Level 3 – AS91586 – 4 Credits – External

Probability Distributions

Written by J Wills – MathsNZ – jwills@mathsnz.com

Achievement	Achievement with Merit	Achievement with Excellence
Apply probability distributions in solving problems	Apply probability distributions, using relational thinking, in solving problems.	Apply probability distributions, using extended abstract thinking, in solving problems.

Contents

Part 1: Difference between Discrete and Continuous Distributions

Distributions fall into two categories, discrete and continuous. Discrete distributions are where we can put items into groups, e.g.: there is 1 white ball, 2 white balls, 3 white balls, etc. Continuous distributions are when numbers can fall anywhere along a scale, e.g.: the boy could be 153.45cm tall or 185.23cm tall or anywhere in between.

Of the distributions that we look at for 3.14 the **Normal**, **Uniform** and **Triangular** distributions are all continuous and the **Binomial** and **Poisson** are discrete distributions. We also occasionally deal with **Discrete Random Variables** which are also obviously discrete.

Exercise 1

For each of the situations below state if a discrete or continuous distribution would be best to model what is going on.

1. The number of white balls selected from a bag
2. The weight of students
3. The heights of kiwi birds
4. How far students can jump
5. How tall buildings in a city are
6. The number of heads when flipping a coin 5 times in a row
7. The circumference of oranges
8. The weight of apples
9. The number of people ahead of you in a queue when you arrive at the bank
10. The number of students in a class
11. The length of bridges around New Zealand
12. The number of traffic lights in a city
13. The number of phone calls received per hour in a call centre

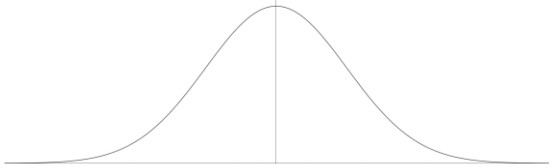
Exercise 1 Answers

1. Discrete
2. Continuous
3. Continuous
4. Continuous
5. Continuous
6. Discrete
7. Continuous
8. Continuous
9. Discrete
10. Discrete
11. Continuous
12. Discrete
13. Discrete

Part 2: The Normal Distribution

In Year 12 maths you should have looked at the normal distribution as part of probability, so we will start with this. The Normal distribution is a continuous distribution that follows what we call a 'bell curve' because the 'probability density function' (a fancy word for the equation that draws the graph, pdf for short) looks like a bell. The equation and the graph for the pdf are shown below.

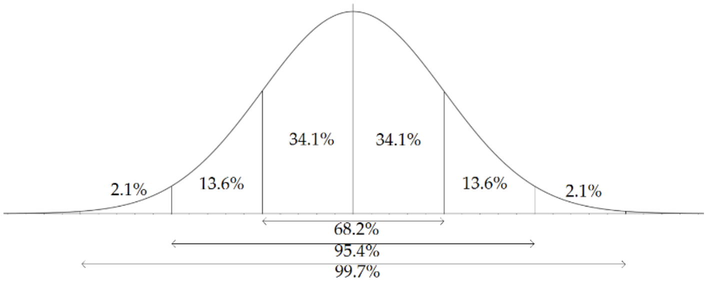
$$f(x,\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Although this formula looks really nasty, we fortunately we do not need to worry about using it as we can use our graphics calculators (or distribution tables) instead.

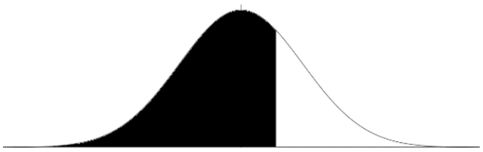
This distribution has two parameters that apply to it (parameters are numbers that tell equation about what it is calculating). The first one is the mean of the population and the second is the standard deviation. The mean is the average of all of whatever we are looking at and is represented by the μ symbol, and the standard deviation is a measure of how spread out it is and is represented by the σ symbol.

This also give us a really cool rule called the '68–95–99.7' rule… 68% of the data is within 1 standard deviation of the mean, 95% is within 2 standard deviations, and 99.7% is within 3 standard deviations as shown in the graphic on the right.



Part 2.1: Up To

The first type of question that we are going to look at is when we want to know the probability that something is less than (or up to) a certain number. We are going to step through an example together using both the graphics calculator and the standard normal tables. The first thing we do is draw a diagram if what we want like the one shown below.



On the graphics calculator you need to go to STAT (2) and then go to DIST (F5) and for this section we want the NORM (F1). This will give us three options: Npd Ncd and InvN. For all the questions where we are trying to find a probability for normal distributions I find it easiest just to use the Ncd (F2). So now… our question.

Example
The weights of Great Spotted kiwi birds are normally distributed with a mean of 2.8 kg and a standard deviation of 0.58 kg. Calculate the probability that a randomly selected kiwi weighs less than 2 kg.

Answer (Graphics Calculator)
As we are after everything **less** than 2kg, our Lower (or minimum amount) is the biggest negative number we can think of… so -99999999999 (the negative button and then a whole bunch of nines). The Upper (or maximum amount) is 2, the σ (or standard deviation) is 0.58 and the μ (or mean) is 2.8.

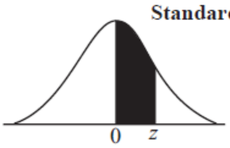
```
Normal C.D
Lower    :-9.999E+11
Upper    :2
σ        :0.58
μ        :2.8
Save Res:None
Execute
```

Pressing the execute button gives us the following screen:

```
Normal C.D
p     =0.08389954
z:Low=-1.724E+12
z:Up =-1.3793103
```

You can ignore the z:Low and z:Up.
So we can see the probability is 0.0839 (3sf) which is our answer.

Standard Normal Distribution



$$\left(Z = \frac{X-\mu}{\sigma}\right)$$

Each entry gives the probability that the standardised normal random variable Z lies between 0 and z.

											Differences								
z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359	4	8	12	16	20	24	28	32	36
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754	4	8	12	16	20	24	28	32	36
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141	4	8	12	15	19	22	27	31	35
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517	4	8	11	15	19	22	26	30	34
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879	4	7	11	14	18	22	25	29	32
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224	3	7	10	14	17	21	24	27	31
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549	3	6	10	13	16	19	23	26	29
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852	3	6	9	12	15	18	21	24	27
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133	3	6	8	11	14	17	19	22	25
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389	3	5	8	10	13	15	18	20	23
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621	2	5	7	9	12	14	16	18	21
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830	2	4	6	8	10	12	14	16	19
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015	2	4	5	7	9	11	13	15	16
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177	2	3	5	6	8	10	11	13	14
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319	1	3	4	6	7	8	10	11	13
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441	1	2	4	5	6	7	8	10	11
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545	1	2	3	4	5	6	7	8	9
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633	1	2	3	3	4	5	6	7	8
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706	1	1	2	3	4	4	5	6	6
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767	1	1	2	2	3	4	4	5	5
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817	0	1	1	2	2	3	3	4	4
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857	0	1	1	2	2	2	3	3	4
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890	0	1	1	1	2	2	2	3	3
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916	0	0	1	1	1	2	2	2	2
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936	0	0	1	1	1	1	1	2	2
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952	0	0	0	1	1	1	1	1	1
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964	0	0	0	0	1	1	1	1	1
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974	0	0	0	0	0	1	1	1	1
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981	0	0	0	0	0	0	0	0	1
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986	0	0	0	0	0	0	0	0	1
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990	0	0	0	0	0	0	0	0	0
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993	0	0	0	0	0	0	0	0	0
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995	0	0	0	0	0	0	0	0	0
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997	0	0	0	0	0	0	0	0	0
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998	.4998	0	0	0	0	0	0	0	0	0
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	0	0	0	0	0	0	0	0	0
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	0	0	0	0	0	0	0	0	0
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	0	0	0	0	0	0	0	0	0
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.5000	.5000	.5000	0	0	0	0	0	0	0	0	0
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	0	0	0	0	0	0	0	0	0

Answer (Tables)
The first thing we need to do is work out the Z value. We do this using the formula $Z = \frac{X-\mu}{\sigma}$.
With our data X=2, μ=2.8, σ=0.58… when we put this into our formula we get:

$$Z = \frac{2 - 2.8}{0.58} = -1.379$$

What we do is we pretend this value is positive and look it up in the table above… we look up the 1.3 in the left column, and then go along to the 7 column giving us .4147… we then look up the 9 in the right group of columns, which is 14. We need to add this onto the last two digits, giving us 0.4147 + 0.0014 which is 0.4161… now what does this number mean? This number is the area under the graph between half way and our point… given that our point is less than the mean, we need to subtract this off 0.5 to find our answer.
Therefore: 0.5 - 0.4161 = 0.0839 which is the same as the answer from the graphics calculator.

Hint: always draw the distribution graph… it helps.

Exercise 2.1

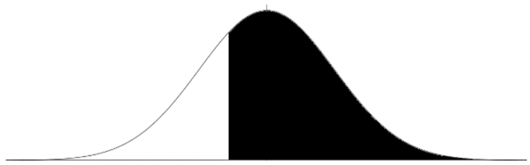
1. The North Island Brown kiwi birds' weights are normally distributed with a mean weight of 2.5 kg and a standard deviation of 0.4 kg. What is the probability that a North Island Brown kiwi will weigh less than 2.7 kg?
2. The length of a pounamu necklaces that have been hand crafted are normally distributed with a mean of 5.3 cm and a standard deviation of 1.1 cm. What is the probability that the pounamu is less than 4.1 cm?
3. The height of women's high heels are normally distributed with a mean of 7.5 cm and a standard deviation of 2.1 cm. What is the probability that the heel is less that 2.9 cm?
4. The amount of milk that a cow produces each day is normally distributed with a mean of 30.3 L and a standard deviation of 3.6 L. What is the probability that the cow produces less than 30 L of milk?
5. The price of bananas in the supermarket is normally distributed with a mean of $2.10 and a standard deviation of $0.40. What is the probability that bananas cost less than $3.10?
6. The weight of a full grown elephant is normally distributed with a mean of 5,200 kg and a standard deviation of 800 kg. What is the probability that the elephant weighs less than 5,000 kg?
7. The length of the leaves on a tree is normally distributed with a mean of 10.4 cm and a standard deviation of 2.1 cm. What is the probability that a leaf is less than 13 cm?
8. The amount of fizzy drink in a bottle is normally distributed with a mean of 1.53 L with a standard deviation of 0.02 L. What is the probability that the bottle has less than 1.50 L of drink in it?

Exercise 2.1 Answers

1. 0.691 (3sf)
2. 0.138 (3sf)
3. 0.0142 (3sf)
4. 0.467 (3sf)
5. 0.994 (3sf)
6. 0.401 (3sf)
7. 0.892 (3sf)
8. 0.0668 (3sf)

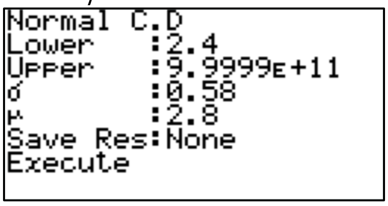
Part 2.2: More Than

In the first section we focused just on finding the probability that something was up to a certain amount. This time we will be looking at the probability that it is more than a certain amount. For these the diagram of what we are trying to find out looks like this one on the right. The method that we use is almost identical to the method we used in the last lot of questions… so an example.



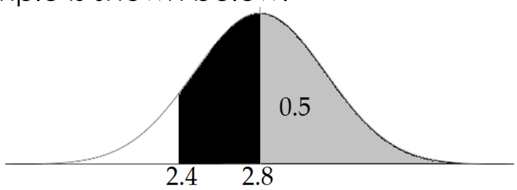
Example
The weights of Great Spotted kiwi birds are normally distributed with a mean of 2.8 kg and a standard deviation of 0.58 kg. Calculate the probability that a randomly selected kiwi weighs more than 2.4 kg.

Answer (Graphics Calculator)
As we are after everything **more** than 2.4kg, our Lower (or minimum amount) is 2.4. The Upper (or maximum amount) is 999999999999 (the nine key pushed a lot of times), the σ (or standard deviation) is 0.58 and the μ (or mean) is 2.8.



Pressing the execute button tells us that the probability is 0.755 (3sf) which is our answer ☺

Answer (Tables)
Again, the first thing we need to do is calculate the Z value. $Z = \frac{2.4-2.8}{0.58} = -0.690$ (3dp)
Again we pretend that the number is positive and look it up in the table. 0.2549… now this is the difference from the middle to the point we want. So we add this onto 0.5 to give us 0.7549 which rounds to 0.755 (3sf) the same as with the graphics calculator.
The sort of diagram that you might use in this example is shown below.



Exercise 2.2

1. The length of phone calls to a call centre are normally distributed with a mean of 2.4 minutes and a standard deviation of 0.8 minutes. What is the probability that a call lasts more than 4 minutes?
2. The cost of a kilo of porter house steak is normally distributed with a mean of \$25 with a standard deviation of \$4.20. What is the probability that the steak costs more than \$32 a kilo?
3. The average size of a 3 bedroom house is normally distributed with a mean of 180 m^2 and a standard deviation of 50 m^2. What is the probability that a 3 bedroom house is more than 200 m^2?
4. The weight lost by a person on a particular diet has a mean of 3.2 kg over 6 months with a standard deviation of 2 kg. What is the probability that someone loses over 5 kg?
5. The size of a hot water bottles are normally distributed with a mean of 1.1 L with a standard deviation of 0.2 L. What is the probability that the hot water bottle can hold more than 0.8 L?
6. The average weight of a new born baby is 3.4 kg with a standard deviation of 0.51 kg. What is the probability that a new born baby weighs more than 4 kg?
7. The flight time from Auckland to Wellington is normally distributed with a mean of 47 minutes and a standard deviation of 5 minutes. What is the probability that the flight lasts for more than an hour?
8. The distance travelled by a commuter in Auckland is normally distributed with a mean of 30 km and a standard deviation of 10 km. What is the probability that a commuter travels more than 43 km each day?

Exercise 2.2 Answers

1. 0.0228 (3sf)
2. 0.0478 (3sf)
3. 0.345 (3sf)
4. 0.184 (3sf)
5. 0.933 (3sf)
6. 0.120 (3sf)
7. 0.00466 (3sf)
8. 0.0968 (3sf)

Part 2.3: Between

The next type of questions we will look at are when we are calculating the probability the value is between two values… this is super easy with the graphics calculator… we just put these numbers into the lower and the upper, with the tables we need to work out two probabilities and then add or subtract them.

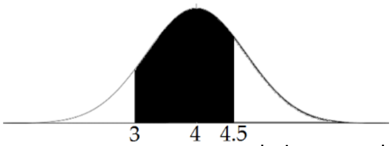
Example
The time taken to brew a coffee is normally distributed with a mean of 4 minutes and a standard deviation of 0.4 minutes. What is the probability it takes between 3 and 4.5 minutes to brew the coffee?

Answer (Graphics Calculator)
You just enter the lower and the upper number in like this:

```
Normal C.D
Lower    :3
Upper    :4.5
σ        :0.4
μ        :4
Save Res:None
Execute
```

And you find the answer is 0.888 (3sf).

Answer (Tables)
The first thing we need to do is draw a diagram like this:



As you can see we need to work out the probability between the lower and the median (3 and 4), and between the median and the upper (4 and 4.5).
First up the lower and the median
$Z = \frac{3-4}{0.4} = -2.5$ giving us p = 0.4938
And the median and the upper
$Z = \frac{4.5-4}{0.4} = 1.25$ giving us p = 0.3944
This gives us a total probability of
0.4938 + 0.3944 = 0.888 (3sf)

Exercise 2.3

1. The gestation period for humans is normally distributed with a mean of 40 weeks and a standard deviation of 1.3 weeks. What is the probability that the gestation period is between 38 and 41 weeks?
2. The height of chicken eggs are normally distributed with a mean of 57 mm and a standard deviation of 4 mm. What is the probability that the chicken egg is between 52 mm and 59 mm tall?
3. The fuel efficiency of new cars in 2008 was normally distributed with a mean of 8.9 L/100 km and a standard deviation of 2.1 L/100 km. What is the probability that the fuel efficiency of a new car was between 5 L/100 km and 7 L/100 km?
4. The average stride length of a marathon runner is normally distributed with a mean of 152 cm and a standard deviation of 20.5 cm. What is the probability that the stride length will be between 150 and 160 cm?
5. The time taken to sell a house is normally distributed with a mean of 34 days and a standard deviation of 10 days. What is the probability that it takes between 20 and 30 days to sell a house?
6. The weight of a forward rugby player is normally distributed with a mean of 111 kg and a standard deviation of 8 kg. What is the probability that a forward rugby player weighs between 120 and 130 kg?
7. The time spent by a person in the shower is normally distributed with a mean of 8 minutes and a standard deviation of 3 minutes. What is the probability that someone spends between 4 and 6 minutes in the shower?
8. The length of cats tails are normally distributed with a mean of 30 cm and a standard deviation of 3 cm. what is the probability that a cat's tail is between 28 cm and 32 cm?

Exercise 2.3 Answers

1. 0.717 (3sf)
2. 0.106 (3sf)
3. 0.151 (3sf)
4. 0.191 (3sf)
5. 0.264 (3sf)
6. 0.122 (3sf)
7. 0.161 (3sf)
8. 0.495 (3sf)

Part 2.4: Inverse Normal

The next thing we need to do is the reverse of what we have done so far… when we are given a probability we need to find the number that it is up to, more than, or between. Let's look at an example.

Example
The time spent by a teacher at work in the holidays is normally distributed with a mean of 20 hours and a standard deviation of 5 hours.

a. What amount of time do the bottom 20% of teachers work?
b. What amount of time do the top 20% of teachers work?
c. What amount of time do the middle 40% of teachers work?

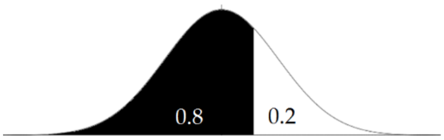
Answer (Graphics Calculator)

a. This time we go to InvN (Inverse Normal) and the tail is to the left (this is the default on the old calculators). The area is 0.2, the mean is 20 and the standard deviation is 5.

```
Inverse Normal
Tail     :Left
Area     :0.2
σ        :5
μ        :20
```

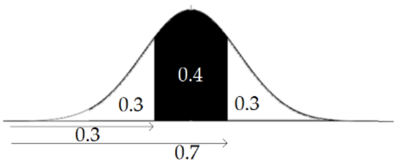
This gives us an answer of 15.8 hours, so in context, the bottom 20% of teachers work less than 15.8 hours.

b. Again we have an inverse normal… if we look at the graph we either want the top 0.2 or the bottom 0.8.



If you have an old calculator you need to put in 0.8, otherwise you can say it is a right tail of 0.2. This gives us an answer of 24.2 which in context means that the top 20% of teachers work more than 24.2 hours in the holidays.

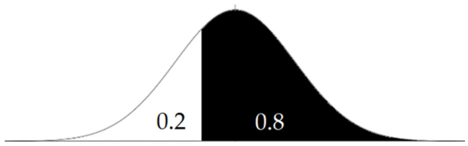
c. On the new calculators this is really easy… you just change the tail to central, otherwise if you draw a diagram you see what is going on and work out the area from the left.



This gives us (17.4, 22.6) which in context means the middle 40% of teachers work between 17.4 hours and 22.6 hours.

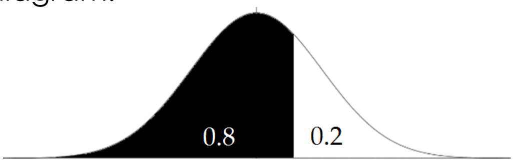
Answer (Tables)

a. The first thing we need to do is draw a diagram.



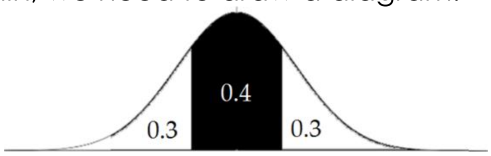
Now, what we need is the area from the middle to the point that we want. In this case the area is 0.3. If we look up 0.3 in the body of the table we can see that the Z score that relates to it is 0.841 as 0.2996 + 3 is the closest we can get to 0.3… now obviously the point that we want is below the mean so the Z score is -0.841. If we put this into our formula we get $-0.841 = \frac{X-20}{5}$
If we solve this we find out that X = 15.8, which in context means the bottom 20% of teachers work less than 15.8 hours in the holidays.

b. Again, the first thing we need to do is draw a diagram.



Again, the area between the mean and the point we want is 0.3, giving us a Z score of 0.841, and because this value is more than the mean, we leave it positive this time. If we put this into our formula we get
$0.841 = \frac{X-20}{5}$
If we solve this we find out that X = 24.2, which in context means the top 20% of teachers work more than 24.2 hours in the holidays.

c. Again, we need to draw a diagram.



With this we can see that the area between the mean and the point that we want is 0.2 in each direction. If we look up 0.2 in the tables we find that the Z score that it lines up with is 0.524 (0.1985+14 is as close to 0.2 as we can get).

We need to solve this again using the same method as before, but this time with **both** the positive and negative numbers.
$0.524 = \frac{X-20}{5}$ and $-0.524 = \frac{X-20}{5}$

Which gives us 22.6 and 17.4 which in context means the middle 40% of teachers work between 17.4 hours and 22.6 hours in the holidays.

Exercise 2.4

1. The time spent in freefall in when skydiving is normally distributed with a mean of 20 seconds and a standard deviation of 2 seconds. What times do the middle 50% of jumpers freefall for?
2. The width of an A4 piece of paper is normally distributed with a mean of 210 mm and a standard deviation of 0.2 mm. Over what size are the top 20% of A4 pieces of paper?
3. The length of a football field is normally distributed with a mean of 100 m and a standard deviation of 3 m. Under what size are the smallest 10% of football fields?
4. The time taken to boil a kettle is normally distributed with a mean of 94 seconds and a standard deviation of 6 seconds. What times do the middle 40% of boils fall between?
5. The temperature in a glass house is normally distributed with a mean of 30 °C and a standard deviation of 3 °C. Under what temperature are the coldest 5% of days?
6. The weight of a block of butter is normally distributed with a mean of 510 g and a standard deviation of 4 g. Over what weight are the top 4% of blocks of butter?
7. The height of rugby players is normally distributed with a mean of 1.86 m and a standard deviation of 0.08 m. Between what weights are the middle 80% of rugby players?
8. The length of an elephant's trunk is normally distributed with a mean of 175 cm and a standard deviation of 10 cm. What length are the shortest 10% of elephant trunks under?

Exercise 2.4 Answers

1. Between 18.7 and 21.3 seconds (3sf)
2. Over 210.17mm (5sf)
3. Under 96.2m (3sf)
4. Between 90.9 and 97.1 seconds (3sf)
5. Under 25.1 °C (3sf)
6. Over 517g (3sf)
7. Between 1.76m and 1.96m (3sf)
8. Under 162cm (3sf)

Part 2.5: Finding the Mean and Standard Deviation

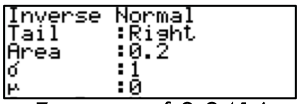
For Merit and Excellence we are sometimes required to find the mean and or the standard deviation based on other pieces of information that is given to us. To do this we need to use the formula that is given at the top of the formula sheet: $Z = \frac{X-\mu}{\sigma}$. Sometimes this will require using simultaneous equations, and sometimes it will just involve a straight use of the formula. Let's look at some examples…

Example 1 – Finding the standard deviation

The amount of time that it takes to change a nappy is normally distributed with a mean of 40 seconds. The worst 20% of nappies take more than 60 seconds to change. Calculate the standard deviation for the time it takes to change a nappy.

Answer (Graphics Calculator)

The first thing we need to do is calculate the Z-score. To do this we are using a 'standard normal' curve… this means for the moment the mean is zero and the standard deviation is one, and we know the right tail is 0.2. We then get:



This gives us a Z score of 0.8416.

We then use the formula $Z = \frac{X-\mu}{\sigma}$ and substitute in the values… $0.8416 = \frac{60-40}{\sigma}$ and solve it to find that σ = 23.8 minutes.

Answer (Tables)

The first thing we need to do is work out the Z-score using the tables just like we did in the last section. The difference between the mean and the point we are after is 0.3 (draw a diagram to help) which when we look it up in the tables we get a Z-score of 0.841. We then use the formula $Z = \frac{X-\mu}{\sigma}$ and substitute in the values… $0.841 = \frac{60-40}{\sigma}$ and solve it to find that σ = 23.8 minutes.

Example 2 – Finding the mean

The marks in a test are normally distributed with a standard deviation of 20%. If the top 10% of students score more than 90%, what is the mean test mark?

Answer (Graphics Calculator)

Again, the first thing we need to do is calculate the Z-score, we do this using the same process as before… Tail: right, Area: 0.1, σ: 1, μ: 0.

This gives us a Z-score of 1.282.

We then substitute in the values into our formula… $1.282 = \frac{90-\mu}{20}$ and find that μ = 64.4%.

Answer (Tables)

We again use the same process, finding a z-score of 1.282. We put this into the formula $Z = \frac{X-\mu}{\sigma}$ and substitute in the values… $1.282 = \frac{90-\mu}{20}$ and solve it to find that μ = 64.4%.

Example 3 – Finding both the mean and the standard deviation.

The heights of Great Spotted Kiwi Birds are normally distributed. The shortest 10% are under 44 cm tall and the tallest 20% are more than 48 cm tall. Calculate the mean and standard deviation of the kiwi's heights.

Answer

In this case because we need to find both the mean and standard deviation, we need to use simultaneous equations. We find the Z score for each part and the form the following equations.

$$-1.281 = \frac{44-\mu}{\sigma} \text{ and } 0.842 = \frac{48-\mu}{\sigma}$$

If we rearrange these to get

$$-1.281\sigma = 44-\mu \text{ and } 0.842\sigma = 48-\mu$$

We can then equate them to:

$$-1.281\sigma - 44 = 0.842\sigma - 48$$

Which we can solve to find σ = 1.884 and then substitute this in to find μ = 46.4.

Exercise 2.5

1. The time an office worker spends at work is normally distributed with a mean of 8 hours. On the busiest 10% of days he spends more than 9 hours at work. Calculate the standard deviation.
2. The time a marathon runner takes to complete a marathon is normally distributed with a mean of 4 hours. The fastest 10% of runners finish in under 2.5 hours. Calculate the standard deviation for the time to complete a marathon.
3. The weight of sugar in a bag is normally distributed with a standard deviation of 4g. If the lightest 5% of bags are less than 500 g, what is the mean weight of sugar in the bag?
4. The time spent watching television in a week is normally distributed with a standard deviation of 2 hours. If the top 5% of people spend more than 10 hours watching TV a week, what is the mean time spent watching TV a week?
5. The lengths of boats in a marina are normally distributed with a mean of 5 m. If the shortest 10% of boats are less than 2 m what is the standard deviation for the length of the boats?
6. The length of a foot long sub is normally distributed. If the shortest 10% of subs are less than 30 cm and the longest 20% are more than 31 cm, what is the mean and standard deviation for the lengths of foot-long subs?
7. The height of 5 year olds are normally distributed. The shortest 20% of 5 year olds are under 100 cm and the tallest 30% are more than 130cm. Calculate the mean and standard deviation for the height of five year olds.
8. The weight of an unladen swallow is normally distributed. The lightest 30% are less than 13.5 g and the heaviest 40% are more than 14.5g. Calculate the mean and standard deviation for the weight of unladen swallows.

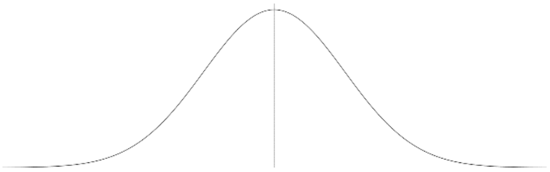
Exercise 2.5 Answers

1. 0.780 hours (3sf)
2. 1.17 hours (3sf)
3. 507g (3sf)
4. 6.71 hours (3sf)
5. 2.34m (3sf)
6. $\sigma = 0.471$cm, $\mu = 30.6$cm (3sf)
7. $\sigma = 22.0$cm, $\mu = 118$cm (3sf)
8. $\sigma = 1.29$g, $\mu = 14.2$g (3sf)

Part 2.6: Assumptions

Occasionally we get asked to compare a graph to a normal distribution, or to fit a normal distribution to a set of data. In this case there are just a couple of things that we need to be seeing happening, or assume they are happening.

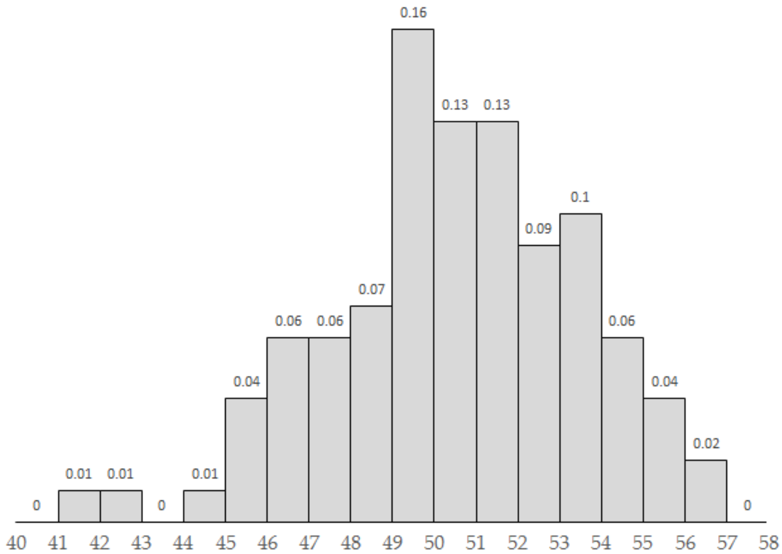
As we can see from the normal distribution graph on the left it is **symmetrical**. It is also **unimodal**. If the data roughly fits both of these things then using a normal distribution is good, otherwise you might want to use a different type of distribution, or at least comment on the differences, depending on the question.



Part 2.7: Mixed Questions

1. The marks in an exam were normally distributed with a mean of 50 and a standard deviation of 13.
    a. What is the probability that a student scored more than 70%?
    b. What is the probability that a student scored between 40% and 60%?
    c. In order to be in the top 10% of students, what mark did you need to score?
2. The height of desks in a classroom is normally distributed with a standard deviation of 2 cm.
    a. If the mean is 70 cm, what is the probability that the height is more than 73 cm?
    b. If the mean is 68 cm, what is the probability that the height is between 67 cm and 68 cm?
    c. If the tallest 20% of desks are more than 72 cm, what was the mean height of the desks?
3. Bumble bees' weights are normally distributed with a mean of 220mg and a standard deviation of 40mg.
    a. What is the probability that a bumble bee weighs less than 150 mg?
    b. What is the probability that a bumble bee weighs between 180 mg and 200 mg?
    c. Over what weight are the heaviest 20% of bumble bees?
4. The time spent on the computer in a certain household is normally distributed.
    a. In one house the mean time is 4 hours with a standard deviation of 1 hour. What is the probability that in a given day the amount of time spent is less than 2.5 hours?
    b. In another house the mean time is 8 hours with a standard deviation of 2.4 hours. What is the probability that in a given day the amount of time spent is between 4 and 5 hours?
    c. In another house in the 10% of days when they use the computer least they spend less than 2 hours on the computer, and the mean amount of time that they spend on the computer is 3 hours. What is the standard deviation for the amount of time that they spend on the computer?
5. The gestation period for dogs are normally distributed.
    a. For one breed the mean time is 62 days with a standard deviation of 1.2 days. What is the probability that the gestation period is less than 60 days?
    b. For another breed the mean time is 61 days with a standard deviation of 0.8 days. What is the probability that the gestation period is between 59 and 63 days?
    c. Another breed the longest 15% of pregnancies are more than 63 days, and the shortest 12% of pregnancies are less than 59 days. Calculate the mean and the standard deviation for this breed.
6. The distance that a jogger runs when doing his training is normally distributed with a mean of 5.3 km.
    a. If the standard deviation is 1.2 km what is the probability that the jogger runs more than 3.5 km?
    b. If the standard deviation is 0.8 km what is the probability that the jogger runs more than 3.5 km?
    c. What is the standard deviation if 10% of the time the jogger runs more than 8 km?
7. The gaps between the draws in a particular brand (Alpha) of desk where a chair is designed to go are normally distributed with a mean of 46 cm and a standard deviation of 1 cm. The gaps between the draws in another brand (Beta) are also normally distributed with a mean of 47 cm. The Beta brand desks are expected to have a gap under 45 cm 13% of the time.
    a. How often would you expect Alpha brand desks to have a gap under 45 cm?
    b. How often would you expect Alpha brand desks to have a gap between 47 cm and 48 cm?
    c. Explain how the standard deviation for the gap between the draws for Beta brand desks compares with the Alpha brand desks.

8. The popliteal length (the length from the ground to the back of the knee when standing) was recorded for a large number of boys. The mean for this data was 50.0 cm and the standard deviation was 2.9 cm. This is shown in the histogram below.



Explain whether a normal distribution would be an appropriate model for the distribution of popliteal length for boys. As part of your explanation, describe the features of the distribution and include at least one calculation.

Part 2.7 Answers

1a. 0.0620 (3sf)
1b. 0.558 (3sf)
1c. 66.7% (3sf)

2a. 0.0668 (3sf)
2b. 0.191 (3sf)
2c. 70.3 cm (3sf)

3a. 0.0401 (3sf)
3b. 0.150 (3sf)
3c. 254mg (3sf)

4a. 0.0668 (3sf)
4b. 0.0579 (3sf)
4c. 0.780 (3sf)

5a. 0.0478 (3sf)
5b. 0.988 (3sf)
5c. σ = 1.81 days, μ = 61.1 days (3sf)

6a. 0.933 (3sf)
6b. 0.988 (3sf)
6c. 2.11 (3sf)

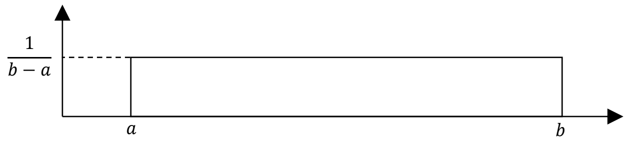
7a. 0.159 (3sf)
7b. 0.134 (3sf)
7c. σ for Beta = 1.78cm (3sf). This means there will be more variation in the gaps for Beta desks than for alpha desks.

8. Normal distributions should be continuous, bell-shaped, unimodal and symmetrical. Looking at this data we can see the data is reasonably symmetrical, with just a couple of smaller points off to the left, which may be due to those boys being younger, and it is reasonably bell-shaped, with the highest bar being between 49 and 50, where the mean is 50, and very clearly unimodal. If it was normally distributed then the probability of a popliteal length being over 52 cm would be 0.245 (3sf) whereas from the histogram we can see it is 0.09+0.1+0.06+0.04+0.02 = 0.31, which is a little bit higher, than the 0.245, but you would expect some variation. Therefore a normal distribution is appropriate for the popliteal length for boys.

Part 3: The Continuous Uniform Distribution

The uniform distribution is another continuous distribution, and is sometimes referred to as the rectangular distribution as it forms a rectangular shape when you draw it out. The probability density function (or pdf for short) is shown below, both as a formula and a diagram. The minimum is 'a' and the maximum is 'b'.

$$f(x) = \begin{cases} \dfrac{1}{b-a}, a < x < b \\ 0, \text{elsewhere} \end{cases}$$



Like with the normal distribution the area under the 'curve' is the probability, so the full rectangle will always have an area of one, as probabilities always add to one. We normally use the uniform distribution when we are only given two pieces of information, the minimum and the maximum.

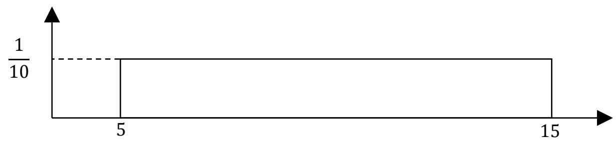
Part 3.1: Finding the Probability

The first thing we need to know how to do is find the probability from a uniform distribution. To do this we find the area of the relevant part of the rectangle. Let's have a look at an example.

Example
The time taken for a component to wear out has a uniform distribution with a minimum of 5 days and a maximum of 15 days.

a. What is the probability that it wears out in less than 7 days?
b. What is the probability that it lasts longer than 10 days?
c. What is the probability that it wears out between 5 and 8 days?

Answer
The first thing that we must do is draw the probability function, as this makes life much easier. Shading in the area that you are trying to find is also helpful.



a. The probability that it is less than 7 days is the probability that it is between 5 days and 7 days. Now we know that the area of a rectangle, and thus the probability, is base × height, so it is $2 \times \frac{1}{10} = 0.2$.
b. The probability that it takes more than 10 days to fail is the probability it is between 10 and 15 days, this means the area, and therefore the probability, is $5 \times \frac{1}{10} = 0.5$.
c. The difference between these two days is 3, so the probability is $3 \times \frac{1}{10} = 0.3$.

Exercise 3.1

1. The time taken to tie a knot has a uniform distribution. The minimum time is 4 seconds and the maximum time is 8 seconds. What is the probability that it:
   a. Takes less than 5 seconds to tie?
   b. More than 6 seconds to tie?
   c. Between 4.5 and 6.4 seconds to tie?
   d. Longer than 10 seconds to tie?
2. The amount of grain that spills from a bag has a uniform distribution with a minimum of 2 g and a maximum of 12 g. What is the probability that:
   a. More than 10 g spill out?
   b. Less than 2 g spill out?
   c. Between 3 and 5 grams spill out?
   d. Less than 2.5 g spill out?
3. The length of pieces of wood have a uniform distribution with a minimum of 3 m and a maximum of 3.2 m. What is the probability the wood is:
   a. More than 3.18 m long?
   b. Less than 3.04 m long?
   c. Between 3.02 m and 3.05 m long?
   d. Between 3.21 m and 3.3 m long?
4. A farmer knows that the heaviest his cows get is 790 kg and the lightest is 660 kg. What is the probability the cow weighs:
   a. More than 750 kg?
   b. Between 700 kg and 750 kg?
   c. More than 300 kg?
   d. Less than 200 kg?
5. The amount spend on my phone bill a month has a minimum of \$80.00 and a maximum of \$93.00. Find the probability the phone bill is:
   a. Between \$80.50 and \$90.50?
   b. Less than \$82.50?
   c. Between \$92.00 and \$94.00?
6. The maximum length of a pen is 12.52 cm and the minimum length is 12.41 cm. Using a normal distribution find the probability the pen is:
   a. More than 12.49 cm?
   b. Less than 12.43 cm?
   c. Between 12.42 and 12.45 cm?
   d. More than 12.6 cm?

7. The maximum distance a rugby player can normally kick a ball is 49 m and the minimum is 10 m. What is the probability the kick is:
    a. More than 10 m?
    b. Less than 40 m?
    c. Between 30 m and 40 m?
    d. More than 36 m?

8. The minimum time I spend each day reading is 10 minutes and the maximum is 3 hours. What is the probability that I spend:
    a. Between 2 and 2.5 hours reading?
    b. More than 1 hour reading?
    c. Less than 40 minutes reading?

Exercise 3.1 Answers

1a. 0.25
1b. 0.5
1c. 0.475
1d. 0
2a. 0.2
2b. 0
2c. 0.2
2d. 0.05
3a. 0.1
3b. 0.2
3c. 0.15
3d. 0
4a. 0.308 (3sf)
4b. 0.385 (3sf)
4c. 1
4d. 0
5a. 0.769 (3sf)
5b. 0.192 (3sf)
5c. 0.0769 (3sf)
6a. 0.273 (3sf)
6b. 0.182 (3sf)
6c. 0.273 (3sf)
6d. 0
7a. 1
7b. 0.769 (3sf)
7c. 0.256 (3sf)
7d. 0.333 (3sf)
8a. 0.176 (3sf)
8b. 0.706 (3sf)
8c. 0.176 (3sf)

Part 3.2: Mean and Standard Deviation

Occasionally we need to find the mean and the standard deviation for a continuous uniform distribution. The formulae for this are not given to you, but are pretty easy to remember. They are:

Mean = $\frac{1}{2}(a+b)$ Standard deviation = $\frac{1}{\sqrt{12}}(b-a)$

Example
The time taken for a component to wear out has a uniform distribution with a minimum of 5 days and a maximum of 15 days. Calculate the mean and the standard deviation.

Answer
Mean = $\frac{1}{2}(a+b) = \frac{1}{2}(5+15) = 10$ and standard deviation = $\frac{1}{\sqrt{12}}(b-a) = \frac{1}{\sqrt{12}}(15-5) = 2.89\ (3sf)$

Exercise 3.2

For each of the following, assuming it is uniform, calculate the mean and the standard deviation.

1. The time taken to tie a knot has a uniform distribution. The minimum time is 4 seconds and the maximum time is 8 seconds.
2. The amount of grain that spills from a bag has a uniform distribution with a minimum of 2 g and a maximum of 12 g.
3. The length of pieces of wood have a uniform distribution with a minimum of 3 m and a maximum of 3.2 m.
4. A farmer knows that the heaviest his cows get is 790 kg and the lightest is 660 kg.
5. The amount spend on my phone bill a month has a minimum of \$80.00 and a maximum of \$93.00.
6. The maximum length of a pen is 12.52 cm and the minimum length is 12.41 cm.
7. The maximum distance a rugby player can normally kick a ball is 49 m and the minimum is 10 m.
8. The minimum time I spend each day reading is 10 minutes and the maximum is 3 hours.

Exercise 3.2 Answers

1. σ = 1.15 s, μ = 6 s (3sf)
2. σ = 2.89 g, μ = 7 g (3sf)
3. σ = 0.0577 m, μ = 3.1 m (3sf)
4. σ = 37.5 kg, μ = 725 kg (3sf)
5. σ = \$3.75, μ = \$86.50 (3sf)
6. σ = 0.0318 cm, μ = 12.5 cm (3sf)
7. σ = 11.3 m, μ = 29.5 m (3sf)
8. σ = 49.1 min, μ = 95 min

Part 3.3: Assumptions

Occasionally we get asked to compare a graph to a uniform distribution, or to fit a uniform distribution to a set of data. In this case there are just a couple of things that we need to be seeing happening, or assume they are happening.

If we are only given the **minimum** and the **maximum** then a uniform distribution is the best distribution to use, but it does assume that the probability is **equally likely** along the entire width of the distribution, and it is **impossible** for it to be outside this range.

Part 3.4: Mixed Questions

1. An ice cream machine is set to pour out ice cream into cups. It is controlled to randomly cut the flow of ice cream between 200 g and 206 g. Find the probability it pours out:
    a. Less than 201 g.
    b. Between 202 g and 205 g.
    c. More than 205.5 g.
2. The maximum weight that can be sent using a standard parcel post bag is 3kg. Assuming the weight of the bags has a uniform distribution:
    a. Justify why a uniform distribution is best to use in this situation and state the value of the parameters (a and b).
    b. What is the probability it weighs more than 2.9 kg?
    c. What is the probability it weighs less than 1.8 kg?
3. The amount of time it takes for a grenade to go off it anywhere between 6 seconds and 8 seconds.
    a. Justify why a uniform distribution is best to use in this situation and state the value of the parameters (a and b).
    b. What is the probability it takes between 6.2 and 6.6 seconds to go off?
    c. What is the probability it goes off in under 5 seconds?
4. The cost to fill my car with petrol is between \$40 and \$90 depending on how long between visits to the petrol station.
    a. Justify why a uniform distribution is best to use in this situation and state the value of the parameters (a and b).
    b. What is the probability it costs more than \$70 to fill my car?
    c. Between what two values do the middle 90% of fill ups cost?
5. The northern explorer train departs Auckland at 7:50 am and arrives in Wellington between 6:20 pm and 6:30 pm.
    a. Justify what distribution would be best to model the time taken to travel between Auckland and Wellington and state the value of the parameters (a and b).
    b. What is the probability it takes more than 10 hours and 5 minutes to complete the journey?
    c. What is the probability it takes less than 9 hours to complete the journey?
6. I always drink more than 1 L of water a day and always drink less than 2.1 L.
    a. Justify what distribution would be best to model the amount of water I drink in a day and state the value of the parameters (a and b).
    b. What is the probability I drink more than 2 L of water in a day?
    c. Between what two values do I drink on the middle 90% of days?
7. The diagonal measurement of a ‘standard’ 40 inch TV is anywhere between 101.2 cm and 102.1 cm.
    a. Justify what distribution would be best to model the diagonal measurement of a ‘standard’ 40 inch TV and state the parameters (a and b).
    b. What is the probability it is actually more than 40 inches (101.6 cm)?
    c. Under what length are the smallest 10% of ‘standard’ 40 inch TVs?

Part 3.4 Answers

1a. 0.167 (3sf)
1b. 0.5
1c. 0.0833 (3sf)

2a. We only have two pieces of information, the maximum weight (b) is 3kg, and the minimum weight (a) would be 0kg. Therefore a uniform distribution is best.
2b. 0.0333 (3sf)
2c. 0.6

3a. We only have two pieces of information, the maximum time (b) is 8 seconds, and the minimum time (a) would be 6 seconds. Therefore a uniform distribution is best.
3b. 0.2
3c. 0

4a. We only have two pieces of information, the maximum amount (b) is \$90, and the minimum amount (a) would be \$40. Therefore a uniform distribution is best.
4b. 0.4
4c. \$42.50 and \$87.50

5a. We only have two pieces of information, the maximum time (b) is 10 hours and 40 minutes, and the minimum time (a) would be 10 hours and 30 minutes. Therefore a uniform distribution is best.
5b. 1
5c. 0

6a. We only have two pieces of information, the maximum amount (b) is 2.1L, and the minimum amount (a) would be 1L. Therefore a uniform distribution is best.
6b. 0.0909 (3sf)
6c. 1.055 and 2.045L

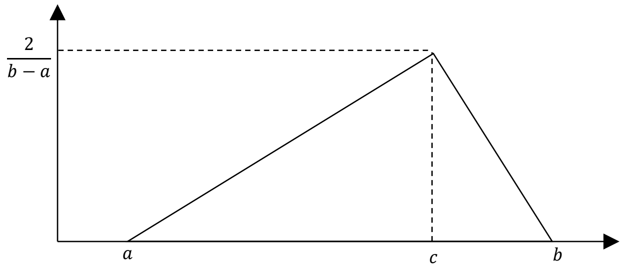
7a. We only have two pieces of information, the maximum diagonal measurement (b) is 102.1, and the diagonal measurement (a) would be 101.2. Therefore a uniform distribution is best.
7b. 0.556 (3sf)
7c. 101.29cm

Part 4: The Triangular Distribution

The big advantage of the triangular distribution over the uniform distribution is that it incorporates a modal value, or outcome that is thought to be most likely to occur. The triangular distribution is widely used in business and project management, and is often referred to as 'three-point estimation' as it is based off three values, a minimum, a maximum, and a mode. Because it is based just off the three values it is sometimes also called the 'lack of knowledge' distribution, as it is often used when the cost of collecting data is high, so we only know these three values, and often in business the modal value is only an estimate.

The pdf for the triangular distribution is shown below, both as a function and a graph. The minimum value is 'a' the maximum value is 'b' and the modal value, or the point most likely to occur, is 'c'.

$$f(x) = \begin{cases} 0, & x < a \\ \dfrac{2(x-a)}{(b-a)(c-a)} & a \le x \le c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & c \le x \le b \\ 0, & x > b \end{cases}$$



Part 4.1: The Left Slope

The first thing we need to know, is that just like the uniform distribution, the probability we are trying to find is always the area under the graph. The formula you will need to know for the area of a triangle is

$$\text{Area} = \frac{1}{2} \times \text{base } \times \text{height}$$

In the pdf formula above there are 4 parts. The first part '0' means the probability when x is less than the minimum is just zero. The part we are going to be focusing on in this section is the part when x is between a and c, this is the left slope and the height is given by the formula: $\frac{2(x-a)}{(b-a)(c-a)}$. To find the area up to a point on this slope you work out the area of the triangle up to this point. Let's look at an example.

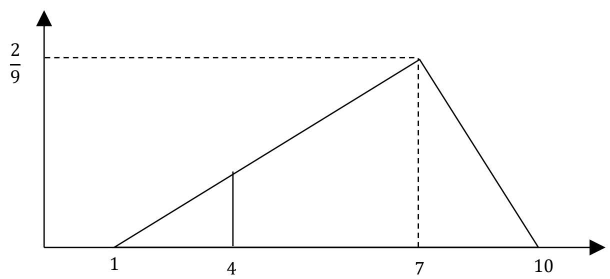
Example

A manager knows the total sales of a product are going to be between \$1 million and \$10 million, and the most likely amount of sales will be \$7 million.

a. What is the probability the sales are less than \$4 million?
b. What is the probability the sales are more than \$4 million?

Answer

The first thing we need to do is draw a diagram and fill in the values.
a = 1 (minimum)
b = 10 (maximum)
c = 7 (most likely)
x = 4 (the point we are looking at)



a. The area of a triangle is $\frac{1}{2} \times \text{base } \times \text{height}$
The base is the difference between 1 and 4 which is 3
Height = $\frac{2(x-a)}{(b-a)(c-a)} = \frac{2(4-1)}{(10-1)(7-1)} = \frac{6}{54} = \frac{1}{9}$
So probability = $\frac{1}{2} \times 3 \times \frac{1}{9} = \frac{1}{6}$
b. The probability that it is **more** than 4 is just 1 minus the probability that it is **less** than 4 so this is $1 - \frac{1}{6} = \frac{5}{6}$

Exercise 4.1

1. The minimum amount of time spent changing a tire is 1 minute and the maximum is 10 minutes. The most likely time is 3 minutes. What is the probability it takes:
    a. Less than 2 minutes to change?
    b. More than 2 minutes to change?
    c. Less than 1.5 minutes to change?
    d. More than 3 minutes to change?
2. The minimum amount a company will spend on advertising is \$100,000 and the maximum is \$800,000. They think they are most likely to spend \$600,000. What is the probability they spend:
    a. Less than \$150,000?
    b. More than \$200,000?
    c. Less than \$300,000?
    d. More than \$500,000?
3. A company knows that the time spent on a project will be between 300 and 800 hours, with the expected time being 500 hours. What is the probability the project takes:
    a. Less than 350 hours?
    b. More than 500 hours?
    c. Less than 450 hours?
    d. More than 200 hours?
4. The time the cookies spent in the oven is between 20 and 22 minutes with the most likely time being 21.5 minutes. What is the probability they spend:
    a. Less than 21 minutes in the oven?
    b. More than 20.5 minutes in the oven?
    c. More than 21.5 minutes in the oven?
    d. Less than 19 minutes in the oven?

Exercise 4.1 Answers

1a. 0.0556 (3sf)	2a. 0.00714 (3sf)	3a. 0.025	4a. 0.333 (3sf)
1b. 0.944 (3sf)	2b. 0.971 (3sf)	3b. 0.6	4b. 0.917 (3sf)
1c. 0.0139 (3sf)	2c. 0.114 (3sf)	3c. 0.225	4c. 0.25
1d. 0.778 (3sf)	2d. 0.543 (3sf)	3d. 1	4d. 0

Part 4.2: The Right Slope

The right slope is very similar to the left slope, but it is the part coming down on the right hand side. The formula for this section is $\frac{2(b-x)}{(b-a)(b-c)}$. Let's look at another example:

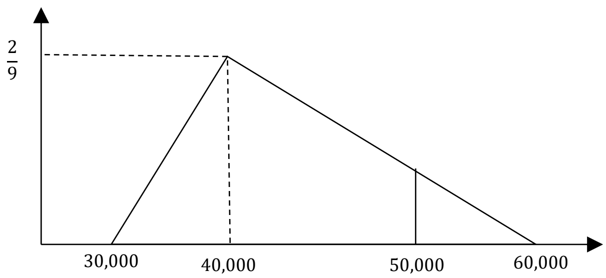
Example

A company knows that the cost of a project is going to be between \$30,000 and \$60,000 with the most likely cost being \$40,000.

a. What is the probability that the project costs more than \$50,000?
b. What is the probability that the project costs less than \$50,000?

Answer

The first thing we need to do is draw a diagram and fill in the values.
a = 30,000 (minimum)
b = 60,000 (maximum)
c = 40,000 (most likely)
x = 50,000 (the point we are looking at)



a. The area of a triangle is $\frac{1}{2} \times \text{base} \times \text{height}$
The base is the difference between 50,000 and 60,000 which is 10,000
Height = $\frac{2(b-x)}{(b-a)(b-c)}$
$= \frac{2(60{,}000-50{,}000)}{(60{,}000-30{,}000)(60{,}000-40{,}000)}$
$= \frac{20{,}000}{600{,}000{,}000} = \frac{1}{30{,}000}$
So probability = $\frac{1}{2} \times 10{,}000 \times \frac{1}{30{,}000} = \frac{1}{6}$

b. The probability that it is **less** than \$50,000 is just 1 minus the probability that it is **more** than \$50,000 so this is $1 - \frac{1}{6} = \frac{5}{6}$

Exercise 4.2

1. The time taken for the kettle to boil is between 1.5 minutes and 3 minutes. The most likely time is 2 minutes. What is the probability the kettle takes:
   a. More than 2.5 minutes to boil?
   b. Less than 2.5 minutes to boil?
   c. More than 2 minutes to boil?
   d. More than 2 minutes 15 seconds to boil?
2. A business knows that it will have a tax bill between \$3 million and \$8 million next year, with the most likely tax bill being \$4 million. What is the probability the tax bill will be:
   a. More than \$5 million?
   b. Less than \$6 million?
   c. More than \$7.5 million?
   d. Less than \$7 million?
3. The minimum time it takes to milk a cow is 4 minutes and the maximum time is 6 minutes. The most common time is 4.5 minutes. What is the probability the milking takes:
   a. More than 5 minutes?
   b. Less than 5.5 minutes?
   c. More than 4.5 minutes?
   d. Less than 6 minutes?
4. The amount of time that employees at a business spend in meetings a year is between 300 and 600 hours, with the most common time of 400 hours. What is the probability employees spend:
   a. More than 500 hours in meetings?
   b. Less than 550 hours in meetings?
   c. More than 600 hours in meetings?
   d. Less than 400 hours in meetings?

Exercise 4.2 Answers

1a. 0.167 (3sf)	2a. 0.45	3a. 0.333 (3sf)	4a. 0.167 (3sf)
1b. 0.833 (3sf)	2b. 0.8	3b. 0.917 (3sf)	4b. 0.958 (3sf)
1c. 0.667 (3sf)	2c. 0. 0125	3c. 0.75	4c. 0
1d. 0.375	2d. 0.95	3d. 1	4d. 0.333 (3sf)

Part 4.3: Combining the Two

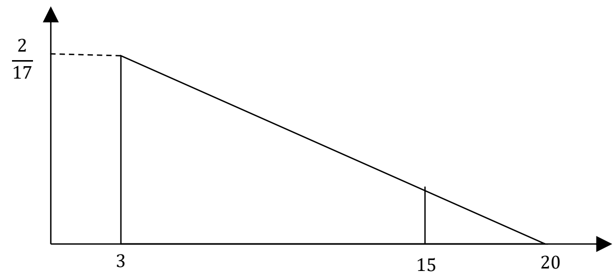
Sometimes the triangles we are using look slightly different (e.g.: a right angle triangle) or the probability we are after spans both the left and the right slopes of the triangle. Here are some examples for you to look at:

Example 1
The amount of time it takes to clean a bathroom has a triangular distribution with a minimum time of 3 minutes and a maximum time of 20 minutes. The most likely amount of time it takes is 3 minutes.

a. What is the probability it takes more than 15 minutes to clean?
b. What is the probability it takes between 10 and 15 minutes to clean?

Answer
The first thing we need to do is draw a diagram and fill in the values.
a = 3 (minimum)
b = 20 (maximum)
c = 3 (most likely)
x = 15, and then 10 (the points we are looking at)

We have a right angled triangle and the left part of the triangle is non-existent… so that means that we just need to look at the right part of the triangle.

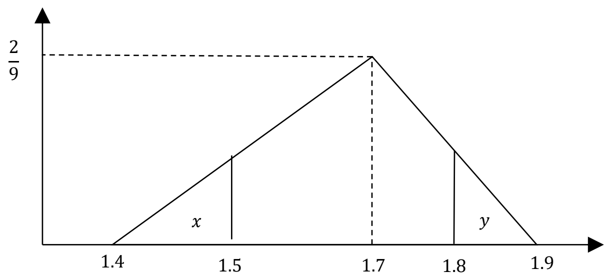


a. The area of a triangle is $\frac{1}{2} \times \text{base} \times \text{height}$.
The base is the difference between 15 and 20 which is 5.
Height = $\frac{2(b-x)}{(b-a)(b-c)} = \frac{2(20-15)}{(20-3)(20-3)} = \frac{10}{289}$
So probability = $\frac{1}{2} \times 5 \times \frac{10}{289} = \frac{25}{289}$

b. To find this we find the difference in the two triangles… the one from 10 to 20 and the one we just worked out from 15 to 20.
Height = $\frac{2(b-x)}{(b-a)(b-c)} = \frac{2(20-10)}{(20-3)(20-3)} = \frac{20}{289}$
So probability = $\frac{1}{2} \times 10 \times \frac{20}{289} = \frac{100}{289}$
Therefore the probability that we are interested in is $\frac{100}{289} - \frac{25}{289} = \frac{75}{289}$.

Example 2
The height of a bookcase is somewhere between 1.4 m and 1.9 m with the most common height being 1.7 m. What is the probability the bookcase is between 1.5 m and 1.8 m tall?

Answer
Again, the first thing we need to do is draw a diagram and fill in the values.
a = 1.4 (minimum)
b = 1.9 (maximum)
c = 1.7 (most likely)
x = 1.5 and 1.8 (the points we are looking at)



This time we need to look at the two triangles. What is probably easiest is to work out the probability of the parts we do not want (marked x and y), and then subtract them from one.

Area x:
Height = $\frac{2(x-a)}{(b-a)(c-a)} = \frac{2(1.5-1.4)}{(1.9-1.4)(1.7-1.4)} = \frac{4}{3}$
Probability = $\frac{1}{2} \times 0.1 \times \frac{4}{3} = \frac{1}{15}$

Area y:
Height = $\frac{2(b-x)}{(b-a)(b-c)} = \frac{2(1.9-1.8)}{(1.9-1.4)(1.9-1.7)} = 2$
Probability = $\frac{1}{2} \times 0.1 \times 2 = \frac{1}{10}$

Therefore the probability the height is between 1.5 m and 1.8 m tall is $1 - \frac{1}{15} - \frac{1}{10} = \frac{5}{6}$

Exercise 4.3

1. The gap between the seats on Trans-Tasman flights is always between 30 cm and 60 cm with the most common gap being 40 cm. What is the probability the gap is:
    a. More than 50 cm?
    b. Between 45 cm and 50 cm?
    c. Between 35 cm and 45 cm?
2. The weight of a pillow is known to be somewhere between 100 g and 250 g, with the most likely weight being 250 g. What is the probability the weight is:
    a. Less than 200 g?
    b. Between 150 and 200 g?
3. I know the amount of water in a cup is between 100 mL and 250 mL, but I think there is 220 mL in the glass. What is the probability the amount of water in the cup is:
    a. Less than 150 mL?
    b. Between 150 mL and 200 mL?
    c. Between 200 mL and 240 mL?
4. The weight of a suitcase must be less than 32 kg. A suitcase is never less than 5 kg. The most common weight of suitcase is 15 kg. What is the probability the suitcase weights:
    a. Less than 10 kg?
    b. Between 4 kg and 10 kg?
    c. Less than the limit for extra baggage charge of 20 kg?
5. The thickness of a magazine is between 3 mm and 6 mm with the most common thickness being 3 mm. What is the probability that the magazine is:
    a. Less than 4 mm thick?
    b. Between 4 mm and 5 mm thick?
    c. More than 5 mm thick?
6. My watch is always within 1 minute of the school bell (in either direction), and it is normally exactly correct. What is the probability my watch is:
    a. Less than 15 seconds out?
    b. Less than 30 seconds out?
7. The amount spent on weddings in New Zealand is normally between \$5,000 and \$60,000 with the most common amount spent being \$28,000. What is the probability that a couple spends:
    a. Less than \$10,000?
    b. Between \$20,000 and \$30,000?
8. The time taken for a person to get ready in the morning is known to be between 10 minutes and 1 hour. Most mornings it takes about 20 minutes. What is the probability it takes:
    a. Less than 15 minutes?
    b. Between 15 minutes and 30 minutes?
    c. Between 30 minutes and 45 minutes?

Exercise 4.3 Answers

1a. 0.167 (3sf)
1b. 0.208 (3sf)
1c. 0.542 (3sf)
2a. 0.444 (3sf)
2b. 0.333 (3sf)
3a. 0.139 (3sf)
3b. 0.417 (3sf)
3c. 0.422 (3sf)
4a. 0.0926 (3sf)
4b. 0.0926 (3sf)
4c. 0.686 (3sf)
5a. 0.556 (3sf)
5b. 0.333 (3sf)
5c. 0.111 (3sf)
6a. 0.4375
6b. 0.75
7a. 0.0198 (3sf)
7b. 0.311 (3sf)
8a. 0.05
8b. 0.5
8c. 0.3375

Part 4.4: Mean and Standard Deviation

Again, just like with the rectangular distribution, sometimes it is helpful to work out the mean and the standard deviation for the triangular distribution.

The formula for the mean is $\frac{a+b+c}{3}$ and the formula for the standard deviation is $\sqrt{\frac{a^2+b^2+c^2-ab-ac-bc}{18}}$.

Example
The time it takes to fly from the Gold Coast to Auckland is always between 2.5 and 3.5 hours, with the normal time being 2.7 hours. What is the mean and standard deviation of the trip time?

Answer
The first thing we need to do is identify a, b and c… a = 2.5, b = 3.5 and c = 2.7. We then just substitute into the formula.
Mean = $\frac{a+b+c}{3} = \frac{2.5+3.5+2.7}{3} = 2.9$ hours.
Standard deviation = $\sqrt{\frac{a^2+b^2+c^2-ab-ac-bc}{18}} = \sqrt{\frac{2.5^2+3.5^2+2.7^2-2.5\times3.5-2.5\times2.7-3.5\times2.7}{18}} = 0.216$ hours (3sf).

Exercise 4.4

The following questions require you to calculate the mean and the standard deviation.

1. A company knows that the time spent on a project will be between 300 and 800 hours, with the expected time being 500 hours. Calculate the mean and the standard deviation for the time spent on a project.
2. The time the cookies spent in the oven is between 20 and 22 minutes with the most likely time being 21.5 minutes. Calculate the mean and the standard deviation of the time the cookies spend in the oven.
3. The minimum time it takes to milk a cow is 4 minutes and the maximum time is 6 minutes. The most common time is 4.5 minutes. Calculate the mean and the standard deviation of the amount of time it takes to milk a cow.
4. A business knows that it will have a tax bill between \$3 million and \$8 million next year, with the most likely tax bill being \$4 million. Calculate the mean and the standard deviation of the tax bill.

The following questions require you to work backwards to find either the minimum, maximum or most common value.

5. The amount spent on advertising is somewhere between \$1 million and \$10 million. The mean amount spent is \$6 million. Calculate the amount that is most likely to be spent, and then the standard deviation for the amount being spent.
6. The weight of lollies taken from the pick and mix on average is 600 grams with a maximum of 810 grams. The most likely amount to be taken is 680 grams. What is the minimum amount that is taken and the standard deviation for the amount taken?
7. The minimum amount of time taken for a project is 24 days, and the most likely time is 50 days. The average time is 47 days. Calculate the maximum amount of time the project is likely to take, and then the standard deviation of the time of the project.
8. The amount of water in a 100 L tank is always between 0 L and 100 L. What is the mean amount of water in the tank if the standard deviation for the amount of water is 21 L?
Hint: work out the mode first.

Exercise 4.4 Answers

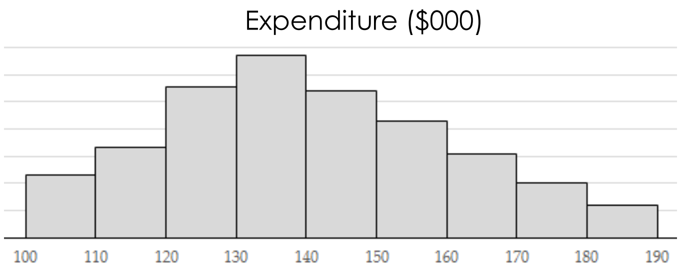
1. σ = 103 hours, μ = 533 hours (3sf)
2. σ = 0.425 min, μ = 21.2 min (3sf)
3. σ = 0.425 min, μ = 4.83 min (3sf)
4. σ = \$1.08 mil (3sf), μ = \$5 mil
5. mode = \$7 mil, σ =\$1.87 mil (3sf)
6. min = 310 g, σ =106 g (3sf)
7. max = 67 days, σ =8.84 days (3sf)
8. mode = 70.9 L (3sf), μ = 57.0 L (3sf)

Part 4.5: Assumptions

With the triangular distribution there are some assumptions that we make. We use it when we know 3 values, a minimum, a maximum and a value that is most likely to occur. It is also really useful when we have a skewed dataset, as a normal distribution is not very good to use when we are looking at a skewed dataset.

Part 4.6: Mixed Questions

1. A company expects its sales to be between \$20 million and \$60 million, but it expects them to be approximately \$30 million.
    a. Choose a distribution to model this situation and justify your choice of distribution.
    b. Find the probability the sales will be less than \$25 million.
    c. Find the probability the sales will be between \$25 million and \$45 million.
2. The length of a rugby practice is always between 40 minutes and 1 hour. Most of the time the practices are approximately 55 minutes long.
    a. Choose a distribution to model this situation and justify your choice of distribution.
    b. Find the probability the practice will be less than 50 minutes.
    c. Find the probability the practice will be more than 45 minutes.
3. The battery life of an iPad is advertised to be 10 hours, and from my experience it always lasts more than 8 hours and less than 13 hours, depending on what I am doing.
    a. Choose a distribution to model this situation and justify your choice of distribution.
    b. What is the probability the battery lasts more than 11 hours?
    c. What is the probability the battery lasts between 9 and 11 hours?
4. The number of calories I consume in a day is normally between 1100 and 2500. The standard adult diet is approximately 2000 calories.
    a. Choose a distribution to model this situation and justify your choice of distribution.
    b. Find the probability I consume less than 1500 calories.
    c. Find the probability I consume more than 2300 calories.
    d. What are the mean and the standard deviation for this distribution?
    e. In the 10% of days that I eat the most, how many calories do I consume?
5. The cost of chicken breast at the supermarket is always between \$10 and \$25 a kilo, but is most often \$19.50 a kilo.
    a. Choose a distribution to model this situation and justify your choice of distribution.
    b. What is the probability the price is more than \$21 a kilo?
    c. What is the probability the price is between \$15 and \$18 a kilo?
    d. On the cheapest 20% of days, what price is it under?
6. The length of a pencil before it is sharpened for the first time is 19 cm and it becomes unusable when it is less than 8 cm. The normal length of a pencil is approximately 15 cm.
    a. Choose a distribution to model this situation and justify your choice of distribution.
    b. A pencil becomes difficult to use if it is less than 10 cm long. What is the probability a pencil is difficult to use?
    c. What range of lengths are the pencils between 90% of the time?
7. Looking at the graph below for a company's expenditure, choose a distribution to model this situation and justify your choice of distribution. The expenditure is in thousand dollar amounts.



Part 4.6 Answers

1a. We are given three pieces of information, the minimum sales (a) is \$20 million, and the maximum sales (b) would be \$60 million, and the most likely amount (c) is \$30 million. Therefore a triangular distribution is best.
1b. 0.0625
1c. 0.75

2a. We are given three pieces of information, the minimum length (a) is 40 minutes, and the maximum time (b) would be 60 minutes, and the most likely time (c) is 55 minutes. Therefore a triangular distribution is best.
2b. 0.333 (3sf)
2c. 0.917 (3sf)

3a. We are given three pieces of information, the minimum length (a) is 8 hours, and the maximum time (b) would be 13 hours, and the most likely time (c) is 10 hours. Therefore a triangular distribution is best.
3b. 0.267 (3sf)
3c. 0.733 (3sf)

4a. We are given three pieces of information, the minimum calories (a) is 1100, and the maximum calories (b) would be 2500, and the most likely calories (c) is 2000. Therefore a triangular distribution is best.
4b. 0.127 (3sf)
4c. 0.0571 (3sf)
4d. $\sigma = 290$, $\mu = 1870$
4e. 2240 calories (3sf)

5a. We are given three pieces of information, the minimum price (a) is \$10, and the maximum price (b) would be \$25, and the most likely price (c) is \$19.50. Therefore a triangular distribution is best.
5b. 0.194 (3sf)
5c. 0.274 (3sf)
5d. \$15.40 (3sf)

6a. We are given three pieces of information, the minimum length (a) is 8cm, and the maximum length (b) would be 19cm, and the most likely length (c) is 15cm. Therefore a triangular distribution is best.
6b. 0.0519 (3sf)
6c. 9.96 cm and 17.5cm (3sf)

7. The data is continuous, has a very distinct minimum, and is definitely skewed to the right. Because the data is continuous a uniform, triangular or normal distribution would be best. Because the data is skewed to the right, and has a very clear peak, a triangular distribution would be best. The minimum value (a) would be \$100,000, the maximum value (b) would be \$190,000 and the modal or most likely value (c) would be \$135,000.

Part 5: Continuity Corrections

All of the distributions that we have done so far have been for continuous data, however often times we want to model discrete data with a continuous distribution. For example we might know the number of units a company sells of a particular product is between a and b, but we don't have any more information than this. We will look at situations where this happens for the uniform, triangular and normal distributions, and then mix it up a bit at the end with a combination of the distributions where you have to choose which distribution you think is best.

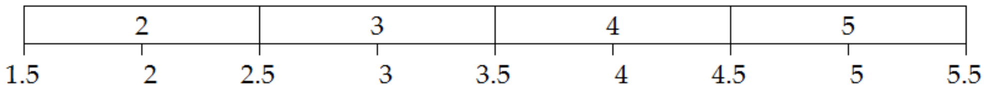
We need to think about what the number is rounded to, for 'number of' type questions it is obviously rounded to the nearest whole number, but sometimes things are measured to the nearest 0.1 m for example. If this is the case then it will tell you what it has been measured to, so you'll know you need to use a continuity correction, but for 'number of' you'll need to always assume it has been rounded to the nearest whole number.

Part 5.1: Uniform Distributions

Let's look at an example for the uniform distribution.

Example
The number of people at a business on a given day is known to be between 2 and 5 people. What is the probability that there will be more than 3 people at the business?

Answer
The first thing we need to do is think about the numbers… the smallest number that rounds up to 2 is 1.5 and the largest number that rounds down to 5 is 5.5… if we were to draw a diagram of what this looked like it would look a bit like this:



You can see that 2 goes from 1.5 to 2.5, 3 goes from 2.5 to 3.5 and so on. This means a = 1.5 and b = 5.5.
Therefore if we want to know the probability there are more than 3 people at the business you need to know from the area of the rectangle from 3.5 (more than 3) to 5.5.
Height = $\frac{1}{b-a} = \frac{1}{5.5-1.5} = \frac{1}{4}$
Probability = base × height = $2 \times \frac{1}{4} = \frac{1}{2}$

Exercise 5.1

1. The number of students in a class is between 25 and 32. What is the probability there are:
    a. Less than 28 students in the class?
    b. More than 30 students in the class?
    c. Exactly 26 students in the class?
2. The height of students is measured to the nearest 0.1 m. The tallest student was 2.1 m and the shortest was 1.5 m. Based on this what is the probability a student is measured to be:
    a. More than 1.7 m tall?
    b. Less than 2 m tall?
    c. Measured to be 1.7 m tall?
3. The weight of vegetables I buy from the supermarket is between 400 g and 500 g measured to the nearest 5 grams. What is the probability the weight is:
    a. More than 450 g?
    b. Less than 470 g?
    c. 400 g or less?
4. The number of pens in my pocket is always less than 10. What is the probability I have:
    a. 1 pen in my pocket?
    b. More than 5 pens in my pocket?
    c. No pens in my pocket?

Exercise 5.1 Answers

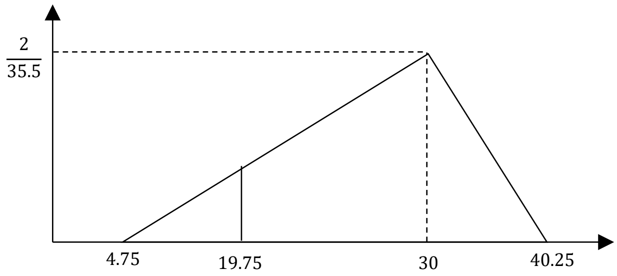
1a. 0.375	2a. 0.571 (3sf)	3a. 0.476 (3sf)	4a. 0.1
1b. 0.25	2b. 0.714 (3sf)	3b. 0.667 (3sf)	4b. 0.4
1c. 0.125	2c. 0.143 (3sf)	3c. 0.0476 (3sf)	4c. 0.1

Part 5.2: Triangular Distribution

Exactly the same principle applies to the triangular distribution. Let's look at an example…

Example
The amount of petrol in my tank is always between 5 L and 40 L. It normally has approximately 30 L in it. I only know the amount of petrol to the nearest 0.5 L. What is the probability there is less than 20 L in my tank?

Answer
Again, we need to think about the largest number that rounds to 5 L and the smallest number that rounds to 40 L… in this case a = 4.75 and b = 40.25. Because our modal value is in the middle we don't need to round this, it will still stay as 30. The probability there is less than 20 L in the tank means that the dial would need to display 19.5 L, so any numbers smaller than 19.75 would display this, so x = 19.75. If we draw a diagram we get:



Doing the calculations height = $\frac{2(19.75-4.75)}{(40.25-4.75)(30-4.75)} = \frac{240}{7171}$

Probability = $\frac{1}{2}$ × base × height = $\frac{1}{2} \times 15 \times \frac{240}{7171} = 0.251\ (3sf)$

Exercise 5.2

1. The number of biscuits in a jumbo packet is always between 35 and 42, but it is most likely to be 38. What is the probability there are:
   a. Less than 37 biscuits?
   b. More than 39 biscuits?
   c. Between 37 and 39 biscuits inclusive?
   d. Exactly 38 biscuits?
2. The amount of water in a glass is between 120 mL and 160 mL measured to the nearest mL. Based on looking at it I think there is 130 mL in it. What is the probability there is:
   a. More than 150 mL in the glass?
   b. Less than 125 mL in the glass?
   c. Between 125 mL and 150 mL in the glass inclusive?
   d. An amount of water that measures to be 125 mL?
3. The number of pages in a legal document is known to be between 10 and 20, but is most likely to be 13. Calculate the probability there are:
   a. More than 18 pages in the document.
   b. Less than 12 pages in the document.
   c. Less than 10 pages in the document.
   d. Between 13 and 18 pages inclusive.
4. The number of pearls on a necklace is somewhere between 45 and 60, the most common number of pearls is 50. What is the probability there are:
   a. Less than 50 pearls on the necklace?
   b. More than 50 pearls on the necklace?
   c. Exactly 50 pearls on the necklace?
   d. Between 53 and 55 pearls inclusive on the necklace?

Exercise 5.2 Answers

1a. 0.143 (3sf)	2a. 0.0800 (3sf)	3a. 0.0485 (3sf)	4a. 0.284 (3sf)
1b. 0.25	2b. 0.0581 (3sf)	3b. 0.104 (3sf)	4b. 0.595 (3sf)
1c. 0.607 (3sf)	2c. 0.862 (3sf)	3c. 0	4c. 0.121 (3sf)
1d. 0.234 (3sf)	2d. 0.0256 (3sf)	3d. 0.718 (3sf)	4d. 0.232 (3sf)

Part 5.3: Normal Distribution

Example
The lengths of piping in a storage shed are normally distributed with a mean of 3 m and a standard deviation of 0.65 m. They are marked with their length rounded to the nearest 0.1 m. What is the probability the piping is labelled as being 2.5 m long?

Answer
This time we leave the mean and the standard deviation alone. The only thing that changes is the values we are looking between. For this question, the smallest number that rounds to 2.5 is 2.45 and the largest number that rounds to 2.5 is 2.55. Therefore we need to find the probability the length is between 2.45 and 2.55. Using either our tables or a calculator we find this is 0.0456.

Exercise 5.3

1. The number of hair ties in a packet is normally distributed with a mean of 15 and a standard deviation of 2. What is the probability there are:
    a. More than 12 hair ties in the packet?
    b. Less than 15 hair ties in the packet?
    c. Exactly 14 hair ties in the packet?
2. The volume of a shampoo bottle is normally distributed with a mean of 383 mL and a standard deviation of 2 mL. If my measuring tools can measure to the nearest 1mL what is the probability there is:
    a. More than 385 mL in the bottle?
    b. Exactly 385 mL in the bottle?
    c. Less than the stated amount of 380 mL in the bottle?
3. The number of cotton buds in a box normally distributed with a mean of 200 and a standard deviation of 30. Calculate the probability the number of cotton buds in the box is:
    a. More than 150.
    b. Less than 290.
    c. Exactly 200.
4. The length of a zips in dresses are normally distributed with a mean of 45 cm and a standard deviation of 10 cm. Zips are only sold in 5 cm increments. What is the probability the zip is:
    a. More than 50 cm long?
    b. Less than 60 cm long?
    c. A 45 cm zip?
    d. A 10 cm zip?

Exercise 5.3 Answers

1a. 0.894 (3sf)	2a. 0.106 (3sf)	3a. 0.951 (3sf)	4a. 0.227 (3sf)
1b. 0.401 (3sf)	2b. 0.121 (3sf)	3b. 0.996 (3sf)	4b. 0.894 (3sf)
1c. 0.175 (3sf)	2c. 0.0000884 (3sf)	3c. 0.0133 (3sf)	4c. 0.197 (3sf)
			4d. 0.000489 (3sf)

Part 5.4: Mixed Questions

1. The age of a pine tree being cut down is between 3 and 40 years with the most common age being 28 years. The age of the tree is only known to the nearest year.
    a. Choose a distribution to model this situation and justify your choice.
    b. What is the probability a tree cut down is more than 30 years old?
    c. What is the probability a tree cut down is less than 20 years old?
    d. What is the probability a tree cut down is between 20 and 30 years old inclusive?
    e. In order to be in the oldest 25% of trees, how old does a tree that is cut down need to be?
2. The volume of nail polish in a bottle is normally distributed with a mean of 40 mL and a standard deviation of 10 mL. They are always measured to the nearest mL.
    a. What is the probability a bottle has less than 25 mL in it?
    b. Calculate the probability a bottle is stated as having 19 mL in it?
    c. What is the probability there is more than 25 mL in the bottle?
    d. Calculate the probability there between 30 mL and 35 mL in the bottle?
    e. What volume do the smallest 5% of bottles have in them?
3. Length of a skipping rope is known to be between 1.4 m and 3.2 m. Skipping ropes are manufactured in 20 cm increments.
    a. Choose a distribution to model this situation and justify your choice.
    b. What is the probability the skipping rope is more than 2 m long?
    c. Calculate the probability the skipping rope is exactly 2.4 m?
    d. What is the probability the skipping rope is less than 1.5 m long?
    e. What length must a skipping rope be in order to be in the largest 15% of ropes?
4. The most common length of song in my music library is 3:27. The maximum length is 14:48 and the shortest length is 0:22. All songs lengths are displayed to the nearest second.
    a. Choose a distribution to model this situation and justify your choice.
    b. What is the probability a song I select is more than 3 minutes long?
    c. What is the probability a song I select is less than 5 minutes long?
    d. If I have 1800 songs in my library, how many are likely to be more than 10 minutes long?
    e. What length must a song be shorter than to be in the bottom 10% of songs?

Part 5.4 Answers

1a. We are given three pieces of information, the minimum age (a) is 3 years (so a=2.5), and the maximum age (b) would be 40 years (so b=40.5), and the most likely age (c) is 28 years. Therefore a triangular distribution is best.
1b. 0.211 (3sf)
1c. 0.298 (3sf)
1d. 0.491 (3sf)
1e. 30 years or over
2a. 0.0606 (3sf)
2b. 0.00440 (3sf)
2c. 0.926 (3sf)
2d. 0.179 (3sf)
2e. 23 mL or less
3a. We only have two pieces of information, the maximum length (b) is 3.2m (b=3.25), and the minimum length (a) would be 1.4m (a=1.35). Therefore a uniform distribution is best.
3b. 0.632 (3sf)
3c. 0.0526 (3sf)
3d. 0.0526 (3sf)
3e. 3m or longer
4a. We are given three pieces of information, the minimum time (a) is 0:22 (so a=21.5 seconds), and the maximum age (b) would be 14:48 (so b=888.5 seconds), and the most likely time (c) is 3:27 (so c=207). Therefore a triangular distribution is best.
4b. 0.843 (3sf)
4c. 0.413 (3sf)
4d. 0.1404 x 1800 = 253 songs
4e. 148 seconds or less (2:28 or less)

Part 6: Discrete Random Variables

Quite often we are given probabilities for particular outcomes, but these do not follow a particular distribution. For these we create a table of with the probability for each particular outcome. This might look something like this:

The likelihood of winning different prizes in a raffle is shown below.

x	\$100	\$50	\$0
P(X=x)	1/100	2/100	97/100

From this we can see the probability of winning \$100 is 1/100, the probability of winning \$50 is 2/100 and the probability of not winning anything is 97/100. As you can see these probabilities all add up to one.

Part 6.1: Finding the Mean

The first thing that we need to do with these sorts of tables is find the mean. To find the mean you multiply each number by the corresponding probability and then add them all up. Let's look at two examples.

Example 1
The likelihood of winning different prizes in a raffle is shown below.

x	\$100	\$20	\$0
P(X=x)	1/100	3/100	q

Find q and calculate the mean of the prize amount.

Answer
The first thing we need to do is find q. As probabilities always add to one,

$$q = 1 - \frac{1}{100} - \frac{3}{100} = \frac{96}{100}$$

So the mean = $100 \times \frac{1}{100} + 20 \times \frac{3}{100} + 0 \times \frac{96}{100} = 1.6$
Therefore the mean (or expected value) of the prize from the raffle ticket is likely to be \$1.60.

Example 2
The length of time it takes to straighten hair is given in the table below.

x	0 ≤ x < 5	5 ≤ x < 10	10 ≤ x < 30
P(X=x)	0.2	0.5	0.3

Calculate the mean time it takes to straighten hair.

Answer
This question is slightly different as we are not given exact values, but ranges. The values we want to use in this situation is the midpoint of the range, so 2.5, 7.5 and 20 respectively.
Mean = $2.5 \times 0.2 + 7.5 \times 0.5 + 20 \times 0.3 = 10.25$
Therefore the mean time it takes to straighten hair is 10.25 minutes (or 10 minutes and 15 seconds).

Note: in order for the raffle to be 'fair' the cost of the ticket should be the same as the mean, and if the raffle is to make money the ticket price must be higher than the mean.

Exercise 6.1

1. The prizes in a lottery are shown in the table below.

x	\$20	\$10	\$0
P(X=x)	0.02	0.1	q

   a. Calculate the value of q
   b. What is the mean of the prizes??

2. Biscuits come in 3 packet sizes, small (12), medium (18) and large (30).

x	12	18	30
P(X=x)	0.3	0.5	0.2

   Calculate the mean number of biscuits in a packet based on the table above.

3. Toyota Corollas come in 3 engine sizes, 35% are 1.5 L engines, 50% are 1.8 L engines and the rest are 2.0 L engines.
   a. What percentage of Toyota Corollas have 2.0 L engines?
   b. Calculate the average engine size.

4. Stockings come in three lengths, small (147 – 161 cm), average (162 – 171 cm) and Tall (172 – 183 cm).

x	147 – 161	162 – 171	172 – 183
P(X=x)	0.6	m	0.2

   Calculate the mean length of the stockings.

Exercise 6.1 Answers

1a. 0.88	2. 18.60	3a. 15%	4. m = 0.2
1b. $1.40		3b. 1.725	so mean = 161.2

Part 6.2: Finding the Standard Deviation

The next thing we need to do is find the standard deviation. This is a two-step process where we need to find $E(X^2)$ (pronounced expected value of x-squared) and E(X) (the expected value of x, which is also the mean), this enables us to find the variance which is the standard deviation squared, so we just square root to get the standard deviation.

The formula you are given is $\sigma = \sqrt{E(X^2) - [E(X)]^2}$.

Example
The likelihood of winning different prizes in a raffle is shown below.

x	$100	$20	$0
P(X=x)	1/100	3/100	*q*

Find q and calculate the mean and standard deviation of the prize amount.

Answer
In part 6.1 we worked out $q = \frac{96}{100}$ and the mean, or E(X) = 1.6. This means to use our formula we need to calculate $E(X^2)$. To do this we need to work out what x^2 is. It is best to add a new row to the table.

x	$100	$20	$0
P(X=x)	1/100	3/100	*q*
x^2	10000	400	0

To work out $E(X^2)$ it is the same process as with the mean except instead of multiplying the probability and x we multiply the probability and x^2. This gives $E(X^2) = 10000 \times \frac{1}{100} + 400 \times \frac{3}{100} + 0 \times \frac{96}{100} = 112$.
We then substitute into the formula $\sigma = \sqrt{E(X^2) - [E(X)]^2} = \sqrt{112 - [1.6]^2} = \sqrt{109.44} = 10.46$ (4sf).
Therefore the mean = $1.60 and the standard deviation = 10.46.

Note: If you are asked to calculate the variance you just do not square root at the end, so for this example the variance is 109.44.

Exercise 6.2

Note, to save doing double calculations these are the same situations as exercise 6.1.

1. The prizes in a lottery are shown in the table below.

x	$20	$10	$0
P(X=x)	0.02	0.1	0.88

What is the standard deviation of the prize from the raffle?

2. Biscuits come in 3 packet sizes, small (12), medium (18) and large (30).

x	12	18	30
P(X=x)	0.3	0.5	0.2

Calculate the standard deviation of the number of biscuits in a packet based on the table above.

3. Toyota Corollas come in 3 engine sizes, 35% are 1.5 L engines, 50% are 1.8 L engines and the rest (15%) are 2.0 L engines. Calculate the standard deviation of the size of Toyota Corolla engines.
4. Stockings come in three lengths, small (147 – 161 cm), average (162 – 171 cm) and Tall (172 – 183 cm).

x	147 – 161	162 – 171	172 – 183
P(X=x)	0.6	*m*	0.2

Calculate the standard deviation of the length of the stockings.

Exercise 6.2 Answers

1. 4.00 (3sf)
2. 6.26 (3sf)
3. 0.179 (3sf)
4. 9.48 (3sf)

Part 6.3: Linear Combinations

You by now may have noticed the formulas to the right on your formula sheet. Let's look at them one at a time to see what they mean and an example for each one. They are designed for when we want to add two distributions together or for when we multiply a distribution or add things happening to a distribution. It is very important that for these sorts of distributions that you work out the variance before working out the standard deviation, as it is very easy to make mistakes which mean that you get the wrong answer if you try and do it all in one step.

$E[aX + b] = aE[X] + b$

$Var[aX + b] = a2Var[X]$

$E[aX + bY] = aE[X] + bE[Y]$

$Var[aX + bY] = a2Var[X] + b2Var[Y]$

If X, Y are independent

E[aX + b] = aE[X] + b
This formula means if I have 'a' times X plus 'b' the expected value, or the mean will be a times the expected value of X plus b.

Var[aX + b] = a²Var[X]
This formula is the matching one for variance (standard deviation squared) and means if I have 'a' times X plus 'b' the variance will be a squared times the expected value of X.

Note, we do not add 'b' on to the variance because the b does not affect how spread out the data is.

Example
A salesman earns commission of $10 for every TV he sells plus a fixed rate of $30 a day. If he sells on average 4.3 TVs per day with a standard deviation of 0.3 TVs, calculate the mean and standard deviation for how much he will earn each day.

Answer
Let's start by working out the mean, or E(aX + b). $\mathrm{E}[aX+b] = a\mathrm{E}[X] + b = 10 \times 4.3 + 30 = \$73.00$
Now let's work out the variance. The Var(X) = sd² = 0.3² = 0.09. $\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X] = 10^2 \times 0.09 = 9$
So the standard deviation = $\sqrt{9} = \$3.00$

E[aX + bY] = aE[X] + bE[Y]
This formula means that if I have 'a' times X plus 'b' times Y then the expected total is 'a' times the expected number of X plus 'b' times the expected number of Y.

Var[aX + bY] = a²Var[X] + b²Var[Y]
This is the matching formula for variance and means the variance of the sum is 'a' squared times the variance of X plus 'b' squared times the variance of Y.

Note: these last two formulas assume that X and Y are independent.

Example
In prescription glasses the weight of the lens is 9 grams with a standard deviation of 2 grams. The frames have an average weight of 6 grams with a standard deviation of 1 gram. If the material for the glass costs $3.20 per gram and the material for the frame costs $2.50 per gram, what is the mean and standard deviation for the cost of the materials in the glasses?

Answer
Lets start with the mean. $\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y] = \$3.20 \times 9 + \$2.50 \times 6 = \$43.80.$
Then the variance = $\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] = 3.20^2 \times 2^2 + 2.50^2 \times 1^2 = 47.21.$
(the 2² and 1² are because the variance = sd²)
This means the standard deviation of the total weight = $\sqrt{47.21} = \$6.87$ (3sf)

Note: when you have '3 lots of' you do not square the number of them, as you are just adding it on a number of times.

Exercise 6.3

1. The length of a worm is on average 7.3 cm with a standard deviation of 0.6 cm.
   a. If 10 worms were laid end to end how long would you expect the chain to be, and what would the standard deviation of this length be?
   b. In reality laying 10 worms end to end is quite tricky, so you need a 1cm gap between them (9cm total). With this gap what would the mean and standard deviation be for total length of the line?
2. The average time it takes for a relay runner to run 400 m is 55 seconds with a standard deviation of 3 seconds, how long would you expect it to take for a team of 4 runners to complete a 4 x 400 m relay, and what would the standard deviation of this time be, if there is a delay at the start of 5 seconds as the first runner gets up to speed?
3. A pair of kiwi birds is being transported. The average weight of a female kiwi is 2.91 kg with a standard deviation of 0.404. The average weight of a male is 2.26 kg with a standard deviation of 0.274. What is the mean and standard deviation for the total cost of transporting one male and one female if the different requirements mean that it costs $125 per kilo to transport males and $135 per kilo to transport females?
4. The weight of an average adult in NZ is 81.3 kg with a standard deviation of 6.2 kg.
   a. If you have 10 NZ adults how much are they likely to weigh in total, and what would the standard deviation be of this total weight?
   b. If 5 adults were in a lift with 4 suitcases with an average weight of 15 kg and a standard deviation of 2 kg, what is the total weight of the passengers and their suitcases?

Exercise 6.3 Answers

1a. μ = 73cm, σ = 1.90
1b. μ = 82cm, σ = 1.90
2. μ = 225 seconds, σ = 6 seconds
3. μ = \$675.35 (3sf), σ = \$64.40 (3sf)
4a. μ = 813 kg, σ = 19.6 kg (3sf)
4b. μ = 466.5 kg, σ = 14.4 kg (3sf

Part 6.4: Mixed Questions

1. Two raffles are being run as part of a fundraiser for a school. The first raffle has 500 tickets with a first prize of $500, a second prize of $300 and a third prize of $100. The second raffle has only 200 tickets with a first prize of $1000 a second prize of $500 and a third prize of $200.
    a. Calculate the mean and the standard deviation of the prize amount for the first raffle.
    b. Calculate the mean and the standard deviation of the prize amount for the second raffle.
    c. Which raffle would you expect to have a higher price?
    d. If the school wants to make at least $1000 from the first raffle and $500 from the second raffle, how much should they sell the tickets to for each raffle? (Prices must be whole dollar amounts)
    e. If I purchased 9 tickets to the first raffle and 4 tickets to the second raffle what would I expect my average winnings to be?
2. A large extended family has 3 boys and 5 girls going to the school ball this year. The mean amount that boys spend on their outfit is $213 with a standard deviation of $127. The mean amount girls spend on their outfit is $401 with a standard deviation of $246.
    a. What is the mean and standard deviation for the total amount spent by this extended family on their outfits?
    b. Assuming these amounts are normally distributed calculate the probability the extended family spends more than $3,000 on ball outfits this year?
    c. What assumptions are made in order for the answer in part b. to be calculated?
3. A business is working on two projects. Project A is a smaller job that is likely to cost between $10,000 and $90,000 but the business expects it to cost approximately $40,000. Project B is a larger job that is likely to cost between $100,000 and $300,000, but the best estimate for its cost is $220,000. The business only has $300,000 in its bank account to pay for the two projects.
   Calculate the mean and standard deviation of total cost of the two projects and using a normal approximation calculate the probability the business has enough money to pay for the two projects.
   State any assumptions you are making as you are doing your calculations.

Part 6.4 Answers

1a. μ = \$1.80, σ = \$26.40 (3sf)
1b. μ = \$8.50, σ = \$79.90 (3sf)
1c. The second raffle
1d. \$4 and \$11
1e. \$50.20

2a. μ =\$2644, σ = \$592 (3sf)
2b. 0.274 (3sf)
2c. We were assuming that the amount spent by one person was independent from the others (which may not be a good assumption in this case as they are all from one family, and how much they spend is likely to be affected by how much others in the family spend).

3. Project 1: σ = \$16,500, μ = \$46,700
Project 2: σ = \$41,100, μ = \$207,000
Combined: σ = \$44,288, μ = \$253,700
P(Total > \$300,000) = 0.148 (3sf)
We have assumed that the cost of one project doesn't affect the cost of the other project, and that a normal distribution is appropriate to model the combined costs.

Part 7: The Binomial Distribution

We use the binomial distribution in situations where the same thing happens a fixed number of times, and the outcome has the same probability each time. On the formula sheet you are given a number of formulas for the binomial distribution. These are:

$$\mathrm{P}(X = x) = \binom{n}{x}\pi^x(1-\pi)^{n-x} \qquad \mu = n\pi \qquad \sigma = \sqrt{n\pi(1-\pi)}$$

Now… what do these letters mean?

n = the number of 'trials' or the number of times the event occurs.
π = the probability of the particular outcome occurring.
x = the actual number that we care about.

We are going to look at several things, but there are two assumptions that this makes… it assumes that each event is independent from each other, and it only works if there are two outcomes, normally the event meets one criteria or it doesn't.

Part 7.1: Probability it is Exactly

Let's look at the first formula, $\mathrm{P}(X = x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$. Fortunately if you have a graphics calculator you don't need to worry about using this formula unless you are going for excellence. This formula works out the probability of an event meeting a criteria exactly x times when it has a probability of π of meeting the criteria and happens n times. Let's look at an example:

Example
I was at an outlet store and there was a watch sale on, the only catch was there was a 40% chance that the watch didn't work. I decided that I would buy 5 watches and then sell on whichever ones did work that I didn't need. What is the probability exactly 3 watches work?

Answer (Graphics Calculator)
Again we go into STATS (2) → DIST (F5) and this time we go into BINM (F5) and when we are working out the probability it is exactly we use Bpd (Binomial Point Distribution). In this case x = 3, Numtrial = 5 and p = 0.6 as the probability of failing is 0.4 so the probability it works is 0.6. This would look like this:

```
Binomial P.D
Data     :Variable
x        :3
Numtrial:5
p        :0.6
```

Which gives us an answer of 0.3456.

Answer (Formula)
We can see that x = 3, n = 5 and π = 0.6 as the probability of failing is 0.4 so the probability it works is 0.6. We substitute these into the formula
$\mathrm{P}(X = x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$ we get
$\mathrm{P}(X = 3) = \binom{5}{3}0.6^3(1-0.6)^{5-3} = 0.3456$
Note: $\binom{n}{x}$ *means* ${}^nC_x$

As you can see these two answers match up.

Exercise 7.1

1. The probability of a person being allergic to peanuts is 0.1. In a group of 10 students what is the probability:
   a. Exactly 3 are allergic?
   b. Exactly 2 are allergic?
   c. Exactly 1 is allergic?
2. In a manufacturing process 12% of items are faulty. What is the probability that there is exactly one faulty item:
   a. In a sample of 30?
   b. In a sample of 20?
   c. In a sample of 10?

Exercise 7.1 Answers

1a. 0.0574 (3sf)
1b. 0.194 (3sf)
1c. 0.387 (3sf)

2a. 0.0884 (3sf)
2b. 0.212 (3sf)
2c. 0.380 (3sf)

Part 7.2: Probability it is Less Than or Up To

The next part is very similar, but this time instead of working out the probability it is exactly, we want to work out the probability it is less than or up to a particular number. Let's look at an example:

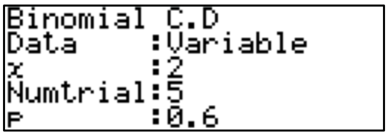
Example
I was at an outlet store and there was a watch sale on, the only catch was there was a 40% chance that the watch didn't work. I decided that I would buy 5 watches and then sell on whichever ones did work that I didn't need. What is the probability that less than 3 watches work?

Answer
The first thing I always do is draw a number line and highlight the numbers I want, this helps avoid making silly mistakes… remember the number line stats at zero.

0 1 2 3 4 5

This is super important… do it every time! Otherwise you will make mistakes.

(Graphics Calculator)
Again we go into STATS (2) → DIST (F5) and this time we go into BINM (F5) and when we are working out the probability it is up to we use Bcd (Binomial Cumulative Distribution). In this case x = 2 (the number we want up to), Numtrial = 5 and p = 0.6. This would look like this:



Which gives us an answer of 0.31744.

(Formula)
We can see that x = 0, 1 or 2, n = 5 and π = 0.6. This means we need to substitute into the formula 3 times, once for each vaule of x… we get

$$P(X = 2) = \binom{5}{2} 0.6^2 (1 - 0.6)^{5-2} = 0.2304$$

$$P(X = 1) = \binom{5}{1} 0.6^1 (1 - 0.6)^{5-1} = 0.0768$$

$$P(X = 0) = \binom{5}{0} 0.6^0 (1 - 0.6)^{5-0} = 0.01024$$

We then add these up giving 0.31744.

As you can see these two answers match up.

Exercise 7.2

1. The probability of having to stop at a traffic light is 0.4. If I drive through 5 sets of traffic lights on the way to work what is the probability:
    a. I don't have to stop
    b. I have to stop at 3 or less traffic lights?
    c. I have to stop at less than 3 traffic lights?
2. In a multi-choice test there are 5 possible answers to each question. What is the probability if you are guessing for all 10 questions that you get:
    a. None right?
    b. Less than 3 questions correct?
    c. 8 or less questions correct?
3. I have a box of 6 calculators. The probability of the calculator working is 0.9. What is the probability that:
    a. Less than 2 work?
    b. Up to 5 work?
    c. None work?
4. Each morning I make myself lunch. 60% of the time I make myself cheese sandwiches, the rest of the time I make something else. What is the probability in one week (5 days) I have:
    a. Less than 3 cheese sandwiches?
    b. Up to 2 days where I have something else for lunch?
    c. 1 or less days where I have cheese sandwiches?

Exercise 7.2 Answers

1a. 0.0778 (3sf)	2a. 0.107 (3sf)	3a. 0.000055	4a. 0.317 (3sf)
1b. 0.913 (3sf)	2b. 0.678 (3sf)	3b. 0.469 (3sf)	4b. 0.337 (3sf)
1c. 0.683 (3sf)	2c. 1.00 (3sf)	3c. 0.000001	4c. 0.0870 (3sf)

Part 7.3: Probability it is More Than

In the last section we look at the probability that up to, or less than a particular number occurred, now we want to know the probability that more than a certain number occurred. To do this we capitalise on the fact that probabilities always add up to one, so we calculate the probability of the bit we don't want (using the up to method) and then subtract it from 1. Let's look at an example.

Example
I roll a dice 8 times, what is the probability I get a 3 at least 5 times?

Answer
Again, the first thing I always do is draw a number line and highlight the numbers I want, this helps avoid making silly mistakes… remember the number line stats at zero.

0 1 2 3 4 5 6 7 8

(Graphics Calculator)
As you can see we want 5-8… the calculator doesn't work this out, so we need to work out up to and including 4 and then subtract this from 1.
We go into STATS (2) → DIST (F5) and this time we go into BINM (F5) and when we are working out the probability it is up to we use Bcd (Binomial Cumulative Distribution). In this case x = 4 (the number we want up to), Numtrial = 8 and p = 1/6. This would look like this:

```
Binomial C.D
Data     :Variable
x        :4
Numtrial:8
p        :0.16666666
```

This gives us an answer of 0.99539.
We then subtract this off one to get 0.00461

(Formula)
We can see that x = 5, 6, 7 or 8, n = 8 and π = $\frac{1}{6}$. This means we need to substitute into the formula 3 times, once for each vaule of x… we get

$$\mathrm{P}(X=5)=\binom{8}{5}\left(\frac{1}{6}\right)^{5}\left(1-\frac{1}{6}\right)^{8-5}=0.0041676$$
$$\mathrm{P}(X=6)=\binom{8}{6}\left(\frac{1}{6}\right)^{6}\left(1-\frac{1}{6}\right)^{8-6}=0.0004168$$
$$\mathrm{P}(X=7)=\binom{8}{7}\left(\frac{1}{6}\right)^{7}\left(1-\frac{1}{6}\right)^{8-7}=0.0000238$$
$$\mathrm{P}(X=8)=\binom{8}{8}\left(\frac{1}{6}\right)^{8}\left(1-\frac{1}{6}\right)^{8-8}=0.0000006$$

We then add these up giving 0.00461.

As you can see these two answers match up.

Exercise 7.3

1. The probability of catching a Frisbee when a long throw is made is 0.4. If two players make 10 throws, what is the probability that:
   a. More than 3 of them are caught?
   b. More than 6 of them are caught?
   c. 6 or more of them are caught?
   d. Exactly 4 were caught?
2. The probability of winning a tennis match is 0.7. If the player plays 10 matches, calculate the probability he:
   a. Wins at least half of them.
   b. Wins 6 or more of them.
   c. Wins all of them.
   d. Loses all of them?
3. The probability of a part being faulty is 1%. If 10 parts are chosen at random for inspection, what is the probability
   a. None of them are faulty?
   b. At least 1 of them is faulty?
   c. More than 5 of them are faulty?
   d. Less than 3 are faulty?
4. I flip a coin a number of times and record my results. What is the probability I get
   a. More than 3 heads if I flip it 5 times?
   b. Less than 2 heads if I flip it 3 times?
   c. 2 heads if I flip it 4 times?
   d. Between 3 and 5 heads (inclusive) if I flip it 8 times?

Exercise 7.3 Answers

1a. 0.618 (3sf)
1b. 0.0548 (3sf)
1c. 0.166 (3sf)
1d. 0.251 (3sf)

2a. 0.953 (3sf)
2b. 0.850 (3sf)
2c. 0.0282 (3sf)
2d. 0.00000590 (3sf)

3a. 0.904 (3sf)
3b. 0.0956 (3sf)
3c. 0.0000000242 (3sf)
3d. 1.00 (3sf)

4a. 0.188 (3sf)
4b. 0.5
4c. 0.375
4d. 0.711 (3sf)

Part 7.4: Mean and Standard Deviation

You may have noticed at the start I gave you three formulas… we are now going to use the last two of these… they are for calculating the mean and the standard deviation.

$$\mu = n\pi \qquad \sigma = \sqrt{n\pi(1-\pi)}$$

With these (which are given to us) we can very quickly and easily calculate the mean and standard deviation for a binomial distribution. Let's look at an example.

Example
I flip a coin 100 times, what is the mean and standard deviation for the number of heads I get?

Answer
The first thing we need to do is identify n and π. As n is the number of trials and we flip the coin 100 times, n = 100 and π is the probability of a head occurring which we know is 0.5. We then just substitute these into the formulas, making $\mu = n\pi = 100 \times 0.5 = 50$ and $\sigma = \sqrt{n\pi(1-\pi)} = \sqrt{100 \times 0.5(1-0.5)} = 5$

Exercise 7.4

For each of the following situations calculate the mean and the standard deviation.

1. The probability of a person being allergic to peanuts is 0.1 and I have a group of 10 students.
2. In a manufacturing process 12% of items are faulty. I need to take a sample of 30 people.
3. The probability of having to stop at a traffic light is 0.4. I drive through 5 sets of traffic lights on the way to work.
4. I have a box of 6 calculators. The probability of the calculator working is 0.9.
5. In a multi-choice test there are 5 possible answers to each question and there are 10 questions.
6. Each morning I make myself lunch. 60% of the time I make myself cheese sandwiches, the rest of the time I make something else. The time period we are looking at is one week (5 days).

Exercise 7.4 Answers

1. $\sigma = 0.949$ (3sf), $\mu = 1$
2. $\sigma = 1.78$ (3sf), $\mu = 3.6$
3. $\sigma = 1.10$ (3sf), $\mu = 2$
4. $\sigma = 0.735$ (3sf), $\mu = 5.4$
5. $\sigma = 1.26$ (3sf), $\mu = 2$
6. $\sigma = 1.10$ (3sf), $\mu = 3$

Part 7.5: Working Backwards and Using the Formula

Often we are given harder situations where we are required to work backwards, i.e. we are given the answer and we need to work out some of the parameters of the situation. Sometimes this is easiest to do by trial and error on the calculator, and other times we need to use the formula. Let's look at an example where we can use the formula, and then you'll get to do a series of questions where you can choose the approach.

Example
In a factory a quality control system is used to check 10 items at regular intervals. If one or more of them is faulty then the machine needs to be fixed. 88.5% of the time the machine is fine (i.e. no items are faulty in the sample). Calculate the percentage of items produced that are faulty based on these outcomes.

Answer
First of all we should write down what we know. We know x = 0, n = 10 and the answer should be 0.885. If we put all this information into the formula we get:

$$\mathrm{P}(X = x) = \binom{n}{x}\pi^x(1-\pi)^{n-x} = \binom{10}{0}\pi^0(1-\pi)^{10-0} = (1-\pi)^{10} = 0.885$$

If we put $(1-\pi)^{10} = 0.885$ into solver on the calculator (EQUA → SOLV) we find π = 0.01214 (4sf) which means 1.214% of the items produced are faulty.

Note: sometimes the easiest way to solve this is just by trial and error in the calculator.

Exercise 7.5

1. The probability of a student being sick on any given day is 5%.
   a. If the probability that 50 students are away from a school is 0.05779 how many students are at the school?
   b. A certain number of students are away today from a school of 550. This number of students being away occurs 6.65% of the time. How many students are away?
2. An office building has 930 light bulbs.
   a. If the probability that 31 light bulbs are blown is 0.0612, what is the probability that a light bulb blows?
   b. On 6.4% of days the number of light bulbs that are blown today happens, what number of light bulbs is blown if the probability a bulb blows is 4%.

Exercise 7.5 Answers

1a. 1000
1b. 30

2a. 0.0300 (3sf) or 0.0369 (3sf)
2b. 35

Part 7.6: Assumptions

When you are asked to justify the use of the binomial distribution you use the acronym FIST. If we look at the example of rolling the dice 6 times and getting a 3 a certain number of times:

Fixed number of trials – I roll the dice 6 times.

Independent – the result from each roll in independent from the other rolls.

Same – the probability of getting a 3 is always $\frac{1}{6}$.

Two Outcomes – there are two possible outcomes, it is either a three or it is not a three.

It is super important that you put all of the sections in context, rather than just stating what each letter means.

Part 7.7: Mixed Questions

1. The probability of a teacher being involved in extracurricular activities is 65%. If there are 130 teachers at a school:
    a. Choose a distribution to model this situation and justify your choice.
    b. What is the probability that at least half of the teachers are involved in extracurricular activities?
    c. What is the expected number of teachers involved in extracurricular activities?
    d. The school knows that the likelihood of having the number of teachers they actually do have involved in extracurricular activities is 1.126%. How many teachers do they have involved?
2. An experiment is being conducted to see if a coin is fair. It is thrown a certain number of times and the results are recorded if it is a heads or a tails.
    a. Choose a distribution to model this situation and justify your choice.
    b. If the coin is a fair coin and is thrown 100 times, what is the probability that less than 40 of them it comes up heads?
    c. If the coin was thrown 100 times, what would the mean and the standard deviation be for the expected number of heads if the coin is fair?
    d. The researchers did this experiment 1000 times, flipping the coin 100 times each experiment and recording the results. On 79 of the experiments 55 heads came up. Based on this what is the probability that the coin shows heads on each flip, and therefore is it a fair coin?
3. The probability that a morning is foggy is 5%.
    a. Choose a distribution to model the number of foggy mornings in a week and justify your choice.
    b. What is the probability that there are no foggy mornings in a week (7 days)?
    c. Calculate the probability that there are more than 2 foggy mornings in a week (7 days).
    d. If the probability of getting 2 foggy days over a period of time is 0.198. How many days is this period of time?
    e. What is the probability that on two consecutive weeks that there are more than two days each week that are foggy? Comment on the validity of any assumptions you have made.

Part 7.7 Answers

1a. A binomial distribution is best. This is because:

- There are a **F**ixed number of teachers at the school (130) so a fixed number of 'trials'
- The probability one teacher is involved is **I**ndependent from the other teachers
- The probability of each teacher being involved is always the **S**ame (0.65)
- There are **T**wo possible outcomes, either the teacher is involved or not.

1b. 1.00 (3sf)

1c. 84.5 so 85 teachers

1d. 95 teachers

2a. A binomial distribution is best. This is because:

- There are a **F**ixed number of coin flips so a fixed number of 'trials'
- The probability one flip is heads is **I**ndependent from the other flips
- The probability of each flip being heads is always the **S**ame (0.5 if it is fair)
- There are **T**wo possible outcomes, either heads or tails.

2b. 0.0176 (3sf)

2c. $\sigma = 5$, $\mu = 50$

2d. 0.542 or 0.558 (3sf) so therefore not a fair coin.

3a. A binomial distribution is best. This is because:

- There are a **F**ixed number of days we will look at so a fixed number of 'trials'
- The probability one flip is foggy can be assumed to be **I**ndependent from the other days
- The probability of each day being foggy is always the **S**ame (0.05)
- There are **T**wo possible outcomes, either the day is foggy or it is not.

3b. 0.698 (3sf)

3c. 0.00376 (3sf)

3d. 21 days

3e. 0.00376 x 0.00376 = 0.0000141

This assumes that each day, and each week is independent of each other. This may not be a valid assumption as bad weather often will hit several days in a row, and in winter you are more likely to get bad weather.

Part 8: The Poisson Distribution

We use the Poisson distribution when there is a certain number **per** something, for example the number of calls to a call centre **per** hour, or the number of worms **per** m^2 of grass. We are given a number of formulas:

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

Now… what do these letters mean?

λ = the mean number per a given amount (it is pronounced lambda)
x = the specific number we are looking at
e = a mathematical constant approximately equal to 2.71828… (it's on your calculator)

We are going to look at several things, but there are a couple of assumptions that this distribution makes, it assumes that the occurrence of the item is random and independent from other occurrences, but the big one is that it is when the number of items or events is proportional to the area or the time etc. It also assumes that the items or events cannot occur simultaneously.

Part 8.1: Probability it is Exactly

Let's look at the first formula: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$. This works out the probability of an event occurring exactly x times in a given period of time or amount if it occurs on average λ times. Let's look at an example:

Example
The number of calls to a call centre is on average 3.2 per minute. What is the probability there are exactly 25 calls in 10 minutes?

Answer
Well, the first thing we need to do is work out the average number of calls in 10 minutes. If there are 3.2 calls per minute, then in 10 minutes there will be $10 \times 3.2 = 32$ on average.

(Graphics Calculator)
We go to STAT (2) → DIST (F5) → Across (F6) → Poisn (F1). We want Ppd as we are after exactly. In this case x = 25 and μ = 32 which looks like this:

```
Poisson P.D
Data      :Variable
x         :25
μ         :32
Save Res:None
```

When we press calculate we get 0.0347 (3sf) which is the probability of getting exactly 25 calls in 10 minutes.

(Formula)
The formula is $(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$. If we substitute x = 25 and λ = 32 into the formula we get

$P(X = 25) = \frac{32^{25} e^{-32}}{25!} = 0.0347$ (3sf)

This gives us the same probability as when we used the calculator.

Exercise 8.1

1. The number of worms per m^2 of garden is on average 35. What is the probability there is exactly 70 worms in 2 m^2 of garden?
2. The number of detentions given out at a school is approximately 20 per day. What is the probability that in a week (5 days) there were exactly 90 detentions given out?
3. There have been 10 accidents on a 2km stretch of motorway in the last year. If there are 52 weeks in a year what is the probability on that 2km stretch there are no accidents this week?
4. The number of sandwiches that I eat for lunch is on average 2.5. Explain whether it is more likely for me to eat two or three sandwiches for lunch on any given day. (I always eat a whole number of sandwiches)

Exercise 8.1 Answers

1. 0.0476 (3sf)
2. 0.0250 (3sf)
3. 0.825 (3sf)
4. $P(x=2) = 0.257$ (3sf), $P(x=3) = 0.214$ (3sf)
So I am more likely to eat 2 sandwiches

Part 8.2: Probability it is Less Than or Up To

As well as working out the probability that it is exactly an amount, we need to know the probability it is up to or less than a certain amount. Let's look at an example:

Example
My car has broken down 10 times in the last 8 years. What is the probability it breaks down less than 3 times in the next year?

Answer
The first thing I always do is draw a number line and highlight the numbers I want, this helps avoid making silly mistakes… remember the number line stats at zero.

0 1 2 3 4 5 …

This is super important… do it every time! Otherwise you will make mistakes.

We then work out that the mean per year is $\frac{10}{8} = 1.25$ and the number we want up to (x) is 2.

(Graphics Calculator)
We go to STAT (2) → DIST (F5) → Across (F6) → Poisn (F1). We want Pcd as we are after up to. In this case x = 2 and μ = 1.25 which looks like this:

```
Poisson C.D
Data     :Variable
x        :2
μ        :1.25
Save Res:None
```

When we press calculate we get 0.868 (3sf) which is the probability of getting less than three break downs in the next year.

(Formula)
We can see that x = 0, 1 or 2 and λ = 1.25. This means we need to substitute into the formula 3 times, once for each vaule of x… we get

$$\mathrm{P}(X = 2) = \frac{1.25^2 e^{-1.25}}{2!} = 0.22383$$

$$\mathrm{P}(X = 1) = \frac{1.25^1 e^{-1.25}}{1!} = 0.35813$$

$$\mathrm{P}(X = 0) = \frac{1.25^0 e^{-1.25}}{0!} = 0.28650$$

We then add these up giving 0.868 (3sf).
As you can see these two answers match up.

Exercise 8.2

1. There are approximately 300 flights per day from Auckland airport. What is the probability there is:
   a. Less than 275 flights in a given day?
   b. Up to 310 flights in a given day?
   c. Exactly 300 flights in a given day?
   d. Less than 500 flights in two days?
2. The number of police callouts to a particular neighbourhood is on average 3 per week (7 days). What is the probability that there are:
   a. Up to 2 callouts in the next 2 days?
   b. Less than 4 callouts in the next 2 days?
   c. Exactly 1 callout in the next 2 days?
   d. No callouts in the next week?
3. The number of trees planted per hectare in a pine plantation is on average 300. What is the probability that in a forestry block there are:
   a. Less than 3100 trees in a 10 hectares?
   b. Up to 4400 trees in a 15 hectares?
   c. Exactly 3000 trees in a 10 hectares?
   d. Less than 100 trees in a half hectare block?
4. The number of bacteria in a probiotic tablet is on average 20,000. What is the probability there are:
   a. Less than 19,900 bacteria in the tablet?
   b. Up to 19,900 bacteria in the tablet?
   c. Exactly 20,000 bacteria in the tablet?
   d. Less than 40,000 bacteria in two tablets?

Exercise 8.2 Answers

1a. 0.0689 (3sf)
1b. 0.730 (3sf)
1c. 0.0230 (3sf)
1d. 0.0000123 (3sf)

2a. 0.944 (3sf)
2b. 0.989 (3sf)
2c. 0.364 (3sf)
2d. 0.0498 (3sf)

3a. 0.965 (3sf)
3b. 0.0686 (3sf)
3c. 0.00728 (3sf)
3d. 0.00000592 (3sf)

4a. 0.239 (3sf)
4b. 0.241 (3sf)
4c. 0.00282 (3sf)
4d. 0.499 (3sf)

Part 8.3: Probability it is More Than

Again, just like with the binomial, we often need to work out the probability that it is more than a certain amount. Let's look at an example.

Example
My car has broken down 10 times in the last 8 years. What is the probability it breaks down more than once in the next year?

Answer
Again, the first thing we should do is draw a number line and highlight the part that we want…

0 1 2 3 4 5 …

And from the last time we did this question we can remember that the mean is 1.25.

(Graphics Calculator)
We go to STAT (2) → DIST (F5) → Across (F6) → Poisn (F1). We want Pcd as we are after up to. In this case x = 1 and μ = 1.25 which looks like this:

```
Poisson C.D
Data     :Variable
x        :1
μ        :1.25
Save Res:None
```

When we press calculate we get 0.6446 (4sf) which is the probability of the part we do not want.
We then do 1 – 0.6446 to find the probability of the car breaking down more than once is 0.355 (3sf).

(Formula)
We can see that x is not 0 or 1 and λ = 1.25. This means we need to substitute into the formula 2 times, once for each vaule of x… we get

$$\mathrm{P}(X = 1) = \frac{1.25^1 e^{-1.25}}{1!} = 0.35813$$

$$\mathrm{P}(X = 0) = \frac{1.25^0 e^{-1.25}}{0!} = 0.28650$$

We then add these up giving 0.6446 (4sf).
We subtract from 1 giving 0.355 (3sf).
As you can see these two answers match up.

Exercise 8.3

1. The number of kiwi birds in a km² in a particular area is on average 20. Calculate the probability there are:
    a. More than 25 kiwi birds in a 1 km² area.
    b. 50 or more kiwi birds in a 3 km² area.
    c. Exactly 58 kiwi birds in a 3 km² area.
    d. Between 18 and 22 kiwi birds (inclusive) in a 1 km² area.
2. The number of books on a shelf in the library is on average 50. What is the probability there are:
    a. 60 or more books on a shelf
    b. More than 90 books on two shelves?
    c. Exactly 95 books on two shelves?
    d. Between 45 and 55 books (inclusive) on a shelf?
3. The number of cups of coffee sold at a café is on average 15 per hour. Calculate the probability the café sells:
    a. More than 20 cups of coffee in a hour.
    b. 10 or more cups of coffee in half an hour.
    c. Less than 5 cups of coffee in 15 minutes.
    d. Between 3 and 5 cups of coffee (inclusive) in 10 minutes.
4. The number of calls to a call centre is on average 10 per minute. What is the probability there are:
    a. More than 12 calls in a minute?
    b. 100 or more calls in 10 minutes?
    c. Less than 10 calls in a minute?
    d. Between 8 and 12 calls (inclusive) in a minute?

Exercise 8.3 Answers

1a. 0.157 (3sf)	2a. 0.0923 (3sf)	3a. 0.0830 (3sf)	4a. 0.208 (3sf)
1b. 0.916 (3sf)	2b. 0.829 (3sf)	3b. 0.224 (3sf)	4b. 0.513 (3sf)
1c. 0.0506 (3sf)	2c. 0.0360 (3sf)	3c. 0.678 (3sf)	4c. 0.458 (3sf)
1d. 0.424 (3sf)	2d. 0.563 (3sf)	3d. 0.414 (3sf)	4d. 0.571 (3sf)

Part 8.4: Mean and Standard Deviation

You may have also noticed that on the formula sheet we are given the formulas for the mean ($\mu = \lambda$) and the standard deviation ($\sigma = \sqrt{\lambda}$). We can use these to very quickly work out the standard deviation, as obviously the mean is already given to us in most situations. Let's look at an example:

Example
My car has broken down 10 times in the last 8 years. What is the mean and the standard deviation for the number of times it is likely to break down in the next year?

Answer
Well… we know from previously the average number of breakdowns is $\frac{10}{8} = 1.25$ and the standard deviation is $\sigma = \sqrt{\lambda} = \sqrt{1.25} = 1.118$ (4sf).

Exercise 8.4

1. The number of worms per m^2 of garden is on average 35. Calculate the mean and standard deviation for the number of worms in:
   a. 1 m^2.
   b. 5 m^2.
   c. 10 m^2.
2. The number of detentions given out at a school is approximately 20 per day. Calculate the mean and standard deviation for the number of detentions in:
   a. 1 day.
   b. 1 week (5 days).
   c. 1 term (10 weeks).
3. There have been 10 accidents on a 2km stretch of motorway in the last year. Calculate the mean and standard deviation for the number of accidents in:
   a. 1 year.
   b. 1 week.
   c. 1 month.
4. The number of sandwiches that I eat for lunch is on average 2.5. Calculate the mean and standard deviation for the number of sandwiches I eat in:
   a. 1 day.
   b. 1 week (7 days).
   c. 1 month (30 days).

Exercise 8.4 Answers

1a. $\mu = 35$, $\sigma = 5.92$ (3sf)
1b. $\mu = 175$, $\sigma = 13.2$ (3sf)
1c. $\mu = 350$, $\sigma = 18.7$ (3sf)

2a. $\mu = 20$, $\sigma = 4.47$ (3sf)
2b. $\mu = 100$, $\sigma = 10$ (3sf)
2c. $\mu = 1000$, $\sigma = 31.6$ (3sf)

3a. $\mu = 10$, $\sigma = 3.16$ (3sf)
3b. $\mu = 0.192$ (3sf), $\sigma = 0.439$ (3sf)
3c. $\mu = 0.833$ (3sf), $\sigma = 0.913$ (3sf)

4a. $\mu = 2.5$, $\sigma = 1.58$ (3sf)
4b. $\mu = 17.5$, $\sigma = 4.18$ (3sf)
4c. $\mu = 75$, $\sigma = 8.66$ (3sf)

Part 8.5: Working Backwards

Again, just like with the binomial, sometimes we are given the 'answer' and need to work backwards to find certain other numbers using the formula, $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$. Let's look at the most common type of question:

Example
A bank knows the likelihood of them having one or more customers arrive in a minute is 0.95. What is the mean number of customers arriving at the bank per minute?

Answer
The first thing we need to do is find the probability that x = 0 which in this case is 0.05. If we put this into the formula $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} = 0.05\ldots$ if we put this last part ($e^{-\lambda} = 0.05$) into solver we can find out that x = 2.9957 which means the average number of customers is likely to be 3 per minute.

Sometimes we will need to work out questions with different parameters (i.e. when x is not zero). If this comes about you need to work out what you are trying to find out, put the numbers into the formula (or the calculator) and find the solution.

Note: sometimes the easiest way to solve this is just by trial and error in the calculator.

Exercise 8.5

1. The average likelihood of having 1 or more defective products on a production line in an hour is 0.993. What is the mean number of defective products produced per hour?
2. A real-estate agent knows that on 4.98% of weeks he will sell no houses. What is the average number of houses that he will sell in **one year**?
3. A company claims that the likelihood of seeing at least one dolphin on its 1 day dolphin sightseeing cruise is 99.9%. What is the average number of dolphins seen on the 1 day cruise?
4. The probability a m^2 of carpet has a certain number of faults is 0.0437. If there are 0.85 faults per m^2 on average, what is the certain number?

Exercise 8.5 Answers

1. 4.96 defective products
2. 3 per week, so 156 per year
3. 6.9 dolphins
4. 3 faults

Part 8.6: Assumptions

When you are asked to justify the use of the Poisson distribution you use the acronym RIPS. If we look at the example of my car breaking down:

Random – the occurrence of breakdowns is random.
Independent – the occurrence of breakdowns is independent from other breakdowns.
Proportional – the occurrence of breakdowns is proportional to the amount of time.
Simultaneous – breakdowns cannot occur at the same time.

It is super important that you put all of the sections in context, rather than just stating what each letter means.

Part 8.7: Mixed Questions

1. The number of dishes dropped by a waiter during his 100 hours training is 4.
    a. Choose a distribution to model this situation and justify your choice.
    b. Calculate the probability:
        i. During the next hour he drops exactly one dish.
        ii. During the next three hours he drops no dishes.
        iii. In his next 100 hours of working he will drop no dishes. State whether the assumptions you made doing this calculation are valid.
    c. In what period of time will the probability of him dropping one dish be exactly 0.1?
2. The number of gumdrops in a 2 L tub of ice cream is on average 40.
    a. Choose a distribution to model this situation and justify your choice.
    b. Calculate the probability:
        i. That in a 300 mL bowl there are less than 3 gumdrops.
        ii. That in five 2 L tubs there are more than 220 gumdrops.
        iii. That a 2 L tub of ice cream has between 30 and 50 gumdrops (inclusive) in it.
    c. If the probability of getting one or more gum drops in a scoop is 0.135 what is the volume of a scoop?
3. A pizza company has just purchased a new machine which puts the toppings on its pizza. It is programmed to put on average 20 pieces of pepperoni on each pizza.
    a. Choose a distribution to model this situation and justify your choice.
    b. Calculate the probability:
        i. There are less than 20 slices of peperoni on a pizza.
        ii. 25 or more slices on a pizza.
        iii. Between 18 and 20 slices (inclusive) on a pizza.
    c. In a batch of 30 pizzas how many would you expect to be on the pizza with the most pieces?
4. The average number of customers that complain about the service in a store is 5 per day.
    a. Choose a distribution to model this situation and justify your choice.
    b. What is the probability that less than 20 customers complain in a week (7 days)?
    c. The manager wants to make it so that on 10% of days there are no customer complaints. How many complaints would there be on average per day if this is the case?

Part 8.7 Answers

1a. A Poisson distribution is best. This is because:
- The occurrence of droppages are **R**andom
- The occurrence of one droppage is **I**ndependent from the other droppages
- The number of droppages is **P**roportional to the amount of time spent waitering
- Two droppages cannot occur **S**imultaneously.

1bi. 0.0384 (3sf)
1bii. 0.887 (3sf)
1biii. 0.0183 (3sf). As the waiter gets more experienced you'd expect them to drop less dishes, so the independent assumption probably isn't valid, neither is the proportional.
1c. mean = 3.577 (4sf) so 89.4 hours (3sf)

2a. A Poisson distribution is best. This is because:
- The occurrence of gumdrops is **R**andom
- The occurrence of one gumdrops is **I**ndependent from the other gumdrops
- The number of gumdrops is **P**roportional to the amount of ice-cream.
- Two gumdrops cannot be in the same place (cannot occur **S**imultaneously).

2bi. 0.0620 (3sf)
2bii. 0.0753 (3sf)
2biii. 0.904 (3sf)
2c. mean = 0.145 (3sf), so volume = 7.25 mL

3a. A Poisson distribution is best. This is because:
- The occurrence of pieces of pepperoni is **R**andom
- The occurrence of one piece of pepperoni is **I**ndependent from the other pieces of pepperoni
- The number of pieces of pepperoni is **P**roportional to the size / number of pizzas
- Two pieces of pepperoni cannot be in the same place (cannot occur **S**imultaneously).

3bi. 0.470 (3sf)
3bii. 0.157 (3sf)
3biii. 0.262 (3sf).
3c. 28

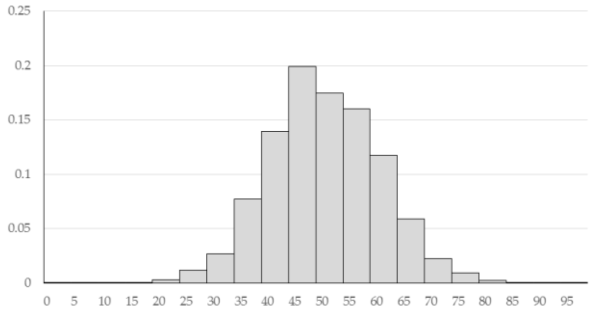
4a. A Poisson distribution is best. This is because:
- The occurrence of complaints is **R**andom
- The occurrence of a complaint is **I**ndependent from the complaints
- The number of complaints is **P**roportional to the amount of time
- Two complaints cannot occur **S**imultaneously.

4b. 0.00232 (3sf)
4c. 2.30 (3sf) complaints.

Part 9: Looking at Graphs

We are sometimes given graphs and ask to choose what distribution best fits. There are a few things that we can do to check. Below we will look at a few examples.

Example (Normal)



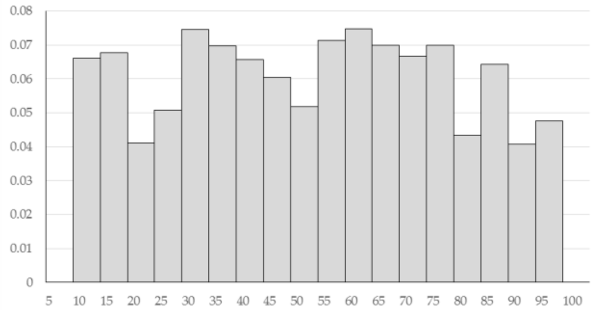
The normal distribution is a continuous distribution and therefore should be represented by a histogram. It should be reasonably even on both sides of the centre, i.e. symmetrical, and generally follow a bell shape with most of the data in the centre and less data towards the edge.

The two parameters we estimate are the mean and the standard deviation.

To estimate the mean look for where the centre of the data is, and if the bars are taller to the right or the left move your mean just slightly in that direction. For example, the graph above has a mean of approximately 48.

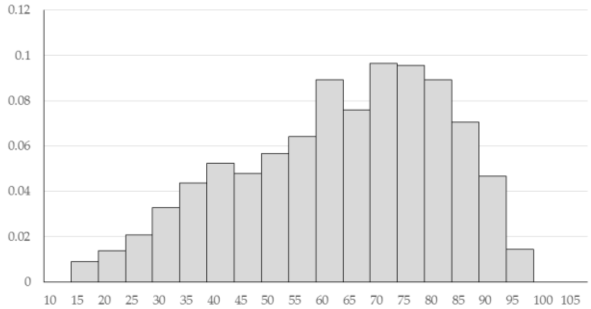
To estimate the standard deviation look for the smallest data point (in this case 20) and the largest data point (in this cast 80, as the data between 80 and 85 is too small to count) and then divide by 6. So in this case I would estimate the standard deviation to be 10 (60 ÷ 6).

Example (Rectangular)

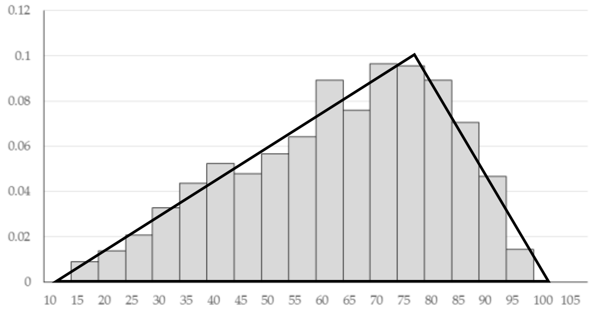


The rectangular distribution is also a continuous distribution and therefore should be represented by a histogram. It should be relatively even along, and then just suddenly stop. The two parameters that are needed are the minimum (in this case 10) and the maximum (in this case 100)

Example (Triangular)

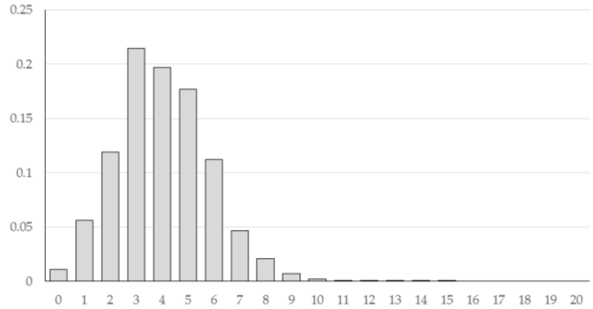


The triangular distribution is also a continuous distribution and therefore should be represented by a histogram. The triangular distribution is most useful when the data is skewed in one direction or the other and we need to work out three parameters, the minimum, the maximum and the most likely. Often the easiest way to do this is to draw a triangle over the data to see where these points occur. For example:



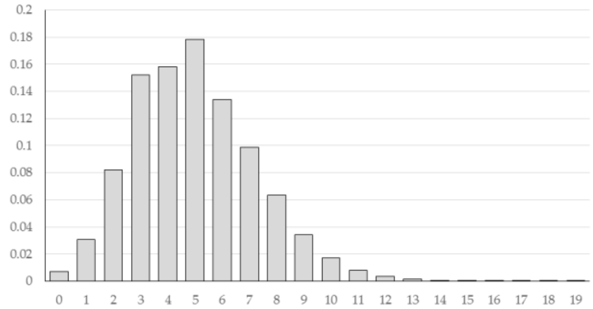
You want the lines of the triangle to evenly cut the open space and the bars as shown above. This would give us a minimum value of 12 (a), a maximum value of 102 (b) and a most likely value of 78 (c).

Example (Binomial)



The binomial distribution is a discrete distribution and therefore it should be represented by a bar graph (rather than a histogram).

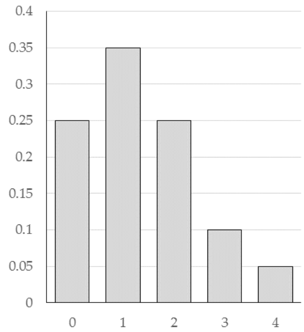
Example (Poisson)



The Poisson distribution is a discrete distribution and therefore it should also be represented by a bar graph (rather than a histogram).

As you may have noticed these two graphs are very similar to each other. The big difference between a binomial distribution and a Poisson distribution is a binomial has a fixed number of trials, so will therefore stop after a certain point, in the graph above you can see that there is no more data after 15, so there must be 15 trials, however the easiest way to work what type of distribution it is will be to look at the **context** of the question that will be given with it.

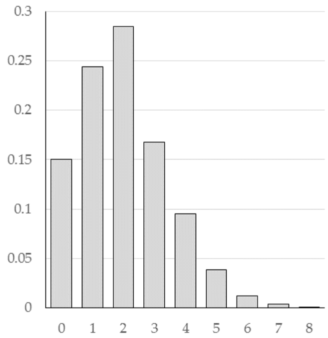
Part 9.1: Calculate the Mean for a Binomial or Poisson Distribution



The only other thing that you may need to do is calculate the mean. If we use the example on the right which is a binomial distribution, the number of trials is 4, and to work out the mean we can use it like a discrete variable.

$$\mu = 0 \times 0.25 + 1 \times 0.35 + 2 \times 0.25 + 3 \times 0.1 + 4 \times 0.05 = 1.35$$

For the binomial distribution we also know that $\mu = n\pi$.
If n = 4 and μ = 1.35 so to work out π we get π = 1.35 ÷ 4 = 0.3375



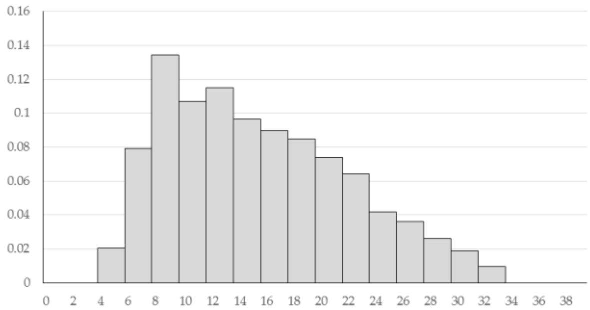
The other thing we sometimes might need to do is work out the mean for a Poisson distribution. To do this, because the Poisson keeps going on forever we need to use the formula $P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} = 0.15.$

If we solve $e^{-\lambda} = 0.15$ either using either algebra or equation mode on the calculator we get λ = 1.897 (4sf)

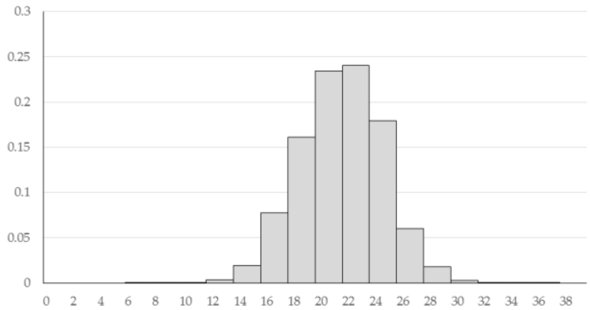
Part 9.2: Mixed Questions

For each of the graphs below state what distribution you think fits it best and give the parameters for it.
Give reasons for your choice of distribution.

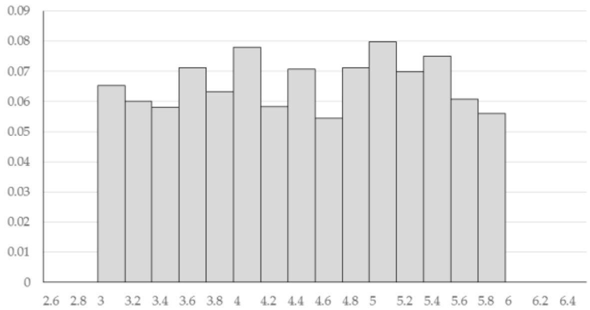
1. This is the amount of money earnt by a shop in an hour.



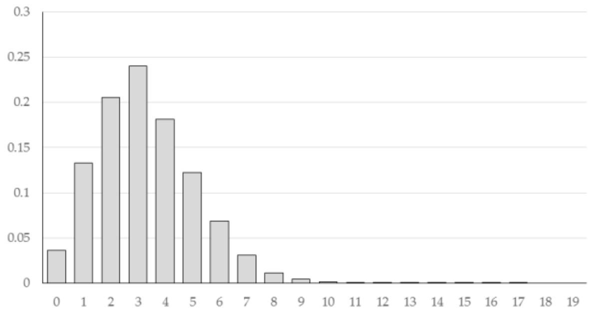
2. This is the length of leaves taken off a tree.



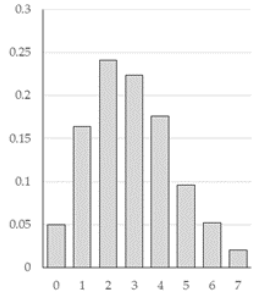
3. This is the amount of time it takes to get a transaction completed at the bank.



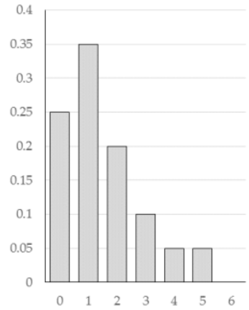
4. This is the number of people that turn up to a shop in an hour.



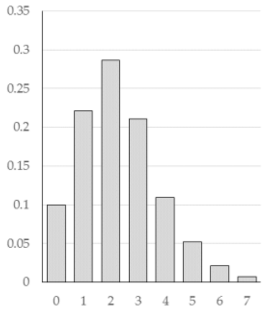
5. The distribution on the right is the number of fish per m^2 of water.



6. The distribution on the right has a fixed number of trials of 6.



7. The distribution on the right is for the mean number of worms per m^2 of garden.



Part 9.2 Answers

1. A triangular distribution would be best as the data is continuous and there is a very clear minimum and maximum, and the data is skewed to the right. The parameters would be minimum (a) is 4, maximum (b) is 34 and the modal value (c) would be 9.

2. A normal distribution would be best as the data is continuous and there is a very clear bell shape. The mean would be 22 and the standard deviation would be approximately 3 (20/6).

3. A rectangular distribution would be best as the data is continuous and the data has a very clear minimum and maximum, and is fairly flat in-between. The parameters would be minimum (a) is 3, maximum (b) is 6.

4. A Poisson distribution is best. This is because:
    - The occurrence of people turning up to the shop is **R**andom
    - The occurrence of one person turning up to the shop is **I**ndependent from the other people turning up to the shop
    - The number of people turning up to the shop is **P**roportional to the amount of time
    - Two people cannot turn up to the shop at exactly the same time (cannot occur **S**imultaneously).

   The mean for this distribution would be approximately 3.2 (2sf) based off a probability no one turns up of 0.04

5. A Poisson distribution is best. This is because:
    - The occurrence of fish is **R**andom
    - The occurrence of one fish is **I**ndependent from the other fish
    - The number of fish is **P**roportional to the amount of water
    - Two fish cannot be in the same place (cannot occur **S**imultaneously).

   The mean for this distribution would be approximately 3.0 (2sf) based off a probability no fish of 0.05

6. A binomial distribution is best. This is because we are told there is a fixed number of trials of 6. So number of trials (n) = 6, mean is 0.25 x 0 + 0.35 x 1 + 0.2 x 2 + 0.1 x 3 + 0.05 x 4 + 0.05 x 5 = 1.5 Therefore the probability for each event (π) is 1.5/6 = 0.25

7. A Poisson distribution is best. This is because:
    - The occurrence of worms is **R**andom
    - The occurrence of one worm is **I**ndependent from the other worms
    - The number of worms is **P**roportional to the amount of garden
    - Two worms cannot be in the same place (cannot occur **S**imultaneously).

   The mean for this distribution would be approximately 2.3 (2sf) based off a probability no worms of 0.1

Note: for excellence we would expect to see some calculations based on these parameters to back up how well the model fits the distribution.