fit a Poisson distribution to the data and comment on the degree of agreement between the calculated and observed values.

(MEI - part)

4 A machine in a factory produces components continuously. Each day a sample of 20 components are selected and tested. Over a period of 30 days the number of defective components in the sample is recorded as follows.

Number of defectives per sample	0	1	2	3	4	> 4
Number of samples	8	9	8	3	2	0

The quality Control Inspector says that he will stop the production if any sample contains 5 or more defective components.

(i) Find the mean and variance of the number of defectives per sample.

(ii) State whether the data can be modelled by the Poisson distribution.

(iii) Using the Poisson distribution with the mean found in part (i), find the probability that on any one day the quality control inspector will stop the production.

5 In a college, the number of accidents to students requiring hospitalisation in one year of 30 weeks is recorded as follows.

Number of accidents requiring hospitalisation each week	0	1	2	3 or more
Frequency	25	4	1	0

The Principal uses these data to assesses the risk of such accidents.

(i) Is the Poisson distribution a suitable model for this assessment?
State the assumptions that need to be made about the data provided for this to be so.

(ii) Assuming that the Poisson distribution is a suitable model calculate the probability that

(*A*) In any one week there will be 3 accidents requiring hospitalisation,

(*B*) In a term of 8 weeks there will be no accidents.

6 A count was made of the red blood corpuscles in each of the 64 compartments of a haemocytometer with the following results:

Number of corpuscles	2	3	4	5	6	7	8	9	10	11	12	13	14
Frequency	1	5	4	9	10	10	8	6	4	3	2	1	1

Estimate the mean and variance of the number of red blood corpuscles in each compartment. Explain how the values you have obtained support the view that these data are a sample from a Poisson population.

Using the Poisson distribution as a model, estimate the expected number of compartments containing 2, 3, 4 and 5 red corpuscles. (MEI, adapted)

7 The coach arriving into Avonford Coach station each morning at 0800 has a capacity of 54 passengers. (On this route standing passengers are not allowed.)
Over a period of time the number of passengers alighting at the Coach Station on 20 days was recorded as follows.

50	50	50	50	50	51	51	51	51	51
51	51	52	52	52	52	52	52	53	54

(i) Show that in this case the Poisson distribution is not a suitable distribution to model the number of passengers arriving each day.

In fact, 50 of the passengers were regular commuters. The random variable, X, is the number of passengers arriving at Avonford on this coach other than the 50 regulars.

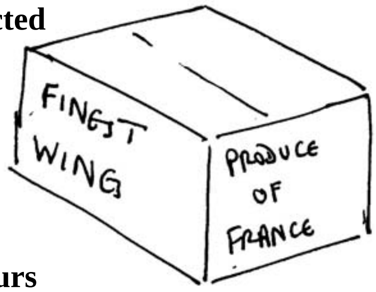
(ii) Show that the Poisson distribution is a reasonable fit for these data.
Use this distribution with your mean for the data above to estimate the probability that on any given day someone has been left behind because the coach was full.

The sum of two or more Poisson distributions

Example 8.4

Special Offer exclusive to Avonford Star readers

A case of the finest specially selected fine wines from France



Buy on line or by telephone

www.buyitonline.com

Or telephone during working hours
0123 456789

A company accepts orders for their product either by telephone or on-line.
The mean number of sales per day are: telephone order: 1.5
on-line purchase: 2.5

What is the probability of the company receiving exactly 2 orders on one day?

SOLUTION
In this example you are interested in the total number of orders per day.
This has mean 1.5 + 2.5 = 4.

If two independent distributions are both Poisson with means λ and μ then the sum of the distributions is Poisson with mean $\lambda + \mu$.
So in this case the distribution of the total orders per day is Poisson with parameter 4.

From the cumulative Poisson distribution tables, the probability of obtaining exactly 2 orders is 0.2381 – 0.0916 = 0.1465.

! The conditions for adding two Poisson distributions together are that the random variables are independent of each other.

Enrichment:
An alternative solutions involves listing the ways in which 2 orders in one day could be obtained by the two methods.

The 2 orders could be as follows.

Written order	Probability ($\lambda = 1.5$)	on line purchase	Probability ($\mu = 2.5$)	Total order probability
0	0.2231	2	0.2565	$0.2231 \times 0.2565 = 0.0572$
1	0.3347	1	0.2052	$0.3347 \times 0.2052 = 0.0687$
2	0.2510	0	0.0821	$0.2510 \times 0.0821 = 0.0206$
			Total:	0.1465

? Explain where the figures in this table come from.

You can imagine that doing such problems this way can rapidly become far too cumbersome. For example, to find the probability of receiving 8 orders would involve a great deal of work!

Example 8.5

A family categorises the mail received each day into three types; financial (bills, communications from the bank, etc), circulars (and all kinds of junk mail) and personal (letters from friends, etc). The three types may each be modelled by independent Poisson distributions with means 1.1, 1.8 and 1.3 items of mail per day.

Find the probability that in any one day

(i) the number of items of mail that are personal or financial exceeds 3,

(ii) the total number of items of mail is less than 5.

SOLUTION

You are given that the delivery of the different types of mail may be modelled by Poisson distributions and it is reasonable to assume that

- they are independent
- they occur randomly
- the probability of a delivery does not vary with time

(i) Personal or financial mail:
In this case the parameter is 1.1 + 1.3 = 2.4
From cumulative Poisson distribution tables
P(number exceeds 3) = 1 – P(number is 3 or less)
= 1 – 0.7787
= 0.2213

(ii) All mail:
In this case the parameter is 1.1 + 1.8 + 1.3 = 4.2

From tables, with $\lambda = 4.2$, P(total number is less than 5)
= P(total number is less than *or equal to* 4) = 0.5898

Exercise 8C

1 The lost property office at a mainline station records items left on trains handed in in two categories: Category A is clothing (umbrella, coats, gloves, etc) and category B is paperwork (briefcase, papers, books, etc).
Over a period of time the mean number of items in category A handed in per day is 2 and the mean number in category B per day is 3.

These distributions may be assumed to be independent of each other and each may be modelled by the Poisson distribution.

Find the probability that

(i) in one day there are no items handed in,

(ii) in one day there is exactly one item handed in,

(iii) in one day there are exactly 2 items handed in.

2 On a particular stretch of road one morning the number of cars travelling into town past a particular point per 10 second interval has a mean of 3.0 and the number of car travelling out of town past the same point has a mean of 1.3 per 10 second interval. Assuming that both may be modelled by independent Poisson distributions find the probability that in any interval of 10 seconds

(i) no cars pass in either direction,

(ii) a total of 5 cars pass.

3 A certain blood disease is caused by the presence of two types of deformed corpuscles, A and B. It is known that if the total number of deformed corpuscles exceeds 6 per 0.001cm^3 of blood then the patient will contract the blood disease.
For a particular family it is estimated that the number of type A per 0.001cm^3 of blood can be modelled by a Poisson distribution with mean 1.3 and the number of type B per $0.001\ \text{cm}^3$ of blood can be modelled by a Poisson distribution with mean 1.6.

(i) Find the probability that one person in this family will contract the disease.

(ii) State any assumptions you have made.

4 In the refectory of a college both tea and coffee are sold. During the "non-teaching" parts of the day the number of cups of coffee and tea sold per five minute interval may be considered to be independent Poisson distributions with means 2.7 and 1.5 respectively.

Calculate the probabilities that, in a given five minute interval,

(i) exactly one cup of coffee and one cup of tea are sold,

(ii) exactly two drinks are sold,

(iii) more than five drinks are sold.

5 The water in a tank is contaminated with bacteria. The bacteria are located at random and independently and the mean number per millilitre of liquid is known to be 1.1.

Find the probability that

(i) a sample of 1 ml of liquid contains more than 2 bacteria,

Five samples, each of 1 ml of liquid, are taken.

(ii) Find the mean number of bacteria for the five samples, taken together.

(iii) Find the probability that there are in total (A) 0, (B) less than 3 bacteria in the five samples.

6 A business man receives an average of 2 e-mails per hour related to his business and 1.5 e-mails per hour on personal matters.

Find the probability that in any randomly chosen hour

(i) he receives no e-mails,

(ii) he receives more than 5 e-mails.

7 To justify the building of the Avonford by-pass, a research company carried out an investigation into the number of lorries passing through the centre of the town on the north-south road.

They found that the mean number of lorries per 5 minute interval were 5.0 travelling north and 3.0 travelling south.

Find the probability that in any given interval of 5 minutes

(i) no lorries passed through travelling north,

(ii) a total of less than 6 lorries passed through,

(iii) a total of at least 6 lorries passed through,

(iv) a total of exactly 6 lorries passed through.

8 An insurance company has 40 000 clients covered for "severe industrial accident". Such accidents are estimated to affect 1 in 200 000 of the population in any year, and they are assumed to occur independently of each other. Let X be the number of claims for severe industrial accident received by the company in any one year.

(i) State the distribution of X, and explain why an appropriate Poisson would give a good approximation.

(ii) Find the probability that the number of claims received by the company in a year is

(A) 0,
(B) 1,
(C) 2 or more.

A second insurance company offers cover on the same terms for severe industrial accident, and it receives an average of 1.2 claims per year.

(iii) Estimate the number of clients insured for severe industrial accident with the second company.

(iv) The two companies merge. Find the probability that the number of claims received by the merged company in a year is

(***A***) 0,
(***B***) 2.

The Poisson approximation to the binomial theorem

Look at this example of the use of the binomial distribution.

Example 8.6

In a certain part of the country 1 in 50 children have accidentally broken a bone in their body by the age of five. If 100 five year old children are chosen at random what is the probability that exactly 4 of the children have suffered a broken bone?

SOLUTION

Using the binomial distribution

This distribution is a binomial distribution with $n = 100$, $p = \frac{1}{50}$ and $q = \frac{49}{50}$.

You met this in Chapter 6

The probability of 4 children having had broken bones is ${}^{100}C_4 \left(\frac{1}{50}\right)^4 \left(\frac{49}{50}\right)^{96} = 0.0902$.

Using the Poisson distribution

A calculation like this can be rather cumbersome. In some cases the cumulative binomial probability table can be used, but this is not always so.

In such cases the Poisson distribution can provide a good approximation to the binomial distribution. For this to be the case you require

- n to be large
- p to be small (and so the event is rare)
- np is not to be large (typically less than 10).

In this example, the mean of the binomial distribution is given by $np = 100 \times \frac{1}{50} = 2$ so for the Poisson approximation, take $\lambda = 2$.

Using the tables gives the probability of 4 children having suffered a broken bone is $0.9473 - 0.08571 = 0.0902$.

You will see that the answer is the same as the exact answer from the binomial distribution to 4 decimal places.

ENRICHMENT

Activity 8.2
Accuracy
Working out the equivalent Poisson distribution terms can be very much simpler than the binomial terms, but they are only approximations. How accurate are they?
An important question is how accurate do they need to be? In most cases you are only modelling a distribution based on a few values so this approximation is often good enough.

This table gives the figures obtained above.

Number of cases	Probability by binomial	Probability by Poisson
0	0.1326	0.1353
1	0.2707	0.2707
2	0.2734	0.2707

In this table the corresponding values have the same first two decimal places (though one of them is not accurate to 2 decimal places).

Consider now the situation where the distribution is B(1000, 0.02).
The mean value is still $np = 1000 \times 0.002 = 2$. So if you redraw the table the right hand column remains the same, but the binomial probabilities are much closer.
For instance,

For instance, $\mathrm{P}(X=0) = \left(\frac{499}{500}\right)^{1000} = 0.1351$, $\mathrm{P}(X=1) = 1000\left(\frac{1}{500}\right)\left(\frac{499}{500}\right)^{999} = 0.2707$,

$$\mathrm{P}(X=2) = 1000.\frac{999}{2}\left(\frac{1}{500}\right)^2\left(\frac{499}{500}\right)^{998} = 0.2709, \text{ etc.}$$

These results can be shown in a table.

Number of cases	Probability by binomial	Probability by Poisson
0	0.1351	0.1353
1	0.2707	0.2707
2	0.2709	0.2707

Complete a similar table for B(10 000, 0.0002) and also for B(100 000, 0.00002)

Example 8.7

The probability that a component produced in a factory is defective is 0.005. If a sample of 500 components are tested, find the probability that

(i) exactly 2 components are defective

(ii) more than 2 components are defective.

SOLUTION
The distribution is binomial with $n = 500$ and $p = 0.005$.
This gives the mean as $np = 500 \times 0.005 = 2.5$

The conditions for using the Poisson distribution have been met since

- p is small
- n is large
- $np = 2.5$. This is not large.

So the Poisson approximation to the binomial is appropriate with $\lambda = np = 2.5$.

So from the tables for the cumulative Poisson distribution with $\lambda = 2.5$.

(i) Probability of 2 defectives = 0.5438 − 0.2873 = 0.2565

(ii) Probability of more than 2 defectives = 1 − 0.5438 = 0.4562

Exercise 8D

1 **(a)** It is known that on a production line the probability that an item is faulty is 0.1. 50 items are chosen at random and checked for faults. Find the probability that there will be no faulty items and also the probability that there will be 3 faulty items using

(i) the binomial distribution,

(ii) a Poisson distribution.

(iii) Comment on your answers.

(b) After an improvement in the production line the probability that an item is faulty is now 0.01. 500 items are chosen at random and checked for faults. Find the probability that there will be no faulty items and also the probability that there will be 3 faulty items using

(i) the binomial distribution,

(ii) a Poisson distribution.

(iii) Comment on your answers.

2 A chemical firm produces bottles of shampoo. It is found over a long period of time that 1 in 50 bottles contains enough impurity to render the shampoo unusable. A random sample of 100 bottles is taken. What is the probability that more than 5 of them will be unusable?

3 The mean number of accidents in a factory is known to be 2.8 per month. Records of recorded accidents are scrutinised for a random sample of 10 months.

(a) Justify the use of the Poisson distribution to model this distribution.

(b) Find the probability that

(i) each of the 10 months had at least one accident,

(ii) in exactly 8 of the months there was at least one accident.

(c) If the first three months of a year are chosen, show that it is rather improbable that in this time interval there will be fewer than 3 accidents or more than 15.

4 It is found that 1 in 200 patients who stay more than 3 days in hospital develop an illness that is unrelated to the cause of their admission.

During one year, 1000 people are in the hospital for more than 3 days. Estimate the probability that more than 5 of them develop another disease.

5 A sociologist claims that only 2.5% of all students from inner city schools go on to university. A group of 300 students are randomly chosen from inner city schools from around the country.

(i) Show why the Poisson distribution may be used as an approximation to the binomial distribution to model this situation.

(ii) Calculate the probability that, if the sociologists' claim is true, from this group of 300 students more than 6 go to university.

6 State the conditions under which it is permissible to use the Poisson distribution as an approximation to the binomial distribution.

It is known that 0.5% of components produced by a factory are defective. Each day a random sample of 200 components is inspected.

(i) Find the probability that there are no defectives in the daily sample.

(ii) Find the probability that there is at least one defective on any day.

(iii) How many components are inspected in 3 days?

(iv) Find the probability that there are at least three defective components in a period of

7 A car has a part that lasts, in safari conditions, on average, 1000 miles before failing. A driver is setting out on a safari of 3000 miles. She wants to know how many spare parts to take with her. She fits a new part before she starts.

(i) The Poisson distribution is used to model this situation. State the value of λ.

(ii) Find the probabilities that

(***A***) the fitted part does not fail,
(***B***) 0 or 1 parts fail,
(***C***) 0, 1 or 2 parts fail,
(***D***) 0, 1, 2 or 3 parts fail.

(iii) Find the number of spare items she should carry to be 95% sure of having enough parts to be able to complete the safari.

8 **(a)** Among the population of a large city, the proportion of people with blue eyes is 0.2. A random group of 10 people are selected. Find the probability that there will be

(i) no person with blue eyes,

(ii) at least two people with blue eyes.

(b) In the same population the proportion of people with a randomly occurring medical condition is 0.002.
A random group of 1000 people is selected. Find the probability that there will be

(i) no person with the condition,

(ii) at least two people with the condition.

9 A man is trying to persuade people to join an organisation. He knows that the probability of being successful in persuading a person at random to join is 0.02.

(i) One day he tries 100 people. What is the probability of recruiting at least one person?

(ii) On another day he tries 200 people. What is the probability of recruiting at least one person?

(iii) How many people should he try in order to be 99% sure of recruiting at least one person?

Key points

1 The Poisson probability distribution

The Poisson distribution is used to model the probability distribution of the numbers of occurrences of an event in a given interval when

- Occurrences are independent.
- Occurrences are random.
- The probability of an occurrence is constant over time.

The Poisson distribution is defined by a parameter, λ.

The probability of 0 occurrences is $e^{-\lambda}$

The probability of 1 occurrence is $\frac{\lambda}{1}e^{-\lambda}$

The probability of 2 occurrences is $\frac{\lambda^2}{2!}e^{-\lambda}$

....

The probability of r occurrences is $\frac{\lambda^r}{r!}e^{-\lambda}$

- The value of the mean is often taken to be the parameter λ. The value of the variance of the distribution is close to λ.

The Poisson distribution is often an appropriate model for the probability distribution of the number of occurrences of a *rare* event.

2 The sum of two Poisson distributions

If two independent random variables both have Poisson distributions with parameters λ and μ, then their sum also has a Poisson distribution and its parameter is $\lambda + \mu$.

3 Approximating the binomial distribution by a Poisson distribution

The Poisson distribution may be used to model a binomial distribution, B(n, p) provided that

- n is large
- p is small
- np is not too large.

Solutions to exercises

Page 2 Discussion point 1

Given $\lambda = 0.6$, the probability of 0 accidents = $e^{-0.6} = 0.5488$

Probability of 1 accident = $0.6e^{-0.6} = 0.3293$

and so on.

The expected frequency figures are the result of multiplying the probability by the number of days, 720..

E.g. $0.5488 \times 720 = 395.1$, $0.3293 \times 720 = 237.1$

and so on.

Page 2 Discussion point 2

The table giving the actual and the expected frequencies are as follows:

Number of accidents per day	0	1	2	3	4	5	> 5
Actual frequency	395	235	73	17	0	0	0
Expected frequency (1 d.p.)	395.1	237.1	71.1	14.2	2.1	0.3	0

The fit seems to be very close and so the use of the Poisson distribution to model this situation is reasonable.

Activity 8.1 Page 2

Occurrences	0	1	2	3
Probability (4 d.p.)	0.0821	0.2052	0.2565	0.2138

Activity 8.2 Page 4

The probabilities will the same as in the table for Activity 8.1.

Page 4 Discussion point

This would be worked out by finding the probability of there being 0, 1 and 2 mistakes, adding and subtracting the total from 1.

i.e. probability of 0 mistakes = $e^{-1.5} = 0.2231$

probability of 1 mistake $= 1.5 \times e^{-1.5} = 0.3347$

probability of 2 mistakes $= \dfrac{1.5^2}{2} \times e^{-1.5} = 0.0.2510$

$\Rightarrow$ probability of 0, 1 or 2 mistakes $= 0.2231 + 0.3347 + 0.3347 = 0.8088$

Soi probability of more than 2 mistakes = $1 - 0.8088 = 0.1912$

Exercise 8A

1 $\lambda = 1.8$

(i) 0.1653 **(ii)** 0.2975 **(iii)** 0.2678 **(iv)** 0.1607 **(v)** 0.1087

2 $\lambda = 2.2$

(i) 0.1108 **(ii)** 0.2438 **(iii)** 0.2681 **(iv)** 0.1967 **(v)** 0.1081 **(vi)** 0.0725

3 $\lambda = 3.5$

(i) 0.0302 **(ii)** 0.1849 **(iii)** 0.2746

4 $\lambda = 2.9$

(i) 0.0550 **(ii)** 0.1596 **(iii)** 0.2314 **(iv)** 0.2237 **(v)** 0.3304

5 $\lambda = 4$

(i) (***A***) 0.0183 (***B***) 0.0733 (***C***) 0.1465 (***D***) 0.1954

(ii) 0.2149

6 $\lambda = 1$

(i) 0.0613 **(ii)** 0.0190

7 $\lambda = 1.9$

Number	0	1	2	3
Probabilities	0.1496	0.4337 – 0.1496 = 0.2841	0.7037 – 0.4337 = 0.2700	0.8747 – 0.7037 = 0.1710

So greatest probability is $r = 1$

8 **(i)** The assumption is that only one particle would ever be found in 1 kg of the molten glass.

(ii) Assuming that the particles occur randomly and independently of each other, then a Poisson model is appropriate with $\lambda = 0.2$

Then $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.8187 = 0.1813$

9 $\lambda = 1.5$

(i) Probability of 0 demands = 0.2251, so in 100 days you would expect there to be 22 with no demands.

(ii) Probability of more than 2 demands = 1 – 0.8088 = 0.1912, so in 100 days you expect there to be 19 with more than 2 demands, meaning that there will be 1 or more demands not met.

Exercise 8B

1 **(i)** For this set of data

$$\sum f = 50, \quad \sum xf = 40, \quad \sum x^2 f = 70$$

$$\Rightarrow \bar{x} = \frac{40}{50} = 0.8, \; s^2 = \frac{70 - 50 \times 0.8^2}{49} = 0.776$$

(ii) It may be assumed that the bacteria occur randomly and independently.

(iii) (***A***) 0.4493 (***B***) 0.0014

2 **(i)** For this set of data

$$\sum f = 10, \quad \sum xf = 23, \quad \sum x^2 f = 73$$

$$\Rightarrow \bar{x} = \frac{23}{10} = 2.3, \; s^2 = \frac{73 - 10 \times 2.3^2}{9} = 2.23$$

(ii) The conditions are that the accidents should be random and independent and it is reasonable to assume that these conditions are satisfied.

(iii) 0.0838
So the answer is no (just!)

3 **(i)**

No.	0	1	2	3	4
Freq.	13	13	8	5	1

(ii) For this set of data
$\sum f = 40, \quad \sum xf = 48, \quad \sum x^2 f = 106$
$\bar{x} = \dfrac{48}{40} = 1.2, \quad s^2 = \dfrac{106 - 40 \times 1.2^2}{39} = 1.24$

(iii) Take $\lambda = 1.2$

Number	0	1	2	3	4
Probs	0.3012	0.6626 – 0.3012 = 0.3614	0.8795 – 0.6626 = 0.2169	0.9662 – 0.8795 = 0.0867	0.9923 – 0.9662 = 0.0261
Exp. freq	12	14	9	3	1

Close agreement - except that expected frequencies do not sum to 40.

(iv) For 2 minutes, $\lambda = 2.4$,
So $P(X = 0) = 0.0907$

4 **(i)** For this set of data
$\bar{x} = 1.4, \quad s^2 = 1.42$

(ii) $\lambda = 1.4 \Rightarrow 0.0537$

5 **(i)** For this set of data
$\bar{x} = 0.2, \quad s^2 = 0.234$
So yes.
The main assumption is that the accidents recorded are random and independent.
This may not be so as accidents may involve more than one person.

(ii) **(*A*)** 0.001

(*B*) 0.8187. So $P(X = 0$ on 8 consecutive weeks$) = 0.8187^8 = 0.2$

6 For this set of data
$\bar{x} = 7, \quad s^2 = 6.92$
Since the mean and the variance are close and on the assumption that the placing of the corpuscles in each compartment was random and independent, a Poisson distribution is a good model.
Expected number: (1 d.p.) 1.4, 3.3, 5.8, 8.2

7 **(i)** For this set of data

$\bar{x} = 51.3, \quad s^2 = 1.168$

The mean is obviously nowhere near the variance so a Poisson distribution is not suitable.

(ii) For this set of data (i.e. subtract 50 from the numbers given)

$\bar{x} = 1.3, \quad s^2 = 1.168$

(N.B. The variance is unchanged, the mean is reduced by 50.)

The mean and the variance are here close in value and so a Poisson model with $\lambda = 1.3$ may be used.

Bus is full if the number of passengers is 50 + 4

So we require the probability that the number of passengers is greater than 4 = 0.0107

Page 14 Discussion point

2 orders could have come from 0 written **and** 2 on-line **or** 1 written **and** 1 on line **or** 2 written **and** 0 on-line.

So to find the probability of 2 orders means adding the three probabilities, each of which are the result of multiplying two probabilities.

Exercise 8C

1 **(i)** 0.0067 **(ii)** 0.0337 **(iii)** 0.0843

2 $\lambda_z = \lambda_x + \lambda_y = 3 + 1.3 = 4.3$

(i) 0.0136

(ii) 0.1663

3 $\lambda_z = \lambda_A + \lambda_B = 1.3 + 1.6 = 2.9$

(i) 0.0287

(ii) The occurrence of A and B are independent and random.

4 **(i)** $\lambda_C = 2.7$; Probability that one cup of tea is sold = 0.1815

$\lambda_T = 1.5$; Probability that one cup of coffee is sold = 0.3347

Probability of one of each = 0.1815 × 0.3347 = 0.0607

(ii) 0.1322 **(iii)** 0.2469

5 **(i)** 0.0996 **(ii)** 5.5 **(iii)** 0.0253

6 **(i)** 0.0302 **(ii)** $P(X > 5) = 0.1424$

7 **(i)** 0.0067 **(ii)** 0.2241 **(iii)** **(A)** 0.0041 **(B)** 0.0884

8 **(i)** Mean = 0.2, Poisson would be a good approximation because n is large, p is small, and claims assumed to be random and independent.

(ii) (***A***) 0.8187 (***B***) 0.1638 (***C***) 0.0175

(iii) 240 000

(iv) (***A***) 0.2466 (***B***) 0.2417

Exercise 8D

1 **(a)** **(i)** 0.0052, 0.1386 **(ii)** 0.0067, 0.1403

(iii) The values are not particularly close - the use of the Poisson distribution to approximate the binomial distribution is not valid in this situation.

(b) **(i)** 0.0066, 0.1402 **(ii)** 0.0067, 0.1403

(iii) The values here are much closer; the must greater value for n means that the approximation is valid.

2 0.0166

3 **(a)** The mean is a reasonable number and it is not possible to determine n or p.

(b) **(i)** 0.5340 **(ii)** 0.1007

(c) $\lambda = 3 \times 2.8 = 8.4$

The probability that there will be fewer than 3 accidents = 0.0323

The probability that there will be more than 15 accidents = 0.0125

4 0.3840

5 **(i)** $p = 0.025$, $n = 300$ gives $np = 7.5$

We may assume that when selected, they will be random and independent.

(ii) 0.6218

6 Random, independent events, n large and p small.

(i) 0.3679 **(ii)** 0.6321 **(iii)** 600 **(iv)** 0.5768

7 **(i)** $\lambda = 3$.

(ii) (***A***) 0.0498 (***B***) 0.1991 (***C***) 0.4232 (***D***) 0.6472

(iii) Probability that up to 6 will fail = 0.9665.

So 5 spares should be carried.

8 **(a)** $\lambda = 2$

(i) 0.1353 **(ii)** 0.5940

(b) $\lambda = 2$ so same answers.

9 **(i)** $\lambda = 100 \times 0.02 = 2$; 0.8647 **(ii)** $\lambda = 200 \times 0.02 = 4$; 0.9817

(iii) We require $P(X \geq 1) = 1 - P(X = 0) \geq 0.99 \Rightarrow P(X = 0) \leq 0.01$

From tables we see that this is when $\lambda = 4.7 = 0.02n \Rightarrow n = 235$