Cambridge University Press
978-1-316-60042-9 — Cambridge International AS and A Level Mathematics: Statistics 2 Coursebook
Steve Dobbs , Jane Miller , Julian Gilbey
Excerpt
More Information

Chapter 1
The Poisson distribution

This chapter introduces a discrete probability distribution which is used for modelling random events. When you have completed it you should

- be able to calculate probabilities for the Poisson distribution
- understand the relevance of the Poisson distribution to the distribution of random events and use the Poisson distribution as a model
- be able to use the result that the mean and variance of a Poisson distribution are equal
- be able to use the Poisson distribution as an approximation to the binomial distribution where appropriate
- be able to use the normal distribution, with a continuity correction, as an approximation to the Poisson distribution where appropriate.

1.1 The Poisson probability formula

Situations often arise where the variable of interest is the number of occurrences of a particular event in a given interval of space or time. An example is given in Table 1.1. This shows the frequency of 0, 1, 2 etc. phone calls arriving at a switchboard in 100 consecutive time intervals of 5 minutes. In this case the 'event' is the arrival of a phone call and the 'given interval' is a time interval of 5 minutes.

Table 1.1 Frequency distribution of number of telephone calls in 5-minute intervals

Number of calls	0	1	2	3	4 or more
Frequency	71	23	4	2	0

Some other examples are

- the number of cars passing a point on a road in a time interval of 1 minute,
- the number of misprints on each page of a book,
- the number of radioactive particles emitted by a radioactive source in a time interval of 1 second.

Further examples can be found in the practical activities in Section 1.4.

The probability distribution which is used to model these situations is called the **Poisson distribution** after the French mathematician and physicist Siméon-Denis Poisson (1781–1840). The distribution is defined by the probability formula

$$\mathrm{P}(X = x) = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

This formula involves the mathematical constant e *which you may have already met in unit P3. If you have not, then it is enough for you to know at this stage that the approximate value of* e *is* 2.718 *and that powers of* e *can be found using your calculator.*

Check that you can use your calculator to show that $\mathrm{e}^{-2} = 0.135\ldots$ *and* $\mathrm{e}^{-0.1} = 0.904\ldots$.

The method by which Poisson arrived at this formula will be outlined in Section 1.2.

This formula involves only one parameter, λ. (λ, pronounced 'lambda', is the Greek letter *l*.) You will see later that λ is the mean of the distribution. The notation for indicating that a random variable X has a Poisson distribution with mean λ is $X \sim \mathrm{Po}(\lambda)$. Once λ is known you can calculate $\mathrm{P}(X = 0)$, $\mathrm{P}(X = 1)$ etc. There is no upper limit on the value of X.

EXAMPLE 1.1.1

The number of particles emitted per second by a radioactive source has a Poisson distribution with mean 5. Calculate the probabilities of

a 0, **b** 1, **c** 2,

d 3 or more emissions in a time interval of 1 second.

a Let X be the random variable 'the number of particles emitted in 1 second'. Then $X \sim \mathrm{Po}(5)$. Using the Poisson probability formula $\mathrm{P}(X = x) = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}$ with $\lambda = 5$, $\mathrm{P}(X = 0) = \mathrm{e}^{-5}\frac{5^0}{0!} = 0.006\,737\ldots = 0.006\,74$, correct to 3 significant figures.

Recall that $0! = 1$ *(see P1 Section 8.3).*

b $\mathrm{P}(X=1)=\mathrm{e}^{-5}\frac{5^1}{1!}=0.03368\ldots=0.0337$, correct to 3 significant figures.

c $\mathrm{P}(X=2)=\mathrm{e}^{-5}\frac{5^2}{2!}=0.084\,22\ldots=0.008\,42$, correct to 3 significant figures.

d Since there is no upper limit on the value of X the probability of 3 or more emissions must be found by subtraction.

$$\begin{aligned}\mathrm{P}(X\geq 3)&=1-\mathrm{P}(X=0)-\mathrm{P}(X=1)-\mathrm{P}(X=2)\\&=1-0.006\,737\ldots-0.033\,68\ldots-0.084\,22\ldots\\&=0.875,\text{ correct to 3 significant figures.}\end{aligned}$$

EXAMPLE 1.1.2

The number of demands for taxis to a taxi firm is Poisson distributed with, on average, four demands every 30 minutes. Find the probabilities of

a no demand in 30 minutes,

b 1 demand in 1 hour,

c fewer than 2 demands in 15 minutes.

a Let X be the random variable 'the number of demands in a 30-minute interval'. Then $X\sim\mathrm{Po}(4)$. Using the Poisson formula with $\lambda=4$,

$\mathrm{P}(X=0)=\mathrm{e}^{-4}\frac{4^0}{0!}=0.0183$, correct to 3 significant figures.

b Let Y be the random variable 'the number of demands in a one-hour interval'. As the time interval being considered has changed from 30 minutes to 1 hour, you must change the value of λ to equal the mean for this new time interval, that is to 8, giving $Y\sim\mathrm{Po}(8)$. Using the Poisson formula with $\lambda=8$,

$\mathrm{P}(Y=1)=\mathrm{e}^{-8}\frac{8^1}{1!}=0.002\,68$, correct to 3 significant figures.

c Again, the time interval has been altered. Now the appropriate value for λ is 2.

Let W be the number of demands in 15 minutes. Then $W\sim\mathrm{Po}(2)$.

$$\mathrm{P}(W<2)=\mathrm{P}(W=0)+\mathrm{P}(W=1)=\mathrm{e}^{-2}\frac{2^0}{0!}+\mathrm{e}^{-2}\frac{2^1}{1!}=0.406,$$

correct to 3 significant figures.

Here is a ***summary*** of the results of this section.

The Poisson distribution **is used as a model for the number, X, of events in a given interval of space or time. It has the probability formula**

$$\mathbf{P}(X=x)=\mathbf{e}^{-\lambda}\frac{\lambda^x}{x!},\quad x=0,1,2,\ldots,$$

where λ is equal to the mean number of events in the given interval.

The notation $X\sim\mathrm{Po}(\lambda)$ indicates that X has a Poisson distribution with mean λ.

Some books use μ rather than λ to denote the parameter of a Poisson distribution. Both alternatives are referred to in the syllabus.

CAMBRIDGE
Cambridge University Press
978-1-316-60042-9 — Cambridge International AS and A Level Mathematics: Statistics 2 Coursebook
Steve Dobbs , Jane Miller , Julian Gilbey
Excerpt
More Information

Exercise 1A

1 The random variable T has a Poisson distribution with mean 3. Calculate

a $P(T = 2)$, **b** $P(T \le 1)$, **c** $P(T \ge 3)$.

2 Given that $U \sim Po(3.25)$, calculate

a $P(U = 3)$, **b** $P(U \le 2)$, c $P(U \ge 2)$.

3 The random variable W has a Poisson distribution with mean 2.4. Calculate

a $P(W \le 3)$, **b** $P(W \ge 2)$, **c** $P(W = 3)$.

4 Accidents on a busy urban road occur at a mean rate of 2 per week. Assuming that the number of accidents per week follows a Poisson distribution, calculate the probability that

a there will be no accidents in a particular week,

b there will be exactly 2 accidents in a particular week,

c there will be fewer than 3 accidents in a given two-week period.

5 On average, 15 customers a minute arrive at the check-outs of a busy supermarket. Assuming that a Poisson distribution is appropriate, calculate

a the probability that no customers arrive at the check-outs in a given 10-second interval,

b the probability that more than 3 customers arrive at the check-outs in a 15-second interval,

6 During April of this year, Malik received 15 telephone calls. Assuming that the number of telephone calls he receives in April of next year follows a Poisson distribution with the same mean number of calls per day, calculate the probability that

a on a given day in April next year he will receive no telephone calls,

b in a given 7-day week next April he will receive more than 3 telephone calls.

7 Assume that cars pass under a bridge at a rate of 100 per hour and that a Poisson distribution is appropriate.

a What is the probability that during a 3-minute period no cars will pass under the bridge?

b What time interval is such that the probability is at least 0.25 that no car will pass under the bridge during that interval?

8 A radioactive source emits particles at an average rate of 1 per second. Assume that the number of emissions follows a Poisson distribution.

a Calculate the probability that 0 or 1 particle will be emitted in 4 seconds.

b The emission rate changes such that the probability of 0 or 1 emission in 4 seconds becomes 0.8. What is the new emission rate?

1.2 Modelling random events

The examples which you have already met in this chapter have assumed that the variable you are dealing with has a Poisson distribution. How can you decide whether the Poisson distribution is a suitable model if you are not told? The answer to this question can be found by considering the way in which the Poisson distribution is related to the binomial distribution in the situation where the number of trials is very large and the probability of success is very small.

Table 1.2 reproduces Table 1.1 giving the frequency distribution of phone calls in 100 5-minute intervals.

Table 1.2 Frequency distribution of number of telephone calls in 5-minute intervals

Number of calls	0	1	2	3	4 or more
Frequency	71	23	4	2	0

If these calls were plotted on a time axis you might see something which looked like Fig. 1.3.

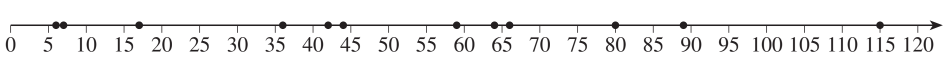

Fig. 1.3 Times of arrival of telephone calls at a switchboard

The time axis has been divided into 5-minute intervals (only 24 are shown) and these intervals can contain 0, 1, 2 etc. phone calls. Suppose now that you assume that the phone calls occur *independently* of each other and *randomly* in time. In order to make the terms in italics clearer, consider the following. Imagine the time axis is divided up into very small intervals of width δt (where δ is used in the same way as it is in pure mathematics). These intervals are so small that they never contain more than one call. If the calls are *random* then the probability that one of these intervals contains a call does not depend on which interval is considered; that is, it is constant. If the calls are *independent* then whether or not one interval contains a call has no effect on whether any other interval contains a call.

Looking at each interval of width δt in turn to see whether it contains a call or not gives a series of trials, each with two possible outcomes. This is just the kind of situation which is described by the binomial distribution (see S1 Chapter 7). These trials also satisfy the conditions for the binomial distribution that they should be independent and have a fixed probability of success.

Suppose that a 5-minute interval contains n intervals of width δt. If there are, on average, λ calls every 5 minutes then the proportion of intervals which contain a call will be equal to $\frac{\lambda}{n}$. The probability, p, that one of these intervals contains a call is therefore equal to $\frac{\lambda}{n}$. Since δt is small, n is large and $\frac{\lambda}{n}$ is small. You can verify from Table 1.2 that the mean number of calls in a 5-minute interval is 0.37 so the distribution of X, the number of calls in a 5-minute interval, is $\mathrm{B}\left(n, \frac{0.37}{n}\right)$.

Finding P(X = 0) Using the binomial probability formula $\mathrm{P}(X = x) = \binom{n}{x} p^x q^{n-x}$, you can calculate, for example, the probability of zero calls in a 5-minute interval as

$$\mathrm{P}(X = 0) = \binom{n}{x}\left(\frac{0.37}{n}\right)^0\left(1 - \frac{0.37}{n}\right)^n.$$

Cambridge University Press
978-1-316-60042-9 — Cambridge International AS and A Level Mathematics: Statistics 2 Coursebook
Steve Dobbs , Jane Miller , Julian Gilbey
Excerpt
More Information

In order to proceed you need a value for n. Recall that n must be large enough to ensure that the δt intervals never contain more than one call. Suppose $n = 1000$. This gives

$$\mathrm{P}(X=0)=\binom{1000}{0}\left(\frac{0.37}{1000}\right)^0\left(1-\frac{0.37}{1000}\right)^{1000}=0.690\,68\ldots.$$

However, even with such a large number of intervals there is still a chance that one of the δt intervals could contain more than one call, so a larger value of n would be better. Try $n = 10\,000$, giving

$$\mathrm{P}(X=0)=\binom{10\,000}{0}\left(\frac{0.37}{10\,000}\right)^0\left(1-\frac{0.37}{10\,000}\right)^{10\,000}=0.690\,72\ldots.$$

Explore for yourself what happens as you increase the value of n still further. You should find that your answers tend towards the value $0.690\,73\ldots$. This is equal to $\mathrm{e}^{-0.37}$, which is the value the Poisson probability formula gives for $\mathrm{P}(X = 0)$ when $\lambda = 0.37$.

This is an example of the general result that $\left(1-\dfrac{x}{n}\right)^n$ *tends to the value* e^{-x} *as n tends to infinity.*

Provided that two events cannot occur simultaneously, allowing n to tend to infinity will ensure that not more than one event can occur in a δt interval.

Finding P(X= 1) In a similar way you can find the probability of one call in a 5-minute interval by starting from the binomial formula and allowing n to increase as follows.

$$\mathrm{P}(X=1)=\binom{n}{1}\left(\frac{0.37}{n}\right)^1\left(1-\frac{0.37}{n}\right)^{n-1}=0.37\left(1-\frac{0.37}{n}\right)^{n-1}.$$

Putting $n = 1000$,

$$\mathrm{P}(X=1)=0.37\left(1-\frac{0.37}{1000}\right)^{999}=0.37\times0.690\,94\ldots=0.255\,64\ldots.$$

Putting $n = 10\,000$,

$$\mathrm{P}(X=1)=0.37\left(1-\frac{0.37}{10\,000}\right)^{9999}=0.37\times0.690\,75\ldots=0.255\,579\ldots.$$

Again, you should find that, as n increases, the probability tends towards the value given by the Poisson probability formula,

$$\mathrm{P}(X=1)=0.37\times\mathrm{e}^{-0.37}=0.255\,57\ldots.$$

Finding P(X= 2), P(X= 3), etc. You could verify for yourself that similar results are obtained when the probabilities of $X = 2$, 3, etc. are calculated by a similar method. A spreadsheet program or a programmable calculator would be helpful.

The general result for $\mathrm{P}(X = x)$ *can be derived as follows. Starting with* $X \sim \mathrm{B}\left(n, \dfrac{\lambda}{n}\right)$,

$$\mathrm{P}(X=x)=\binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}=\frac{n(n-1)(n-2)\ldots(n-x+1)}{x!}\times\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$=\frac{\lambda^x}{x!}\times\frac{n-1}{n}\times\frac{n-2}{n}\times\ldots\times\frac{n-x+1}{n}\times\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

Cambridge University Press
978-1-316-60042-9 — Cambridge International AS and A Level Mathematics: Statistics 2 Coursebook
Steve Dobbs , Jane Miller , Julian Gilbey
Excerpt
More Information

Now consider what happens as n gets larger. The fractions $\frac{n-1}{n}, \frac{n-2}{n}$*, etc. tend towards 1. The term* $\left(1-\frac{\lambda}{n}\right)^{n-x}$ *can be approximated by* $\left(1-\frac{\lambda}{n}\right)^{n}$ *since x, a constant, is negligible compared with n and, as you have seen previously, this tends towards* $e^{-\lambda}$.

Combining these results gives

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

The assumptions made in the derivation above give the conditions that a set of events must satisfy for the Poisson distribution to be a suitable model. They are listed below.

The Poisson distribution is a suitable model for events which

- **occur randomly in space or time,**
- **occur singly, that is events cannot occur simultaneously,**
- **occur independently, and**
- **occur at a constant rate, that is the mean number of events in a given time interval is proportional to the size of the interval.**

EXAMPLE 1.2.1

For each of the following situations state whether the Poisson distribution would provide a suitable model. Give reasons for your answers.

a The number of cars per minute passing under a road bridge between 10 a.m. and 11 a.m. when the traffic is flowing freely.

b The number of cars per minute entering a city-centre car park on a busy Saturday between 9 a.m. and 10 a.m.

c The number of particles emitted per second by a radioactive source.

d The number of currants in buns sold at a particular baker's shop on a particular day.

e The number of blood cells per ml in a dilute solution of blood which has been left standing for 24 hours.

f The number of blood cells per ml in a well-shaken dilute solution of blood.

a The Poisson distribution should be a good model for this situation as the appropriate conditions should be met: since the traffic is flowing freely the cars should pass independently and at random; it is not possible for cars to pass simultaneously; the average rate of traffic flow is likely to be constant over the time interval given.

b The Poisson distribution is unlikely to be a good model: if it is a busy day the cars will be queuing for the car park and so they will not be moving independently.

Cambridge University Press
978-1-316-60042-9 — Cambridge International AS and A Level Mathematics: Statistics 2 Coursebook
Steve Dobbs , Jane Miller , Julian Gilbey
Excerpt
More Information

c The Poisson distribution should be a good model provided that the time period over which the measurements are made is much longer than the lifetime of the source: this will ensure that the average rate at which the particles are emitted is constant. Radioactive particles are emitted independently and at random and, for practical purposes, they can be considered to be emitted singly.

d The Poisson distribution should be a good model provided that the following conditions are met: all the buns are prepared from the same mixture so that the average number of currants per bun is constant; the mixture is well stirred so that the currants are distributed at random; the currants do not stick to each other or touch each other so that they are positioned independently.

e The Poisson distribution will not be a good model because the blood cells will have tended to sink towards the bottom of the solution. Thus the average number of blood cells per ml will be greater at the bottom than the top.

f If the solution has been well shaken the Poisson distribution will be a suitable model. The blood cells will be distributed at random and at a constant average rate. Since the solution is dilute the blood cells will not be touching and so will be positioned independently.

1.3 The variance of a Poisson distribution

In Section 1.2 the Poisson probability formula was deduced from the distribution of $X \sim \mathrm{B}\left(n, \frac{\lambda}{n}\right)$ by considering what happens as n tends to infinity. The variance of a Poisson distribution can be obtained by considering what happens to the variance of the distribution of $X \sim \mathrm{B}\left(n, \frac{\lambda}{n}\right)$ as n gets very large. In S1 Section 8.3 you met the formula $\mathrm{Var}(X) = npq$ for the variance of a binomial distribution. Substituting for p and q gives

$$\mathrm{Var}(X) = n \times \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right) = \lambda\left(1 - \frac{\lambda}{n}\right).$$

As n gets very large the term $\frac{\lambda}{n}$ tends to zero. This gives λ as the variance of the Poisson distribution. Thus the Poisson distribution has the interesting property that its mean and variance are equal.

For a Poisson distribution $X \sim \mathrm{Po}(\lambda)$

mean $= \mu = \mathrm{E}(X) = \lambda$,

variance $= \sigma^2 = \mathrm{Var}(X) = \lambda$.

The mean and variance of a Poisson distribution are equal.

The equality of the mean and variance of a Poisson distribution gives a simple way of testing whether a variable might be modelled by a Poisson distribution. The mean of the data in Table 1.2 has already been used and is equal to 0.37. You can verify that the variance of these data is 0.4331. These values, which are both 0.4 to 1 decimal place, are

Cambridge University Press
978-1-316-60042-9 — Cambridge International AS and A Level Mathematics: Statistics 2 Coursebook
Steve Dobbs , Jane Miller , Julian Gilbey
Excerpt
More Information

sufficiently close to indicate that the Poisson distribution may be a suitable model for the number of phone calls in a 5-minute interval. This is confirmed by Table 1.4, which shows that the relative frequencies calculated from Table 1.2 are close to the theoretical probabilities found by assuming that $X \sim \text{Po}(0.37)$. (The values for the probabilities are given to 3 decimal places and the value for $\text{P}(X \geq 4)$ has been found by subtraction.)

Note that if the mean and variance are not approximately equal then the Poisson distribution is not a suitable model. If they are equal then the Poisson distribution may be a suitable model, but is not necessarily so.

Table 1.4 Comparison of theoretical Poisson probabilities and relative frequencies for the data in Table 1.2

x	Frequency	Relative frequency	$\text{P}(X = x)$
0	71	0.71	$e^{-0.37} = 0.691$
1	23	0.23	$e^{-0.37}0.37 = 0.256$
2	4	0.04	$e^{-0.37}\dfrac{0.37^2}{2!} = 0.047$
3	2	0.02	$e^{-0.37}\dfrac{0.37^3}{3!} = 0.006$
≥ 4	0	0	0
Totals	100	1	1

Exercise 1B

1 For each of the following situations, say whether or not the Poisson distribution might provide a suitable model.

a The number of raindrops that fall onto an area of ground of 1 cm^2 in a period of 1 minute during a shower.

b The number of occupants of vehicles that pass a given point on a busy road in 1 minute.

c The number of flaws in a given length of material of constant width.

d The number of claims made to an insurance company in a month.

2 Weeds grow on a large lawn at an average rate of 5 per square metre. A particular metre square is considered and subdivided into smaller and smaller squares. Copy and complete the table below, assuming that no more than 1 weed can grow in a sub-division.

Number of sub-divisions	P(a sub-division contains a weed)	P(no weeds in a given square metre)
100	$\dfrac{5}{100} = 0.05$	$0.95^{100} = 0.005\,921$
10 000	$\dfrac{5}{10\,000} =$	
1 000 000		
100 000 000		

Compare your answers to the probability of no weeds in a given square metre, given by the Poisson probability formula.

CAMBRIDGE
Cambridge University Press
978-1-316-60042-9 — Cambridge International AS and A Level Mathematics: Statistics 2 Coursebook
Steve Dobbs , Jane Miller , Julian Gilbey
Excerpt
More Information

3 The number of telephone calls I received during the month of March is summarised in the table.

Number of telephone phone calls received per day (x)	0	1	2	3	4
Number of days	9	12	5	4	1

a Calculate the relative frequency for each of $x = 0, 1, 2, 3, 4$.

b Calculate the mean and variance of the distribution. (Give your answers correct to 2 decimal places.) Comment on the suitability of the Poisson distribution as a model for this situation.

c Use the Poisson distribution to calculate $P(X = x)$, for $x = 0, 1, 2, 3$ and ≥ 4 using the mean calculated in part (b).

d Compare the theoretical probabilities and the relative frequencies found in part (a).

Do these figures support the comment made in part (b)?

4 The number of goals scored by a football team during a season gave the following results.

Number of goals per match	0	1	2	3	4	5	6	7
Number of matches	5	19	9	5	2	1	0	1

Calculate the mean and variance of the distribution. Calculate also the relative frequencies and theoretical probabilities for $x = 0, 1, 2, 3, 4, 5, 6, \geq 7$, assuming a Poisson distribution with the same mean. Do you think, in the light of your calculations, that the Poisson distribution provides a suitable model for the number of goals scored per match?

5 The number of cars passing a given point in 100 10-second intervals was observed as follows.

Number of cars	0	1	2	3	4	5
Number of intervals	47	33	16	3	0	1

Do you think that a Poisson distribution is a suitable model for these data?

1.4 Practical activities

1 Traffic flow

In order to carry out this activity you will need to make your observations on a road where the traffic flows freely, preferably away from traffic lights, junctions, etc. The best results will be obtained if the rate of flow is one to two cars per minute on average.

a Count the number of cars which pass each minute over a period of one hour and assemble your results into a frequency table.

b Calculate the mean and variance of the number of cars per minute. Comment on your results.

c Compare the relative frequencies with the Poisson probabilities calculated by taking X equal to the mean of your data. Comment on the agreement between the two sets of values.

