Inverse Poisson problems

In a Poisson distribution, the mean number of occurrences, λ, can be found if $P(X = 0)$ is known, either directly, or in terms of the complementary event $P(X \geq 1)$.

Example

There is a 98% chance that there will be at least one destructive tornado in Townsend, Tennessee during the tornado season. On average, how many destructive tornadoes are there in Townsend during the tornado season?

Solution

If there is a 98% probability of 1 or more tornadoes, there must be a 2% probability of there being 0 tornadoes, i.e. $P(X = 0) = 0.02$

$\frac{e^{-\lambda}\lambda^0}{0!} = 0.02$ — substituting $x = 0$ in $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$e^{-\lambda} = 0.02$ — since $0! = 1$ and $\lambda^0 = 1$

$-\lambda = \ln 0.02$ — taking natural log (ln) of both sides

$\lambda = -\ln 0.02$

$\lambda = 3.912$ (4 d.p.)

Hence there are 3.9 destructive tornadoes on average per tornado season in Townsend.

In other problems you may be required to rearrange the formula $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ in order to find λ or x for a given probability.

Example

An insurance company has found that its policy holders are three times more likely to make one claim per year than they are to make two claims per year. Assuming a Poisson distribution for the number of claims per year, find the standard deviation for the annual number of claims made by this insurance company's policy holders.

Solution

Let X = the number of claims in a year, where X is Poisson with parameter λ.

$P(X = 1) = \frac{e^{-\lambda}\lambda^1}{1!}$ — substituting $x = 1$ in $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$P(X = 2) = \frac{e^{-\lambda}\lambda^2}{2!}$ — Poisson formula with $x = 2$

From the information given:

$\frac{e^{-\lambda}\lambda^1}{1!} = 3 \times \frac{e^{-\lambda}\lambda^2}{2!}$

$1 = \frac{3\lambda}{2}$ — dividing both sides by $\lambda e^{(-\lambda)}$ (also $1! = 1$ and $2! = 2$)

$\lambda = \frac{2}{3}$

So, the standard deviation is:

$\sigma = \sqrt{\frac{2}{3}}$ — standard deviation $= \sqrt{\lambda}$

$= 0.8165$ (4 d.p.)

ESA PUBLICATIONS ISBN 978-1-990015-44-1

Answers p. 114

1. During the school year, the number of phone calls into the school office is found to have a constant average rate, and in 3% of the hours there are no calls. What is the average rate of calls to the school office?

2. Harry the waiter drops plates regularly. His employer has been noting his clumsiness and has discovered that Harry is only able to go through an hour-long period without an accident 32% of the time. Find the average number of dishes Harry drops in one hour.

3. Over summer many orchards employ casual workers to help with picking. These workers however often leave without giving notice. Each day during the season the probability of two or more casual workers leaving without notice is 0.15 and the probability of one casual worker leaving without notice is 0.55. How many casual workers could the orchard expect to leave without notice each day?

4. *1day* customers purchase discounted items on line using their credit cards. *1day* has found that purchases made using stolen cards occur randomly every hour and that 8% of hours have at least one purchase attempted using a stolen card.

   Calculate for *1day* the mean number of purchases attempted per hour using a stolen credit card.

5. The number of patients per hour through an emergency department is known to have a Poisson distribution. Over a long period of time it is found that patients arrive in 80% of all hours.

   a. Calculate the mean number of patients per hour.

   If there are more than eight patients per hour a second nurse is called in.

   b. Calculate the probability that, when a second nurse is called in, there were ten patients in that hour.

 ISBN 978-1-990015-44-1

6. Sinead's problem-solving rate can be modelled by a Poisson distribution. After much study Sinead found that her probability of *not* completing any problems in an hour was 0.00002. What was her average problem solving rate?

7. If there is too much rain during their final growth, apricots will develop brown spots and hence be unfit for sale. After a wet season 68% of apricots were found to have brown spots.

a. Given that the numbers of brown spots on an apricot have a Poisson distribution, find the mean number of brown spots per apricot.

b. If an apricot has more than one brown spot it has to be rejected. How many apricots will be rejected out of a batch of 1 000 apricots?

8. A textile manufacturer estimates there is an average of 2 faults in every 100 metres of fabric produced. What is the maximum number of metres (i.e. complete metres) that you could buy and be at least 90% sure that there would be no faults in the fabric?

9. At a small museum, the number of visitors per hour has a Poisson distribution. The probability of 2 visitors in an hour is double the probability of 3 visitors in an hour.

a. Find the mean and standard deviation of the number of visitors per hour.

b. Find the probability of at least 2 visitors per hour.

ESA PUBLICATIONS ISBN 978-1-990015-44-1

10. There are many wasp nests in the Marlborough region. With a view to establishing an eradication programme, a researcher is analysing wasp nests in a particularly badly affected part of the region. The researcher has divided each part of this area into squares of area 10 m^2, and has counted the number of wasp nests within each square.

The results are as follows:

Number of wasp nests per 10 m^2 (x)	0	1	2	3	4	5	6
Frequency (f)	2	18	48	29	16	4	3

The numbers of wasp nests per 10 m^2 can be modelled using a Poisson distribution.

On repeating his research on a second day, using different sized squares, he calculated a probability of finding no nests in a square as being 0.486. Based on the data in the table above, what area of squares did the researcher use on the second day, in order to have a probability of 0.486 of finding no nests? Justify your answer using mathematical reasoning.

 ISBN 978-1-990015-44-1

Continuous probability distributions

A **continuous random variable**, X, has values which are found by *measuring*, so X can take on any value in a range of values, e.g. X = height (in cm) of a plant. Probabilities for a continuous random variable are worked out for various **intervals** of values, e.g. $P(X \geq 8)$ or $P(5 < X < 9)$.

Note: There are infinitely many values that a continuous random variable can take, so the probability that X equals a single value is zero, e.g. $P(X = 1) = 0$. Hence $P(X > 8)$ is the same as $P(X \geq 8)$; and $P(5 < X < 9) = P(5 \leq X < 9) = P(5 \leq X \leq 9)$ etc.

Since the values that a continuous random variable can take cannot be listed, the probability distribution of a continuous random variable cannot be shown in a table. Instead, probabilities are represented by areas under a curve (whose equation is called the **probability density function**).

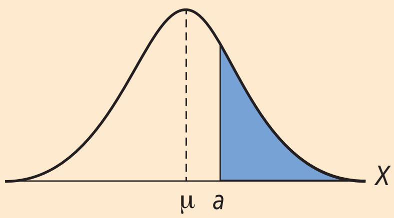
- The total area under the curve is 1 (it is certain that X lies somewhere in the complete interval of values represented).
- The probability that X lies in a given interval is equal to the area below the curve sitting over that interval. (The area above a single value, k, will be zero so $P(X = k) = 0$).

One example of a continuous probability distribution is the **normal distribution.**

Normal distribution

If a random variable has a normal distribution then its probability distribution is a bell-shaped curve defined by two **parameters** – the mean μ and the standard deviation σ.

- The **mean** of the distribution determines the location of the centre of the graph: 50% of the area under the curve is below the mean.
- The **standard deviation** determines the height and width of the graph – the smaller the standard deviation, the taller and narrower the curve.
- The probability that a random variable X is greater than a is equal to the *area under the normal curve* above the interval from a to positive infinity (shown shaded in diagram).



For a normally distributed random variable:

- about 68% of values lie within 1 standard deviation of the mean
- about 95% of values lie within 2 standard deviations of the mean
- about 99.7% of values lie within 3 standard deviations of the mean.

Since each normally distributed random variable X has its own mean, μ, and standard deviation, σ, a **standardised** normal random variable Z is used when calculating probabilities.

ESA PUBLICATIONS ISBN 978-1-990015-44-1

To standardise X the following formula is used.

$$Z = \frac{X - \mu}{\sigma}$$

Each z-value gives the number of standard deviations a given x-value is from its mean.

For example, if X has mean 50 and standard deviation 10, then $X = 65$ converts to $Z = \frac{65 - 50}{10} = 1.5$, i.e. 65 is 1.5 standard deviations above the mean.

A table of probabilities for the **standard normal distribution** can be used to work out probabilities of the form $P(0 < Z < z)$. The symmetry of the graph can be used to determine other probabilities. Alternatively, a graphics calculator may be used.

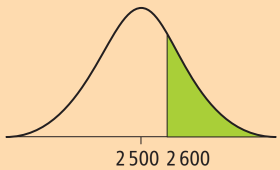
Example

The average annual number of hours of sunshine in Blenheim is 2 500 hours, with a standard deviation of 120 hours.

Find the probability that there will be more than 2 600 sunshine hours in Blenheim in any year.

Solution

Always draw a sketch of the normal curve first and shade the required area.



Let X be the number of hours of sunshine in Blenheim in a randomly selected year.

Require $P(X > 2\,600)$

Using tables:

$P(X > 2600) = P(Z > \frac{2\,600 - 2\,500}{120})$ — standardising using $Z = \frac{X - \mu}{\sigma}$

$= P(Z > 0.8333)$

$= P(Z > 0) - P(0 < Z < 0.8333)$ — splitting into parts to suit tables

$= 0.5 - 0.2975$

$= 0.2025$

Using a graphics calculator (CASIO *CFX*9750G *PLUS*):

To find $P(X > 2600)$ press:

Menu/Stat/Dist(F5)/Norm(F1)/Ncd(F2)/Lower:2600/Upper:9999/σ:120/μ:2500/EXE

to get 0.20232

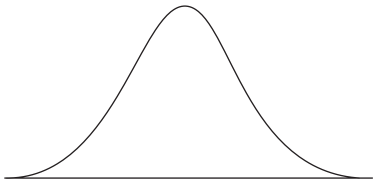
Note: Any large number (at least 4 standard deviations above the mean) can be used for the upper bound.

ESA PUBLICATIONS ISBN 978-1-990015-44-1

It is advisable to draw a sketch and shade appropriately for each problem.

Answers p. 115

1. Spaghetti is sold in tins which are filled by a machine. The actual amount of spaghetti in a tin is normally distributed with a mean weight of 400 g and a standard deviation of 8 g.
Find the probability that, for a randomly selected tin, the weight of spaghetti is:



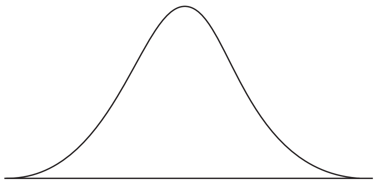
a. less than 390 g

b. between 380 g and 400 g

c. more than 405 g

d. between 410 g and 420 g.

2. The annual earnings of employees in a factory are approximately normally distributed with a mean of \$47 900 and a standard deviation of \$18 000.
What percentage of employees in the factory earn:



a. less than \$45 000?

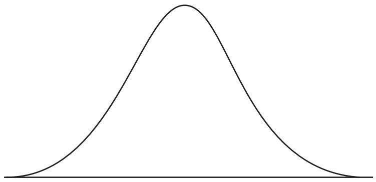
b. between \$45 000 and \$65 000?

c. more than \$70 000?

d. less than \$45 000 or more than \$60 000?

 ISBN 978-1-990015-44-1

3. In 2013, the number of metres run by an All Black in a game was approximately normally distributed with a mean of 432 m, with a standard deviation of 10.5 m.



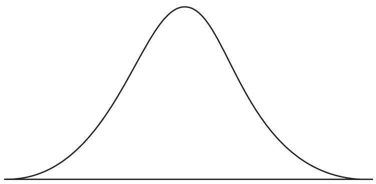
a. What percentage of All Blacks ran more than 450 m in a 2013 game?

b. What is the probability an All Black ran between 425 m and 435 m in a 2013 game?

c. What is the probability an All Black ran less than 430 m during a 2013 game?

d. In a team of 15 All Blacks, how many players would you expect to have run less than 430 m during a 2013 game?

4. The times taken to fit a particular component on a car assembly line are normally distributed with mean 3.75 minutes and standard deviation 0.6 minutes.



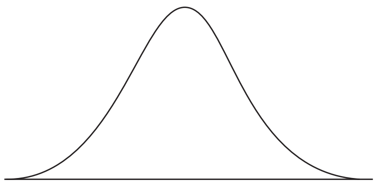
a. Find the probability that it takes more than 3 minutes to fit a car with one of these components.

b. Find the probability that one of these components takes less than $2\frac{1}{2}$ minutes or more than 4 minutes to fit.

c. Three randomly selected components are fitted. Find the probability that all three take at least 3.5 minutes to fit.

ESA PUBLICATIONS ISBN 978-1-990015-44-1

5. The *Mainly Music* company makes compact discs (CDs) of width 10 cm, for CD cases that are 10.4 cm wide. The width of a CD is normally distributed with a standard deviation of 0.3 cm. The company manufactures 5 000 CDs every hour.



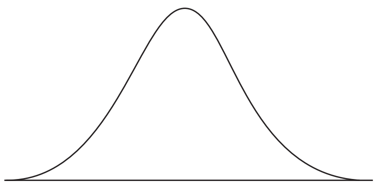
a. What percentage of the CDs would you expect to be more than 10.2 cm wide?

b. In one hour of manufacture, how many CDs would you expect to have a width between 9.9 cm and 10.1 cm?

c. About how many discs per hour will not be able to fit into the cases?

d. The company fine-tunes the machine to reduce the standard deviation by 50%. By what percentage does this reduce the number of CDs per hour that do not fit in the cases? Give answer to nearest whole percentage.

6. The duration of tennis matches at a Wimbledon event can be modelled by the normal distribution, with a mean of 140 minutes and a standard deviation of 15.8 minutes.



a. Calculate the percentage of matches that could be expected to last less than 125 minutes or longer than 150 minutes.

b. Calculate the probability that the next three matches at this Wimbledon event will each last more than $2\frac{1}{2}$ hours.

 ISBN 978-1-990015-44-1

7. The wages of workers in New Zealand are normally distributed with a mean of \$52 000 and standard deviation of \$12 000.

   a. What percentage of people earn more than \$60 000?

   b. What percentage of people earn between \$40 000 and \$55 000?

   c. What percentage of people earn between \$45 000 and \$48 000?

   d. What is the probability that two randomly selected people both earn less than \$30 000?

8. To enter university in America, students can sit two entry tests, the SAT and the ACT. The scores in the SAT are normally distributed with a mean of 500 and a standard deviation of 70, while scores in the ACT have a mean of 18 and a standard deviation of 4.

   a. What is the probability a student sitting the SAT test scores more than 600?

   b. If 10 000 students sit the SAT, how many students would be expected to score more than 650?

   c. A student sits both the SAT and ACT. What is the probability she scores more than 600 in the SAT but less than 15 in the ACT?

   d. Two students sit the SAT. What is the probability they both score under 450?

 ISBN 978-1-990015-44-1

ANSWERS

Discrete probability distributions (page 3)

1. **a.** $E(X) = 2.8$, $Var(X) = 1.16$, $SD(X) = 1.077$
   b. $E(X) = 28$, $Var(X) = 116$, $SD(X) = 10.77$
   c. The values of x in the distribution in **b.** are 10 times the values in the distribution in **a.** So, the expected value and standard deviation in **b.** are 10 times the values of the expected value and standard deviation in **a.** The variance in **b.** is 100 times the variance in **a.** since the variance is a squared value.
2.

x	1	2	3	4
$P(X = x)$	0.25	0.25	0.25	0.25

$E(X) = 2.5$, $Var(X) = 1.25$, $SD(X) = 1.118$

3. Probability distribution is

x	2	4	6	8
$P(X = x)$	0.25	0.25	0.25	0.25

$E(X) = 5$, $Var(X) = 5$, $SD(X) = 2.236$

4. $a = 0.3$, $\mu = E(X) = 2.85$, $Var(X) = 0.9275$, $SD(X) = 0.9631$
5. $b = \frac{1}{8}$, $\mu = E(Y) = 3$, $Var(Y) = 1.25$, $SD(Y) = 1.118$
6. $8c = 1$ so $c = \frac{1}{8}$
   $\mu = E(W) = 2.125$, $Var(W) = 1.109$, $SD(W) = 1.053$
7. $h = 0.2$; $E(T) = 1.7$; $Var(T) = 1.91$; $SD(T) = 1.38$

Experimental probability distributions (page 6)

1. **a.** Lowest value = 1, highest value = 5
   b. Skewed to the right
   c. Mode = 2; uni-modal
   d. 2.7
2. **a.** Lowest value = 4, highest value = 11
   b. Skewed to the left
   c. Mode = 11; uni-modal
   d. 9.71
3. **a.** Lowest value = 0, highest value = 6
   b. Symmetrical
   c. Mode = 3; uni-modal
   d. 3
4. **a.** Lowest value = 0, highest value = 8
   b. Minimal skew
   c. Modes = 0, 4; bi-modal
   d. 3
5. **a.** 9%
   b. 8.61 minutes
   c. 0.8979
   d. 0.9476 minutes (4 s.f.)

Expected value applications (page 10)

1. **a.**

g	45	15	5	–5
$P(G = g)$	0.05	0.05	0.05	0.85

   b. E(gain) = –\$1.00 (the person could expect, on average, to lose \$1.00 per game if a large number of games were played).
   c. 1 000 games
   d. Game is not fair as expected gain is not zero.
   e.

g	$50 - k$	$20 - k$	$10 - k$	$-k$
$P(G = g)$	0.05	0.05	0.05	0.85

$E(G) = 0.05(50 - k) + 0.05(20 - k) + 0.05(10 - k) - 0.85k$
$= 4 - k$

Game fair if $E(G) = 0$ giving $k = 4$

Make the price \$4.00 per game.

2.

t	2	3	4	5	6	7	8	9	10
$P(T = t)$	0.04	0.08	0.12	0.16	0.2	0.16	0.12	0.08	0.04

Expected sum of numbers is 6

3. This is a fair game.
   P(HT or TH) = P(TT or HH) = 0.5
   E(gain) = –0.50 × 0.5 + 0.50 × 0.5 = 0
4. **a.** X = \$ winnings from a single game;
   $P(\text{win}) = P(\text{double}) = \frac{6}{36} = \frac{1}{6}$

x	0	10
$P(X = x)$	$\frac{5}{6}$	$\frac{1}{6}$

   b. $E(X) = 0 \times \frac{5}{6} + 10 \times \frac{1}{6} = 1\frac{2}{3}$ dollars
   Over 9 games expected winnings $= 9 \times 1\frac{2}{3} = \$15$
   c. Expected gain = 15 – 18 = –\$3 (loss of \$3)
   d. Game not fair so change the prize money to \$12
5. **a.** A loss of \$2.50 per game over a large number of games
   b. A profit of \$2.50 per game over a large number of games
   c. A loss of \$50
   d. \$2.50

Expectation algebra (page 13)

1. **a.** Y takes the value 3 whenever X takes the value 1. So, the probability that $Y = 3$ is the same as the probability that $X = 1$.
   b. 10.5 **c.** 26.244 **d.** 5.123 (4 s.f.)
   e. $E(3X) = 3 \times E(X) = 10.5$; $Var(3X) = 9 \times Var(X) = 26.244$; $SD(3X) = 3 \times SD(X) = 5.123$ (4 s.f.)
2. **a.** 8.5 **b.** 2.916
   c. 1.708 (4 s.f.) **d.** 17.5
   e. 26.244 **f.** 5.123 (4 s.f.)
   g. –2.5 **h.** 1.708 (4 s.f.)
3. **a.** \$14.15 **b.** \$3.54
   c. \$17.69 (2 d.p.) **d.** \$4.42
   e. \$16.65 **f.** \$3.54
   g. \$17.48 **h.** \$4.25 (2 d.p.)

Combinations of random variables (page 17)

1. $E(T) = 80$ L, $Var(T) = 0.32$ L, $SD(T) = 0.57$ L
2. $E(T) = 60$ kg, $SD(T) = 1.342$ kg (4 s.f.)
3. $E(T) = 120$ m, $Var(T) = 0.003$ m, $SD(T) = 0.05477$ m (4 s.f.)
4. $E(T) = 100$ kg, $Var(T) = 0.1125$ kg, $SD(T) = 0.3354$ kg (4 s.f.)
5. $E(D) = 35$ g, $SD(D) = 0.3606$ g (4 s.f.)
6. $E(T) = 3\,700$ g, $Var(T) = 113$ g, $SD(T) = 10.63$ g (4 s.f.)
7. $E(T) = 41$ kg, $SD(T) = 0.2417$ kg (4 s.f.)

 ISBN 978-1-990015-44-1